

**KLASIFIKASI RETENSI KARYAWAN DENGAN ALGORITMA
C4.5 DAN TEORI HIERARKI KEBUTUHAN MASLOW
(STUDI KASUS : PT NEURONWORKS INDONESIA)**

TESIS

Disusun sebagai salah satu syarat untuk
memperoleh gelar Magister Komputer
dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Oleh:

AGUNG WIDYANGGA RETHANINDITA

NPM: 2019210095



**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2021**

**KLASIFIKASI RETENSI KARYAWAN DENGAN ALGORITMA
C4.5 DAN TEORI HIERARKI KEBUTUHAN MASLOW
(STUDI KASUS : PT NEURONWORKS INDONESIA)**

Oleh:

AGUNG WIDYANGGA RETHANINDITA

NPM: 2019210095

Bandung, 17 September 2021
Menyetujui,

Dr. Hery Heryanto, S.Kom., M.Kom.
Pembimbing

**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2021**

*To my mother, my wife Sabrina,
my daughter Alisha and my son Alif*

ABSTRAK

KLASIFIKASI RETENSI KARYAWAN DENGAN ALGORITMA C4.5 DAN TEORI HIERARKI KEBUTUHAN MASLOW (STUDI KASUS : PT NEURONWORKS INDONESIA)

Oleh:
AGUNG WIDYANGGA RETHANINDITA
NPM: 2019210095

Karyawan merupakan aset paling berharga bagi PT. Neuronworks Indonesia (Neuron) yang bergerak di bidang Jasa Konsultan Teknologi Informasi. Neuron mengimplementasikan teori hierarki kebutuhan Maslow untuk mengelola retensi karyawan dengan tujuan mempertahankan karyawan demi menjaga performansi dan untuk mencapai visi misi perusahaan. Walaupun teori hierarki kebutuhan Maslow merupakan teori awal berkaitan motivasi manusia namun masih relevan saat ini untuk membantu pengelolaan retensi karyawan di era modern khususnya perusahaan jasa. Neuron harus memperhatikan implementasi teori tersebut yang meliputi pemenuhan kebutuhan fisiologis, kebutuhan akan rasa aman dan ketenangan, hubungan sosial dan rasa cinta yang baik antar sesama karyawan, adanya sistem penghargaan dan pengakuan kepada karyawan serta pengadaan kegiatan untuk aktualisasi diri.

Evaluasi implementasi pengelolaan retensi karyawan menggunakan teori hierarki kebutuhan Maslow dapat dilakukan menggunakan *data mining* dengan metode klasifikasi. Beberapa penelitian *data mining* di bidang manajemen Sumber Daya Manusia (SDM) tidak menyelaraskan dengan teori manajemen SDM. Oleh karena itulah penelitian ini mengajukan solusi penggunaan klasifikasi berbasis teori hierarki kebutuhan Maslow sehingga mampu menghasilkan pengetahuan yang tepat bagi perusahaan untuk membantu perbaikan pengelolaan retensi karyawan.

Salah satu algoritma yang terkenal dalam klasifikasi adalah pohon keputusan khususnya algoritma C4.5, beberapa penelitian *data mining* di bidang manajemen SDM juga menggunakan klasifikasi dengan algoritma C4.5 yang menghasilkan tingkat akurasi tinggi. Oleh karena itu penelitian ini melakukan proses klasifikasi dengan algoritma C4.5 dengan label data status *resign* menggunakan *dataset* berdasarkan teori hierarki kebutuhan Maslow yang berelasi dengan program kerja di Neuron yang diekstraksi dari berbagai sumber data. Implementasi pemodelan *dataset* teori hierarki kebutuhan Maslow dengan algoritma pohon keputusan C4.5 tersebut menggunakan Rapidminer dengan percobaan tanpa dan menggunakan parameter *pruning* dan *prepruning*.

Evaluasi performansi percobaan tersebut menggunakan *cross validation*, menemukan bahwa algoritma C4.5 dengan parameter *pruning* dan *prepruning* mempunyai performansi terbaik dengan tingkat akurasi mencapai 96,12% dibandingkan tanpa parameter tersebut. Model dengan tingkat akurasi terbaik tersebut menghasilkan sebuah pengetahuan untuk membantu pengelolaan retensi karyawan dengan memberikan rekomendasi perbaikan pada tingkat hierarki kebutuhan Maslow secara berurutan berdasarkan prioritas yaitu aktualisasi diri, hubungan sosial dan cinta, rasa aman dan ketenangan serta kebutuhan fisiologis dasar. Kegiatan yang paling berpengaruh terhadap keputusan *resign* karyawan adalah kegiatan *upgrading* dan pelatihan. Sedangkan kegiatan penting lainnya yang perlu ditingkatkan adalah kegiatan ekstrakurikuler dan perbanyak kegiatan untuk menciptakan keakraban dan kesolidan karyawan.

Kata Kunci: neuron, sumber daya manusia, retensi karyawan, teori hierarki kebutuhan maslow, *data mining*, klasifikasi, pohon keputusan, algoritma C4.5

ABSTRACT

CLASSIFICATION OF EMPLOYEE RETENTION WITH C4.5 ALGORITHM AND MASLOW'S HIERARCHY OF NEEDS THEORY (CASE STUDY : PT NEURONWORKS INDONESIA)

By:

AGUNG WIDYANGGA RETHANINDITA
NPM: 2019210095

Employees are the most valuable asset for PT. Neuronworks Indonesia (Neuron), which is engaged in Information Technology consulting services. Neuron implements Maslow's Hierarchy of Needs theory to manage employee retention with the aim of retaining employees in order to maintain performance and to achieve the company's vision and mission. Although Maslow's Hierarchy of Needs theory is an early theory related to human motivation, it is still relevant today to help manage employee retention in the modern era, especially in service companies. Neurons must pay attention to the implementation of the theory, which includes the fulfillment of physiological needs, the need for security and tranquility, good social relations and love between fellow employees, the existence of a reward and feeling of accomplishment for employees, and the provision of activities for self-actualization.

Evaluation of the implementation of employee retention management using Maslow's Hierarchy of Needs theory can be done using data mining with the classification method. Some data mining research in the domain of Human Resource management (HR) does not combine and harmonize HR management theory with knowledge in the field of information technology. Therefore, this study proposes a solution using classification based on Maslow's Hierarchy of Needs theory so that it is able to produce appropriate knowledge for companies to improve employee retention management.

One of the well-known algorithms in classification is the decision tree, especially the C4.5 algorithm, which is also used in some data mining research in the domain of HR management with high accuracy performance results. Therefore, this study carried out the modeling process using the classification with the C4.5 algorithm with the resign status data label using a dataset based on Maslow's Hierarchy of Needs theory, which was related to the work program in Neuron extracted from various data sources. The implementation of Maslow's Hierarchy of Needs theory dataset modeling with the C4.5 decision tree algorithm uses Rapidminer with experiments without and using pruning and prepruning parameters.

Performance evaluation of the experiment using cross validation found that the C4.5 algorithm with pruning and prepruning parameters had the best performance with an accuracy rate of 96,12% compared to without these parameters. The model with the best level of accuracy generates knowledge to help manage employee retention by providing recommendations for improvement at the level of Maslow's hierarchy of needs sequentially based on priorities, namely self-actualization, social relationships and love, security and tranquility, and basic physiological needs. The activities that have the most influence on employee resignation decisions are upgrading and training activities. Meanwhile, other important activities that need to be improved are extracurricular activities and more activities to create employee intimacy and solidarity.

Keywords : neuron, human resource, employee retention, Maslow's Hierarchy of Needs theory, data mining, classification, decision tree, C4.5 algorithm

KATA PENGANTAR

Segala puji bagi Allah SWT atas limpahan rahmat dan petunjuk-Nya sehingga penulis dapat menyelesaikan tesis dengan judul : “KLASIFIKASI RETENSI KARYAWAN DENGAN ALGORITMA C4.5 DAN TEORI HIERARKI KEBUTUHAN MASLOW (Studi Kasus : PT Neuronworks Indonesia)” yang merupakan salah satu persyaratan untuk memperoleh gelar magister dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI pada program pascasarjana Magister Sistem Informasi.

Banyak pihak yang telah membantu sehingga tesis ini dapat terselesaikan dengan baik, untuk itu penulis mengucapkan dan memanjatkan puji syukur kepada Allah SWT yang telah memberikan kenikmatan berupa kesehatan dan karunia yang tak terhingga. Akhir kata ucapan terimakasih diucapkan kepada:

1. Bapak Dr. Hery Heryanto, S.Kom., M.Kom. sebagai pembimbing yang telah membantu dan membimbing penulis dalam menyelesaikan tesis ini.
2. Orang tua tercinta dan tersayang, ayahanda almarhum Ir. Roes Wardjito dan ibu Rahyuni Ekaningsih karena kasih sayangNya Ananda bisa sampai seperti saat ini.
3. Istri tercinta Sabrina yang telah pengertian dan memberikan dorongan positif serta anak-anakku tersayang Irdina Alisha Jasmin dan Muhammad Alif Ziaulhaq yang selalu memberikan energi bagi penulis untuk menyelesaikan tesis ini.
4. Rekan-rekan tim HCM di PT Neuronworks Indonesia yang telah meluangkan waktunya untuk membantu penulis dalam penyediaan dan pengumpulan data.
5. Pimpinan dan seluruh rekan-rekan di PT Neuronworks Indonesia yang selalu mendukung dan memberikan semangat.
6. Seluruh Bapak dan Ibu dosen program Magister Sistem Informasi STMIK LIKMI Bandung, atas ilmu yang tiada ternilai harganya.
7. Teman-teman seperjuangan seluruh mahasiswa S2 STMIK LIKMI yang selalu semangat dalam menyelesaikan studi, khususnya Irvan Hasbi Taufiq, yang telah bersama-sama bahu membahu, mengingatkan, memotivasi dan saling menguatkan untuk penyelesaian studi ini.

8. Seluruh staf pengelola dan karyawan program Magister Sistem Informasi STMIK LIKMI Bandung.
9. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu dalam penulisan tesis ini.

Bandung, 17 September 2021

Penulis

DAFTAR ISI

ABSTRAK.....	iii
<i>ABSTRACT</i>	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR RUMUS.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Ruang Lingkup Penelitian	4
1.5 Metode Penelitian.....	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	5
2.1 <i>Data Mining</i>	5
2.1.1 Tahap Pemahaman Bisnis	8
2.1.2 Tahap Pemahaman Data	8
2.1.3 Tahap Penyiapan Data	9
2.1.4 Tahap Pemodelan	13
2.1.5 Tahap Evaluasi	13
2.1.6 Tahap <i>Deployment</i>	14
2.2 Tipe <i>Data Mining</i>	14
2.2.1 Klasifikasi	16
2.2.2 Asosiasi	17
2.2.3 <i>Clustering</i>	17

2.3	<i>Algoritma Decision Tree</i>	18
2.4	<i>Algoritma C4.5</i>	21
2.5	<i>Cross Validation</i>	23
2.6	<i>Confusion Matrix</i>	24
2.7	Manajemen Sumber Daya Manusia	26
2.8	Retensi Karyawan	28
2.9	Teori Hierarki Kebutuhan Maslow	31
2.9.1	Kebutuhan Fisiologis.....	32
2.9.2	Rasa Aman dan Ketenangan.....	32
2.9.3	Hubungan Sosial dan Cinta	33
2.9.4	Penghargaan dan Pengakuan	34
2.9.5	Aktualisasi Diri.....	34
2.10	Rapidminer	35
2.11	Hasil Penelitian Terdahulu	36
BAB III OBJEK DAN METODOLOGI PENELITIAN		40
3.1	Profil PT Neuronworks Indonesia.....	40
3.1.1	Visi Misi dan Produk Layanan.....	41
3.1.2	Struktur Organisasi	42
3.1.3	Proses Retensi Karyawan.....	46
3.2	Metodologi Penelitian	48
3.3	Data Penelitian	51
BAB IV HASIL DAN PEMBAHASAN		54
4.1	Tahapan Pemahaman.....	54
4.1.1	Pemahaman Proses Bisnis	54
4.1.2	Pemahaman Data	58
4.2	Penyiapan Data.....	61
4.2.1	Ekstraksi Data	62
4.2.2	<i>Data Integration</i> dan <i>Data Reduction</i>	68
4.2.3	<i>Data Cleaning</i>	69

4.2.4	<i>Data Transformation</i>	73
4.2.5	<i>Preprocessing Lanjutan</i>	79
4.3	Pemodelan	80
4.4	Evaluasi	84
4.4.1	Penggunaan <i>Cross Validation</i>	85
4.4.2	Hasil Pengukuran Performansi	86
4.4.3	Interpretasi Model Terpilih.....	89
BAB V KESIMPULAN DAN SARAN		94
5.1	Kesimpulan.....	94
5.2	Saran	95
DAFTAR PUSTAKA.....		97
LAMPIRAN-LAMPIRAN		101

DAFTAR GAMBAR

Gambar 2.1	Tahapan proses penemuan pengetahuan.....	6
Gambar 2.2	Tahapan <i>data mining</i> berdasarkan CRISP-DM	7
Gambar 2.3	Contoh pohon keputusan untuk permainan Golf	19
Gambar 2.4	Proses pembagian data pada <i>Cross Validation</i>	24
Gambar 2.5	Teori hierarki kebutuhan Maslow	31
Gambar 2.6	<i>Magic quadrant data science and machine learning</i>	35
Gambar 3.1	Logo PT Neuronworks Indonesia.....	40
Gambar 3.2	Intepretasi logo dan keunggulan perusahaan.....	42
Gambar 3.3	Struktur Organisasi PT. Neuronworks Indonesia.....	43
Gambar 3.4	Proses bisnis pengelolaan retensi di Neuron	47
Gambar 3.5	Alur tahapan penelitian	49
Gambar 4.1	Komposisi pendapatan Neuron.....	55
Gambar 4.2	Komposisi karyawan operasional dan non operasional	56
Gambar 4.3	Penyelarasan proses <i>data mining</i> dan teori hierarki Maslow	57
Gambar 4.4	Sebaran data tingkat pendidikan sebelum <i>data cleaning</i>	70
Gambar 4.5	Sebaran data tingkat pendidikan setelah <i>data cleaning</i>	70
Gambar 4.6	Sebaran data asal sekolah sebelum <i>data cleaning</i>	71
Gambar 4.7	Sebaran data asal sekolah setelah <i>data cleaning</i>	71
Gambar 4.8	Hasil <i>import dataset</i> dan pengecekan statistik <i>missing values</i>	80
Gambar 4.9	Proses pemodelan penentuan target <i>role</i> pada atribut <i>resign</i>	81
Gambar 4.10	Proses pemodelan menggunakan C4.5	81
Gambar 4.11	Hasil pemodelan C4.5 tanpa <i>pruning</i> dan <i>prepruning</i>	82
Gambar 4.12	Hasil pemodelan C4.5 dengan <i>pruning</i>	83
Gambar 4.13	Hasil pemodelan C4.5 dengan <i>pruning</i> dan <i>prepruning</i>	84
Gambar 4.14	Evaluasi pemodelan menggunakan <i>Cross Validation</i>	85
Gambar 4.15	Detail proses evaluasi di dalam <i>Cross Validation</i>	86
Gambar 4.16	Grafik hasil evaluasi 3 percobaan pemodelan	87

DAFTAR TABEL

Tabel 1.1	Data karyawan <i>resign</i>	1
Tabel 2.1	Teknik-teknik yang dipakai dalam <i>data mining</i> bidang SDM	15
Tabel 2.2	Performansi algoritma Klasifikasi analisis data SDM	21
Tabel 2.3	<i>True</i> dan <i>false positives</i> dan <i>negatives</i>	25
Tabel 2.4	Rumus pengukuran hasil proses klasifikasi	26
Tabel 2.5	Faktor retensi karyawan	29
Tabel 2.6	Penelitian terkait <i>data mining</i> SDM terdahulu	36
Tabel 3.1	Program kerja Neuron berdasarkan Hierarki Maslow	48
Tabel 3.2	Atribut-atribut data dan sumber data	52
Tabel 4.1	Contoh variasi data tiap tahun	58
Tabel 4.2	Pemahaman <i>dataset</i> pengelolaan retensi karyawan	60
Tabel 4.3	<i>Dataset</i> demografi	63
Tabel 4.4	<i>Dataset</i> hierarki Maslow kebutuhan fisiologis	63
Tabel 4.5	<i>Dataset</i> hierarki Maslow rasa aman dan ketenangan.....	65
Tabel 4.6	<i>Dataset</i> hierarki Maslow hubungan sosial dan cinta.....	66
Tabel 4.7	<i>Dataset</i> hierarki Maslow penghargaan atau pengakuan	66
Tabel 4.8	<i>Dataset</i> hierarki Maslow aktualisasi diri	67
Tabel 4.9	Transformasi <i>dataset</i> demografi karyawan	74
Tabel 4.10	Transformasi <i>dataset</i> hierarki Maslow kebutuhan fisiologis	74
Tabel 4.11	Transformasi <i>dataset</i> hierarki Maslow rasa aman dan ketenangan	75
Tabel 4.12	Transformasi <i>dataset</i> hierarki Maslow hubungan sosial dan cinta	77
Tabel 4.13	Transformasi <i>dataset</i> hierarki Maslow penghargaan dan pengakuan.....	78
Tabel 4.14	Transformasi <i>dataset</i> hierarki Maslow aktualisasi diri	78
Tabel 4.15	Hasil performansi pohon keputusan algoritma C4.5.....	86
Tabel 4.16	<i>Confusion matrix</i> algoritma C4.5 tanpa <i>pruning</i> dan <i>prepruning</i>	88
Tabel 4.17	<i>Confusion matrix</i> algoritma C4.5 hanya dengan <i>pruning</i>	88
Tabel 4.18	<i>Confusion matrix</i> algoritma C4.5 dengan <i>pruning</i> dan <i>prepruning</i>	88

Tabel 4.19 Keterhubungan model pengetahuan dan hierarki Maslow	90
Tabel 4.20 Jumlah kemunculan di model Pohon Keputusan.....	92
Tabel 4.21 <i>Rule</i> hasil algoritma pohon keputusan C4.5	92

DAFTAR RUMUS

Rumus 2.1 Menghitung nilai <i>Gain</i>	22
Rumus 2.2 Menghitung nilai <i>Entropy</i>	23

DAFTAR LAMPIRAN

Lampiran 1	<i>Dataset</i> siap untuk pemodelan (<i>transforming data</i>)	102
Lampiran 2	Detail pengecekan <i>dataset</i> secara statistik	112
Lampiran 3	Tangkapan layar percobaan pemodelan	117
Lampiran 4	Hasil pemodelan berbentuk deskripsi (<i>rule model</i>)	119
Lampiran 5	Tangkapan layar hasil evaluasi performansi algoritma C4.5	124

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sumber Daya Manusia (SDM) adalah jantung dari sebuah organisasi dan manajemen SDM adalah salah satu pilar sebuah perusahaan. Pengelolaan SDM yang baik menjadi sangat penting. Beberapa masalah dalam pengelolaan SDM antara lain merekrut karyawan yang tidak sesuai, penentuan pemimpin yang salah, penetapan penghargaan atau hukuman yang tidak sesuai dan kompensasi yang tidak dapat memotivasi karyawan serta program-program manajemen SDM yang tidak efektif dalam mengelola karyawan. Hal di atas dapat memicu adanya kegagalan dalam pengelolaan retensi karyawan. Retensi karyawan sangat kritikal untuk kesehatan jangka panjang sebuah perusahaan (Lahida, et.al., 2017).

PT Neuronworks Indonesia (Neuron) adalah perusahaan jasa konsultan IT dimana aset paling berharga di perusahaan jasa tersebut adalah SDM. Neuron berusaha melakukan program kerja yang dapat memelihara kenyamanan dan motivasi SDM sehingga retensi karyawan dapat terjaga dengan baik. Neuron menggunakan teori hierarki kebutuhan Maslow dalam pengelolaan SDM yang ditujukan untuk memotivasi karyawan, baik untuk mempertahankan loyalitas karyawan sehingga mampu mempertahankan karyawan terbaik sampai meningkatkan performansi kerjanya. Lima tahun terakhir terdapat 21 karyawan yang mengundurkan diri, dengan rata-rata per tahun ada 3-4 karyawan yang mengundurkan diri sebagaimana terlihat pada Tabel 1.1 Data karyawan *resign*. Hal ini tentu saja mengganggu proses produksi dan penyelesaian pekerjaan untuk menjaga performansi perusahaan. Kebutuhan akan analisa yang tepat sejauh mana konsep teori hierarki kebutuhan Maslow diterapkan menjadi penting sehingga pengelolaan SDM berdasarkan teori hierarki kebutuhan Maslow tersebut dapat ditingkatkan dan semakin menjaga motivasi karyawan khususnya untuk mempertahankan loyalitas karyawan.

Tabel 1.1 Data karyawan *resign*

	2016	2017	2018	2019	2020	2021
Jumlah <i>Resign</i>	1	1	6	5	5	3

Pengelolaan SDM akan dapat dikerjakan dengan baik jika data-data berkaitan SDM tersedia dan dapat dianalisa untuk membuat sebuah keputusan. Data-data seperti jumlah kompensasi atau penghargaan apa yang memotivasi dan seberapa besar motivasi karyawan untuk disiplin masuk kerja dapat membantu pembuatan kebijakan pengelolaan SDM dengan menganalisa data-data tersebut. *Human Resource Analytics* (Data Analitik SDM) merupakan sebuah pendekatan multidisiplin untuk menggabungkan metodologi untuk meningkatkan kualitas keputusan yang berkaitan individu dengan tujuan untuk meningkatkan performansi SDM dan organisasi (Mishra S N, et.al., 2016). Data analitik SDM tidak hanya membuat SDM lebih baik, tetapi hal tersebut menyebabkan bisnis secara keseluruhan menjadi lebih baik (Meena dan Parimalarani, 2019). Di beberapa penelitian, analisa data SDM ini memakai *data mining* sebagaimana salah satunya diungkapkan oleh Jaffar, Waheed dan Kanwal (2019). Selalu sulit dan terkadang di luar kendali departemen SDM untuk mempertahankan karyawan yang terampil, tetapi *data mining* dapat berperan untuk memprediksi karyawan yang ingin meninggalkan perusahaan sehingga departemen SDM mendapatkan strategi intervensi yang tepat atau mencari alternatif solusinya (Jaffar Z, et.al., 2019). Namun sayangnya, banyak praktisi SDM tidak memahami data analitik atau *Big Data*, sedangkan di sisi lain praktisi data analitik tidak memahami pengelolaan SDM sehingga kemampuan analisis yang disediakan oleh Sistem Informasi SDM terbaru gagal dalam menyediakan kemampuan analisis strategi SDM (Angrave, et.al., 2016). Untuk itulah, penelitian ini mencoba menggabungkan teori dasar pengelolaan SDM yaitu teori hierarki kebutuhan Maslow dan klasifikasi dalam *data mining* untuk menemukan pengetahuan kemudian memberikan rekomendasi-rekomendasi pengelolaan retensi karyawan di Neuron.

Data mining telah lama digunakan untuk menghasilkan sebuah pengetahuan baru dari sejumlah data yang banyak dimana sulit dilakukan analisa data secara manual oleh manusia. *Data mining* adalah sebuah tahapan proses penemuan pengetahuan yang berisi analisis dan metode untuk menemukan pola, kecenderungan, dan hubungan implisit data-data yang ada (Sikaroudi dan Ghousi, 2015). Strategi dan metode pengelolaan SDM dibutuhkan untuk mengukur dan memodelkan data dalam membangun model analitik

dalam *data mining* yang sesuai dengan metode pengelolaan SDM. Data-data dikumpulkan sebagai *dataset* kemudian dihubungkan ke dalam pemahaman data sebelum dilakukan sebuah analisa dan prediksi untuk membantu membuat pengetahuan pengelolaan retensi karyawan. *Dataset* yang akan digunakan pada penelitian ini berdasarkan teori hierarki kebutuhan Maslow.

Ilmu psikologi dan pengelolaan SDM mengenal teori hierarki kebutuhan Maslow dimana jika perusahaan dapat memenuhi kebutuhan manusia dimulai dari tingkat paling bawah sampai tingkat paling atas maka karyawan akan termotivasi untuk meningkatkan kinerjanya dan menjaga tingkat retensi karyawan. Penerapan teori hierarki kebutuhan Maslow dapat mengurangi tingkat *turnover* karena perusahaan mengetahui apa yang mereka butuhkan, motivasi karyawan dalam mengerjakan pekerjaannya juga meningkat dan secara jelas dapat mempertahankan karyawan (Lahida, et.al., 2017). Berdasarkan permasalahan dan referensi di atas penulis menggunakan teori hierarki kebutuhan Maslow untuk mendesain *data model* SDM yang selanjutnya dijadikan *dataset* kemudian digunakan sebagai bahan analisis dalam data analitik menggunakan *machine learning* di dalam *data mining* dengan algoritma klasifikasi untuk dibuat sebuah kesimpulan dan rekomendasi yang berkaitan dengan pengelolaan retensi karyawan. Hal tersebut dapat menghasilkan klasifikasi, karyawan mana yang loyal dan tidak berdasarkan tingkat hierarki kebutuhan Maslow serta faktor apa yang mempengaruhinya.

Informasi hasil klasifikasi dari *data mining* dengan *dataset* berdasarkan teori hierarki kebutuhan Maslow dapat digunakan untuk mengetahui tingkat hierarki kebutuhan Maslow mana yang perlu dilakukan lebih baik terhadap karyawan yang mempunyai peluang berhenti bekerja. Informasi dan pengetahuan tersebut akan berguna bagi Neuron untuk membantu pengelolaan retensi karyawan. Pada tahun 2019, Jaffar melakukan penelitian *data mining* untuk mengukur performansi algoritma klasifikasi menggunakan data SDM dengan hasil bahwa algoritma J48 memiliki performansi paling tinggi. Di dalam Rapidminer, aplikasi *data science* yang berdasarkan Gartner masuk dalam kuadran *visioner*, algoritma J48 dikenal dengan algoritma C4.5. Algoritma C4.5 seperti halnya J48 merupakan algoritma klasifikasi yang paling populer dan dikenal memiliki akurasi yang cukup baik.

Proses klasifikasi berdasarkan *dataset* teori hierarki kebutuhan Maslow di Neuron akan menggunakan salah satu algoritma *data mining* yaitu C4.5. Berdasarkan latar belakang tersebut, penelitian ini mengambil judul yaitu "*Klasifikasi Retensi Karyawan dengan Algoritma C4.5 dan Teori Hierarki Kebutuhan Maslow (Studi Kasus : PT Neuronworks Indonesia)*".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas dapat dirumuskan permasalahan, antara lain:

1. Atribut apa saja yang dijadikan acuan dalam melakukan dan mempengaruhi klasifikasi data retensi karyawan berbasis teori hierarki kebutuhan Maslow di PT Neuronworks Indonesia?
2. Bagaimana membuat rekomendasi perbaikan pengelolaan retensi karyawan di PT Neuronworks Indonesia menggunakan klasifikasi berbasis teori hierarki kebutuhan Maslow?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan sebelumnya, maka penelitian ini mempunyai tujuan, antara lain:

1. Memodelkan *dataset* berdasarkan teori hierarki kebutuhan Maslow yang dapat digunakan dalam proses klasifikasi.
2. Membuat rekomendasi perbaikan pengelolaan retensi karyawan di PT Neuronworks Indonesia menggunakan klasifikasi berbasis teori hierarki kebutuhan Maslow.

1.4 Ruang Lingkup Penelitian

Ruang lingkup penelitian agar yang dibahas pada penelitian ini lebih fokus maka mempunyai beberapa batasan sebagai berikut:

1. Algoritma untuk pemodelan menggunakan 1 algoritma *data mining* yaitu C4.5.

2. Rekomendasi pengetahuan yang diusulkan berkaitan dengan pengelolaan retensi karyawan.

1.5 Metode Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif karena sesuai dengan latar belakang yang telah dipaparkan sebelumnya bahwa dengan proses *data mining* maka mengharuskan data yang terlibat mengandung atribut kuantitatif dan pengolahan data tersebut menggunakan algoritma *data mining* yang merupakan representatif dari perhitungan matematis. Metode penelitian yang dilakukan mengikuti proses standar pengembangan *data mining* yaitu CRISP-DM yang mempunyai tahapan antara lain:

1. Tahap Pemahaman Bisnis

Tahap ini merupakan proses penentuan rumusan masalah serta tujuan penelitian kemudian diikuti melakukan studi literatur berkaitan *data mining* dan manajemen SDM agar memahami bidang yang akan diteliti khususnya berkaitan retensi karyawan dan teori hierarki kebutuhan Maslow.

2. Tahap Pemahaman Data

Proses yang dilakukan pada tahap ini yaitu mempelajari, mengenali, memahami data yang terlibat dan menentukan data yang akan digunakan beserta lokasi objek penelitian dimana data tersebut diperoleh.

3. Tahap Penyiapan Data

Tahap pengumpulan data-data sesuai atribut yang telah ditentukan sesuai pemahaman di tahap 2, kemudian melakukan proses *pre-processing* agar data dapat diolah menggunakan proses *data mining*.

4. Tahap Pemodelan

Pada tahap ini, data yang telah disiapkan pada tahap penyiapan data digunakan untuk melakukan implementasi algoritma *data mining* menggunakan pohon keputusan algoritma C4.5 untuk menghasilkan model pengetahuan.

5. Tahap Evaluasi

Proses pengukuran kuantitatif terhadap hasil pemodelan berdasarkan tahap pemodelan. Setelah pengukuran dilakukan interpretasi hasil evaluasi dan memilih model yang mempunyai nilai akurasi tertinggi serta menerjemahkan model tersebut ke pengetahuan yang akan dipakai. Pemilihan model dan penentuan pengetahuan yang akan direkomendasikan dimasukkan dalam kesimpulan dan saran yang diletakkan di sistematika penulisan Bab V Kesimpulan dan Saran.

1.6 Sistematika Penulisan

Berdasarkan latar belakang, rumusan masalah dan tujuan penelitian maka penelitian ini menggunakan sistematika penulisan sebagai berikut:

BAB I : Pendahuluan

Bab ini berisi penjelasan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, batasan penelitian, metode penelitian dan sistematika penulisan.

BAB II : Tinjauan Pustaka

Bab ini menjelaskan mengenai teori-teori yang berkaitan dan digunakan dalam penulisan tesis antara lain *data mining* dan pengelolaan SDM khususnya pengelolaan retensi karyawan.

BAB III : Objek dan Metode Penelitian

Bab ini menjelaskan gambaran PT Neuronworks Indonesia sebagai tempat penelitian dilakukan dan sumber data untuk penelitian yang digunakan, alur penelitian dengan tahapannya serta atribut data yang akan digunakan berdasarkan demografi karyawan dan data yang berelasi dengan teori hierarki kebutuhan Maslow.

BAB IV : Hasil dan Pembahasan

Bab ini berisi uraian analisa proses bisnis kaitannya dengan data, proses penyiapan data dan pembahasan proses klasifikasi menggunakan algoritma C4.5 untuk pengelolaan retensi menggunakan *dataset* yang ditentukan di Bab III.

BAB V : Kesimpulan dan Saran

Pada Bab ini penulis menyampaikan kesimpulan dan saran dari hasil penelitian sebagai hasil akhir dari penelitian tema *data mining*. Kesimpulan untuk menjawab rumusan masalah berdasarkan hasil dari pembahasan Bab IV.

BAB II

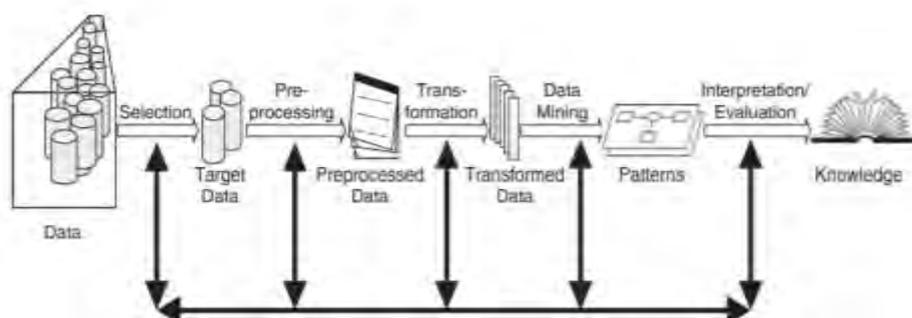
TINJAUAN PUSTAKA

2.1 *Data Mining*

Data mining adalah studi mengumpulkan, membersihkan, memproses, menganalisis, dan mendapatkan wawasan yang berguna dari data (Aggarwal, 2015). Menurut Gartner Group, *data mining* adalah proses menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah sejumlah besar data yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika (Larose, 2014). Sedangkan menurut Evangelous Simoudis bahwa *data mining* merupakan sebuah pendekatan lintas disiplin antar bidang-bidang pengetahuan yang menyumbangkan beberapa teknik untuk digunakan bersama-sama mulai dari *machine learning*, pengenalan pola, statistik, basis data dan visualisasi untuk mengatasi masalah ekstraksi informasi dari sebuah basis data yang besar (Larose, 2014).

Teknik *data mining* digunakan karena di dalam sebuah basis data yang besar, terdapat informasi-informasi ataupun pengetahuan-pengetahuan yang dapat berguna untuk membuat sebuah keputusan jika dilakukan pengolahan data dan analisa data tersebut. Manusia mampu melakukan analisa data jika data yang dilakukan analisa tidak cukup besar, namun jika data tersebut berjumlah ribuan, analisa secara manual yang dilakukan oleh manusia dengan menyisir satu-satu data dan menggabungkan antar data serta melakukan analisa akan membutuhkan waktu sehari-hari. Pemrosesan secara manual tersebut juga memungkinkan terjadi kesalahan yang lebih banyak dibandingkan dilakukan pemrosesan menggunakan mesin. Berbeda dengan cara manual, data mining tidak akan bermasalah jika dengan bertambahnya hari, data semakin tumbuh dan semakin bervariasi yang menambah kompleksitas data karena di dalam data mining terdapat *machine learning* yang mampu beradaptasi. Algoritma yang ada secara otomatis akan tetap menyesuaikan pertambahan data tersebut karena sudah terdefiniskan sebelumnya. Oleh karena itu *data mining* hadir untuk melakukan pemrosesan data sekaligus melakukan analisa data dan membuat sebuah kesimpulan-kesimpulan menggunakan algoritma baik

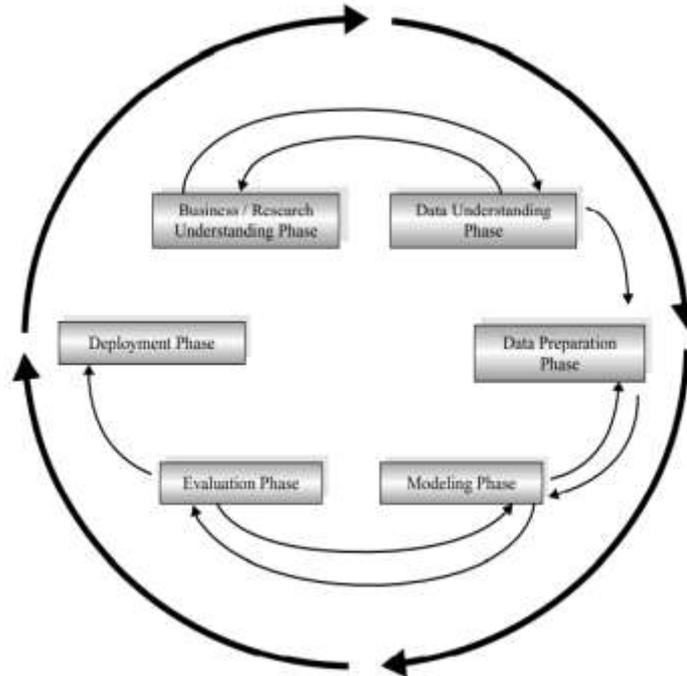
berdasarkan statistika, pengenalan pola atau metode lainnya. *Data mining* dapat digunakan dalam membantu analisa data sebuah organisasi baik untuk analisa di bidang pemasaran, *customer relationship management*, analisa keuangan dan juga yang berkaitan dengan perilaku manusia atau yang dalam sebuah perusahaan disebut manajemen SDM. *Data mining* digunakan untuk *customer relationship management* umum untuk menganalisis perilaku pelanggan untuk mengelola gesekan dan memaksimalkan nilai pelanggan yang diharapkan (Provost dan Fawcett, 2013). *Data mining* dan analisa prediktif sangat dibutuhkan untuk teknik pengambilan keputusan untuk manajemen SDM dalam sebuah organisasi (Mishra dan Lama, 2016). Teknik data mining dapat digunakan untuk menganalisis kumpulan data dan prediksi yang efektif untuk domain dan sub domain khususnya pada manajemen SDM (Changkakati dan Chayanika, 2020).



Gambar 2.1
Tahapan proses penemuan pengetahuan
Sumber : (Mariscal, et.al., 2010)

Data mining hanya merupakan salah satu proses dalam proses penemuan pengetahuan atau biasa disebut *Knowledge Discovery in Database (KDD)*. *Knowledge Discovery* telah didefinisikan sebagai ekstraksi yang diperlukan dan penting dari informasi implisit, yang sebelumnya tidak diketahui dan informasi yang berpotensi berguna dari data (Bramer, 2016). Secara umum, di dunia industri dan pers, *data mining* digunakan untuk merujuk pada keseluruhan proses KDD (Mariscal, et.al., 2010). Banyak orang memperlakukan *data mining* sebagai sinonim untuk istilah lain yang populer digunakan untuk KDD, sementara yang lain melihat *data mining* hanya sebagai langkah penting dalam proses penemuan pengetahuan (Han J, et.al,2012). Gambar 2.1 menunjukkan proses penemuan pengetahuan. Walaupun *data mining* dapat dibidang merupakan sentral dari

keseluruhan proses *Knowledge Discovery*, namun proses lainnya juga sangat penting untuk menjamin keberhasilan memperoleh penemuan pengetahuan. Proses lainnya tersebut seperti ditunjukkan pada Gambar 2.1, antara lain proses seleksi data dari sumber data, interpretasi data yang akan diolah dan *pre-processing data* serta proses transformasinya.



Gambar 2.2
Tahapan *data mining* berdasarkan CRISP-DM
Sumber : (Larose, 2014)

Pada tahun 1996, analisis yang mewakili DaimlerChrysler, SPSS, dan NCR mengembangkan proses standar dari *data mining* sebagai strategi pemecahan masalah baik untuk tujuan bisnis atau tujuan penelitian yang diberi nama *Cross-Industry Standard Process for Data Mining* (CRISP-DM) atau dalam bahasa Indonesia yang berarti Proses Standar Lintas Industri untuk *Data Mining*. Pada tahun 2000, Chapman menyebutkan bahwa penggunaan metodologi CRISP-DM merupakan pertimbangan secara standar *de facto* dalam dunia *data mining*. KdNuggets.com pada tahun 2007 menyatakan bahwa CRISP-DM adalah metodologi yang paling banyak digunakan untuk mengembangkan proyek *data mining* (Mariscal, et.al., 2010). CRISP-DM memiliki sebuah proses standar

yang terdiri dari 6 tahap proses yaitu Tahap Pemahaman Bisnis, Pemahaman Data, Penyajian Data, Pemodelan, Evaluasi dan *Deployment*.

2.1.1 Tahap Pemahaman Bisnis

Pada tahap ini pelaku penerapan *data mining*, baik perorangan, perusahaan atau organisasi harus memahami kebutuhan penggunaan *data mining*. Kebutuhan tersebut dapat diterjemahkan pada hal tujuan untuk memecahkan masalah, permasalahan apa yang harus dipecahkan, dan hasil yang diharapkan serta batasan masalah. Hal tersebut harus dilakukan di awal dengan jelas dan menyiapkan langkah-langkah strategis yang diperlukan untuk mencapai kebutuhan yang sudah didefinisikan. Larose (2014) menyebutkan dengan urutan langkah-langkah dalam tahap pemahaman bisnis antara lain pendefinisian dengan jelas tujuan proyek, bisnis atau penelitian *data mining* kemudian terjemahkan tujuan tersebut ke dalam batasan masalah dan perumusan masalah, yang terakhir adalah menyiapkan strategi untuk mencapai tujuan tersebut. Tahap pemahaman bisnis merupakan bagian dari keahlian di mana kreativitas analisis memainkan peran besar (Provost dan Fawcett, 2013). Tahapan ini merupakan tahapan yang penting sebelum melanjutkan ke tahapan selanjutnya karena akan mempengaruhi apa yang harus dilakukan dan bagaimana melakukan proses-proses selanjutnya sehingga tujuan yang diinginkan tercapai.

2.1.2 Tahap Pemahaman Data

Pengembang *data mining* harus menyiapkan data yang akan diolah, diproses dan dianalisa. Data dapat dikumpulkan lalu dikenali karakteristiknya dengan melakukan penyisiran dan pemilihan data yang terpakai sesuai tahap pemahaman bisnis. Proses penyisiran dan pemilihan data tersebut membuat kebutuhan pemahaman data yang akan diolah menjadi penting. Tahap pemahaman data yaitu sebuah proses untuk mengenali karakteristik data, memahami hubungan data dengan proses bisnis dan mengetahui data mana saja yang dibutuhkan untuk memenuhi tujuan proses penemuan pengetahuan melalui *data mining*. Penting untuk memahami kekuatan dan keterbatasan data karena

jarang ada kecocokan yang tepat antara data dengan masalah dalam bisnis yang akan dipecahkan (Provost dan Fawcett, 2013).

Pengolahan data dalam *data mining* dalam suatu permasalahan bisnis akan berbeda-beda tergantung dari masing-masing permasalahan bisnis dan tujuan pengolahan data tersebut. *Data mining* bukanlah sihir, tanpa pengawasan manusia yang terampil penggunaan perangkat lunak *data mining* secara buta hanya akan memberikan jawaban yang salah untuk pertanyaan yang salah yang diterapkan pada jenis data yang salah (Larose, 2014). Wang (2009), Bole dan kawan-kawan (2011) menyatakan bahwa keberhasilan *data mining* praktis sangat bergantung pada ketersediaan jumlah dan jenis data yang tepat (Strohmeier dan Piazza, 2013). Oleh karena itu menjadi penting proses identifikasi data yang dibutuhkan dengan mencari data-data yang paling relevan yang akan digunakan untuk penyelesaian masalah. Sumber-sumber data juga harus diidentifikasi untuk menyediakan data-data yang dibutuhkan. Tahapan pemahaman baik pemahaman bisnis dan pemahaman data membutuhkan analisa manusia sehingga dibutuhkan bekal pengetahuan dan ketrampilan yang berkesesuaian untuk menunjang keberhasilan penemuan pengetahuan dengan *data mining*.

2.1.3 Tahap Penyiapan Data

Tahap pemahaman data membantu pengembang untuk menyiapkan data-data dari sumber data karena telah diketahui data-data mana saja dan seperti apa saja yang akan diolah dan diproses di *data mining*. Seorang pengembang *data mining* pada tahap penyiapan data dihadapkan dengan sejumlah data yang besar dari dokumen mentah, basis data, log sistem, data transaksi dan data lainnya yang berkaitan dengan bisnis atau nilai yang akan diolah dalam *data mining*, dimana data-data mentah ini harus diubah menjadi sekumpulan data dalam format yang dapat diproses pada tahap selanjutnya. Contoh umum persiapan data adalah mengonversi data ke format tabel, menghapus atau menyimpulkan nilai yang hilang, dan mengonversi data ke tipe yang berbeda (Provost dan Fawcett, 2013). Tahap persiapan data adalah proses banyak tahap yang terdiri dari beberapa langkah yang berdiri sendiri, beberapa atau semuanya dapat digunakan dalam

aplikasi tertentu, langkah-langkah tersebut antara lain fitur ekstraksi dan portabilitas, *data cleaning*, dan yang terakhir tahap *data reduction, selection* dan *transformation* (Aggarwal, 2015).

1. Fitur ekstraksi dan portabilitas.

Pada beberapa aplikasi yang merupakan sumber data untuk *data mining*, ekstraksi data dapat dilakukan secara sederhana dari basis data, misalnya memakai metode akses standar seperti ODBC (Bramer M, 2016). Namun bagian terberat dalam proses penyiapan data ini adalah merubah data-data yang ada dari berbagai sumber menjadi bentuk data standar yang memungkinkan untuk dilakukan analisa. Data-data tersebut bisa dari sebuah *log* sistem, laporan yang di atas kertas, atau yang diekstrak dari transkrip wawancara dari narasumber. Data mentah yang terdiri dari beberapa sumber yang berbeda maka perlu dilakukan integrasi data sebelum dilakukan proses *mining*. Dalam beberapa kasus di mana data diperoleh dari berbagai sumber, perlu diintegrasikan ke dalam satu basis data untuk diproses (Aggarwal, 2015). Proses integrasi ini termasuk menyamakan *platform* yang digunakan untuk menyimpan data yang telah disatukan dari berbagai sumber tersebut. Proses ini disebut *data integration*.

Data integration menggabungkan data dari berbagai sumber ke dalam penyimpanan data yang koheren misalnya seperti *data warehouse* (Han J, et.al, 2012). Proses integrasi ini dapat mengurangi redudansi dan inkonsistensi *dataset*, hal tersebut dapat membantu memperbaiki akurasi dan kecepatan pada proses tahap pemodelan. Jika proses integrasi ini dilakukan dari beberapa sumber basis data yang bisa diakses secara bersamaan dengan satu aplikasi, maka metode *Extract Transform Load* (ETL) dapat dilakukan. Sistem ETL biasanya digunakan untuk mengintegrasikan data dari beberapa aplikasi, biasanya dikembangkan dan didukung oleh vendor yang berbeda atau di-*hosting* pada perangkat keras komputer yang terpisah (Thirumagal R, et.al., 2014).

2. *Data cleaning*

Data dalam dunia nyata cenderung tidak lengkap, ada *noise*, dan tidak konsisten, hal ini berkontribusi pada data yang tidak akurat (Han J, et.al, 2012). Oleh karena itu

data-data yang dikumpulkan haruslah data yang bersih agar dapat dilakukan pemodelan. Data yang tersedia untuk di-*mining* adalah data mentah, yaitu data asli yang didapat dalam format berbeda, berasal dari sumber yang berbeda, *noise* data bisa jadi tidak relevan, data harus diproses terlebih dahulu sebelum menerapkan teknik *data mining* (Jaffar, et.al., 2019). Contoh umum persiapan data adalah mengkonversi data ke format tabel, menghapus atau menyimpulkan nilai yang hilang, dan mengkonversi data ke tipe yang berbeda (Provost dan Fawcett, 2013). Data yang diekstraksi mungkin memiliki entri yang salah atau hilang, oleh karena itu beberapa catatan mungkin perlu dihapus atau entri yang hilang mungkin perlu diperkirakan dan inkonsistensi mungkin perlu dihilangkan (Aggarwal, 2015). Proses ini biasa disebut *cleaning*. *Data cleaning* dapat diterapkan untuk menghilangkan nilai yang hilang, *noise* dan memperbaiki inkonsistensi dalam data (Han J, et.al, 2012).

Aggarwal (2015) menyebutkan beberapa aspek dalam *data cleaning* antara lain mengatasi nilai yang hilang, mengatasi entri data yang tidak benar, satu lagi adalah pembuatan skala dan normalisasi. Proses memperbaiki nilai yang hilang, antara lain mengabaikan *tuple*, isi nilai yang hilang secara manual, gunakan konstanta global untuk mengisi semua nilai yang hilang seperti label "tidak diketahui" atau "-", atau ganti semua nilai yang hilang dengan ukuran seperti rata-rata atau nilai tengah. Pembetulan data yang tidak konsisten dan tidak benar dapat dilakukan dengan cara menemukan dahulu data yang tidak konsistensi menggunakan data deteksi, lalu menggunakan pengetahuan yang ada dilacak dari data yang tidak konsisten, mana data yang benar dan mana data yang salah. *Noise* adalah kesalahan acak atau varian yang berbeda dalam sebuah variable yang diukur (Han J, et.al, 2012). Setelah mengetahui data yang salah maka dilakukan isolasi data yang salah tersebut, kemudian menyamakan data yang salah tersebut dengan data yang paling terdekat. Metode tersebut disebut *Binning*. *Binning* bisa dengan melakukan perhitungan rata-rata sekitar data *binning* atau mengambil nilai tengahnya untuk dijadikan nilai baru pengganti data yang salah tersebut.

3. *Data reduction, selection dan transformation*

Proses *data mining* akan lebih baik jika data yang diproses membantu proses *data mining* menjadi lebih efisien dan mudah. Salah satu pemrosesan data yang akan membantu proses *data mining* adalah *data reduction*. Kumpulan data yang akan diproses mungkin berukuran besar, maka dimungkinkan untuk dikurangi ukurannya baik dari segi jumlah baris atau dimensinya (Aggarwal, 2015). *Data reduction* adalah upaya memperoleh representasi sekumpulan data yang telah dikurangi volume datanya menjadi jauh lebih kecil (Han J, et.al, 2012). *Data reduction* dapat dilakukan misalnya dari segi jumlah baris yang besar maka dapat dikurangi dengan menggunakan data *sampling*, sedangkan dari segi dimensi pengurangan data dapat dilakukan dengan cara mengurangi jumlah kolom data atau pemilihan *subset* fitur atau menggunakan *data transformation*. Ketika jumlah atribut menjadi besar, selalu ada risiko bahwa hasil yang diperoleh hanya akan memiliki akurasi yang kecil dan sebenarnya akan kurang dapat diandalkan dibanding jika hanya sebagian kecil dari atribut yang digunakan (Brammer M, 2016).

Data transformation (misalnya, normalisasi) dapat diterapkan, dimana data diskalakan agar berada dalam kisaran yang lebih kecil seperti 0,0 hingga 1,0, hal ini dapat meningkatkan efisiensi algoritma *data mining* (Han J, et.al, 2012). Dalam *data transformation*, data dirubah bentuk ke dalam nilai yang sesuai untuk di-*mining* misalnya data gaji karyawan setiap bulan dapat dibuat rata-rata untuk menentukan gaji dalam 1 tahun, proses tersebut biasa disebut *aggregation*. Sedangkan *discretization*, misalnya data umur penduduk dari umur 1 sampai umur 80 dirubah menjadi per kategori seperti bayi, balita, anak, remaja, dewasa, orang tua dan lansia. Konsep hierarki untuk nominal data juga dapat diterapkan, seperti alamat dapat dirubah ke hierarki data yang lebih tinggi. Data alamat yang lengkap menggunakan jalan, rt/rw, nomor rumah dapat disimpulkan ke hierarki data yang tinggi yaitu kota. Proses *data transformation* dan *data cleaning* bisa saja dilakukan bersamaan ketika proses menyisir dan menseleksi data.

2.1.4 Tahap Pemodelan

Memodelkan data adalah langkah utama dari *data mining*, dimana data yang ada dimodelkan sebagai bentuk prediksi-prediksi yang berdasarkan sebuah nilai-nilai yang terkandung di dalamnya. Rumus prediksi data tersebut bisa berupa matematis atau juga bisa pernyataan logis seperti sebuah aturan (Provost dan Fawcett, 2013).

Pada tahap pemodelan pengembang memilih teknik pemodelan berupa algoritma yang sesuai dan cocok, lalu digunakan pada data-data yang sudah disiapkan. Pada tahap ini, berbagai teknik pemodelan dipilih dan diterapkan, dan parameternya dikalibrasi ke nilai optimal (Mariscal, et.al., 2010). Teknik pemodelan yang digunakan tidak hanya satu, namun dapat beberapa teknik-teknik yang berbeda yang digunakan untuk memecahkan suatu masalah yang sama dan masing-masing teknik tersebut biasanya membutuhkan *requirement* tertentu dalam bentuk data yang lebih spesifik. Oleh karena itu, terkadang ketika sudah dalam melakukan proses pemodelan dan menemukan data yang kurang cocok untuk salah satu teknik pemodelan, maka dapat dilakukan langkah mundur untuk kembali ke tahap persiapan data khususnya data *pre-processing* untuk melakukan penyesuaian tipe data. Teknik pemodelan ditentukan juga berdasarkan tujuan, rumusan masalah dan karakteristik data.

2.1.5 Tahap Evaluasi

Sebelum melanjutkan ke tahap terakhir yaitu tahap *deployment*, penting untuk mengevaluasi hasil pemodelan secara lebih menyeluruh, dan meninjau langkah-langkah yang dijalankan untuk menyusun model, untuk memastikan model tersebut mencapai tujuan bisnis dengan benar (Mariscal, et.al., 2010). Tujuan dari tahap evaluasi adalah untuk menilai hasil *data mining* secara ketat dan untuk mendapatkan keyakinan bahwa mereka valid dan dapat diandalkan sebelum melanjutkan (Provost dan Fawcett, 2013). Hasil pengembangan perlu diperiksa apakah informasi yang dihasilkan dari pemrosesan model data yang ditemukan bertentangan dengan tujuan atau pemecahan permasalahan yang telah ditentukan pada tahapan pemahaman bisnis atau tidak. Jika tidak sesuai maka dapat dilakukan beberapa hal seperti memperbaiki data dalam tahapan pemahaman data atau

penyiapan data, memilih model atau teknik pengolahan data atau bisa juga menerima hasil ini dengan membuat kesimpulan yang mungkin dapat bermanfaat dalam perusahaan atau penelitian. CRISP-DM mempunyai standar proses, terlihat pada Gambar 2.2 bahwa terdapat proses umpan balik dari tahap evaluasi ke tahap model, ini terjadi ketika pengembang *data analitik* perlu melakukan pemodelan ulang setelah melihat hasil tingkat akurasi pada tahap evaluasi.

Confusion matrix dapat digunakan untuk mengevaluasi kualitas klasifikasi, misal untuk masalah dua kelas, ini menunjukkan positif benar, negatif benar, positif palsu, dan negatif palsu (Han J, et.al, 2012). Evaluasi pengukuran efektifitas klasifikasi dapat dilakukan dengan cara membandingkan model yang berbeda, memilih yang terbaik untuk sekumpulan data tertentu, penyetelan parameter, dan menggunakan meta-algoritma (Aggarwal, 2015). Terdapat tiga pendekatan dalam melakukan evaluasi proses klasifikasi antara lain *holdout*, *bootstrap*, dan *cross validation* (Aggarwal C, 2015).

2.1.6 Tahap *Deployment*

Tahap ini merupakan tahap akhir dimana hasil dari *data mining* dapat ditampilkan oleh pengguna yang membutuhkan atau *stakeholder* lainnya. Bahkan jika tujuan dari tahap pemodelan *data mining* tersebut adalah untuk meningkatkan pengetahuan tentang data, perlu untuk mengatur pengetahuan yang diekstraksi, serta menyajikannya dengan cara yang bermanfaat bagi pelanggan sehingga pelanggan bisa menggunakan pengetahuan tersebut (Mariscal, et.al., 2010). Penampilan hasil dari *data mining* bisa berupa langkah-langkah taktis atau visualisasi sebagai rekomendasi strategis pemangku kepentingan. Visualisasi bisa berupa grafik, tabel atau peta sesuai kebutuhan dan tujuan yang telah ditetapkan pada tahap pertama.

2.2 Tipe *Data Mining*

Pemilihan tipe *data mining* dilakukan saat tahap pemodelan menurut proses standar CRISP-DM. Penting untuk memiliki beberapa pemahaman tentang ide dasar *data mining*, termasuk jenis teknik dan algoritma yang ada karena hal itu merupakan bagian dari

keahlian dimana sebagian besar ilmu pengetahuan dan teknologi yang akan berperan dan digunakan (Provost dan Fawcett, 2013). Banyak algoritma *data mining* tidak bekerja dengan baik ketika mereka memiliki banyak atribut atau fitur. Oleh karena itu, sebelum dilakukan proses *data mining* perlu untuk memilih salah satu jenis algoritma *data mining* (Jaffar, et.al., 2019). *Data mining* terdiri dari sekumpulan teknik yang dapat digunakan untuk mengekstrak pengetahuan yang relevan dan menarik dari data yang memiliki beberapa tugas seperti penambangan aturan asosiasi, klasifikasi dan prediksi, serta *clustering* (Al-Radaideh dan Al Nagi, 2012). Sedangkan menurut Mishra dan Lama (2016), ada enam tugas yang tercakup dalam *data mining*; deteksi anomali, asosiasi, *clustering*, klasifikasi, regresi, dan *summarization*.

Tabel 2.1 Teknik-teknik yang dipakai dalam *data mining* bidang SDM

Metode	Bidang				N
	Performansi	Retensi	Rekrutasi dan seleksi	Lainnya	
Klasifikasi menggunakan C4.5, C5.0, ID3, <i>Random Forest tree</i>	Al-Radaidah dan Nagi, Azar Et Al, Gupta dan Sharma,	Alao Et Al, Chien dan Chen, Attri, Zhao Et Al,		Ancheta Et Al,	8
<i>Clustering</i>		Fan Et Al,			1
Neural networks		Zhao Et Al, Sexton Et Al, Tamizharasi	Jantan Et AL,		4
Teori Fuzzy	Yee dan Chen, Shah dan Ladhake,	Zhao Et Al,	Lad Et Al, Samalia Et Al, Shaout dan Yousif,	Kwak Et Al,	7
Lainnya	Vale Et Al,	Zhao Et Al, Khera dan Divya	Kalugina Et Al, Li Et Al,		5
N	6	11	6	2	25

Sumber : (Changkakati dan Chayanika, 2020)

Pada bidang SDM beberapa peneliti telah melakukan serangkaian penelitian penggunaan *data mining*. Seperti yang diungkapkan oleh Changkakati dan Chayanika (2020) yang menyatakan bahwa berbagai teknik dan metode *data mining* digunakan untuk menganalisis kumpulan data dan prediksi yang efektif untuk domain khusus SDM dan sub domain yang ada di dalamnya. Terlihat pada Tabel 2.1 bahwa sub bidang retensi pada

SDM diminati oleh peneliti untuk dilakukan penelitian dan metode *data mining* yang banyak digunakan adalah klasifikasi. Hal tersebut menunjukkan bahwa bidang retensi menarik dan perlu dilakukan penelitian. Sedangkan metode klasifikasi dengan algoritma-algoritmanya dianggap cocok untuk menyelesaikan masalah dalam menemukan pengetahuan pengelolaan retensi di bidang *data mining*.

2.2.1 Klasifikasi

Klasifikasi adalah proses menemukan model (fungsi) yang menggambarkan dan membedakan kelas atau konsep data (Han, Kamber dan Pie. 2012). Masalah klasifikasi adalah mempelajari struktur kumpulan contoh data, yang sudah dipartisi ke dalam kelompok, yang disebut sebagai kategori atau kelas (Aggarwal, 2015). Pengelompokan pada masalah klasifikasi ini misalnya bisa berupa sekelompok pembeli membeli produk tertentu atau dalam kampanye bisa terlihat sekelompok pemilih cenderung untuk memilih calon presiden yang mana. Saat ini banyak peneliti telah menggunakan tipe klasifikasi *data mining* pada bidang manajemen SDM seperti yang ada pada Tabel 2.1. Ancheta dan rekan penulis (2012) mencoba untuk memprediksi kinerja karyawan baru berdasarkan sekelompok aturan yang dirancang menggunakan algoritma klasifikasi berbasis aturan tertentu (Changkakati dan Chayanika, 2020). Gupta dan Sharma (2017) juga menggunakan tipe klasifikasi ini untuk mempelajari prediksi performansi karyawan bagian marketing memakai algoritma *decision tree* dan *naïve Bayesian*. Sedangkan Zhao dan kawan-kawan (2018) setahun setelah Gupta dan Sharma meneliti penggunaan salah satunya algoritma *decision tree* yang merupakan tipe klasifikasi untuk memprediksi *turn over* karyawan.

Proses klasifikasi terdiri dari beberapa tahap yaitu *training* dan evaluasi akurasi atau *testing*. Tahap pertama yaitu *training* dilakukan proses pengaksesan data menggunakan algoritma *data mining* untuk mencocokkan data dengan berbagai karakteristik data model yang telah didefinisikan sebelumnya. Sedangkan tahap evaluasi akurasi atau *testing*, dari *target* atribut yang telah diprediksi dilakukan evaluasi berdasarkan pemrosesan algoritma apakah sesuai atau tidak. Dengan klasifikasi, model yang dihasilkan akan dapat

memprediksi kelas untuk data yang diberikan tergantung pada informasi yang dipelajari sebelumnya dari data historis (Al Radaideh dan Al Nagi, 2012). Masalah klasifikasi lebih kuat daripada *clustering* karena menangkap gagasan pengelompokan yang ditentukan pengguna dari kumpulan data contoh (Aggarwal, 2015). Pengetahuan yang didapatkan oleh algoritma klasifikasi dapat diekspresikan dengan berbagai cara, seperti aturan-aturan, *decision tree*, *bayesian network* dan sebagainya (Jaffar, et.al., 2019).

2.2.2 Asosiasi

Asosiasi adalah teknik yang menjadi dasar teknik *data mining* lainnya yang bertujuan untuk menemukan aturan asosiasi item kriteria data satu dengan kriteria data lainnya. Masalah klasik *data mining* pola Asosiasi didefinisikan dalam konteks data supermarket yang berisi kumpulan barang yang dibeli oleh pelanggan, yang merujuk sebagai suatu transaksi (Aggarwal, 2015). Dari sebuah supermarket pengguna *data mining* dapat mengetahui pola transaksi pembeli ketika membeli barang, sehingga dapat membantu manajemen supermarket untuk mengelola dan meletakkan produk untuk meningkatkan penjualannya. Pembeli mempunyai kecenderungan pola transaksi misalnya suka membeli roti tawar, beserta mentega dan coklat meses. Kombinasi dan pengelompokan itu akan bisa membantu manajemen supermarket untuk meletakkan produk yang sekelas secara berdampingan sehingga memudahkan dan membuat nyaman pembeli sehingga meningkatkan penjualan. Asosiasi terdiri dari 2 tahap yaitu menemukan kombinasi sekumpulan item yang sering muncul dan kemudian mendefinisikan kondisi apa yang menyebabkan suatu sekumpulan item tersebut sering muncul dan hasil dari kemunculan sekumpulan item tersebut. Analisis Asosiasi adalah penelitian untuk menemukan korelasi implisit antar item, kejadian simultan atau pola yang sering terjadi (Sikaroudi dan Ghousi, 2015).

2.2.3 Clustering

Clustering berkaitan dengan pengelompokan bersama objek yang mirip satu sama lain dan tidak mirip dengan objek milik kelompok lain (Bramer, 2016). Beberapa contoh

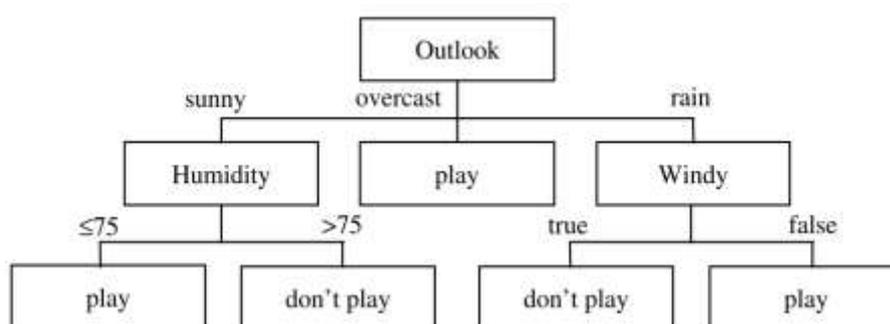
penerapan *clustering* seperti pengelompokan ciri-ciri pembeli dalam supermarket yang memiliki perilaku beli suatu produk, pengelompokan pasien di rumah sakit yang memiliki gejala yang sama, atau pengelompokan perusahaan yang mempunyai performansi keuangan yang sama. Dalam bidang manajemen SDM ditemukan penelitian tentang *clustering* seperti yang tertera pada Tabel 2.1, Fan dan rekannya (2012) mengilustrasikan penggunaan *data mining* menggunakan *clustering* dengan *machine learning* untuk memprediksi *turn over* tenaga profesional (Changkakati dan Chayanika, 2020). Teknik *clustering* membagi sejumlah besar data menjadi sejumlah kecil kelompok sangat membantu untuk meringkas data kemudian memahaminya dalam penggunaan aplikasi *data mining*. *Clustering* adalah metode *unsupervised*, tidak memiliki kelas atau fitur target dan semua fitur data dilihat sebagai jarak, bukan kategori (Sikaroudi dan Ghousi, 2015).

2.3 Algoritma *Decision Tree*

Algoritma *Decision Tree* (Pohon Keputusan) adalah salah satu teknik yang paling banyak digunakan, karena membuat pohon keputusan dari data yang diberikan dengan menggunakan persamaan sederhana terutama tergantung pada perhitungan rasio keuntungan, yang secara otomatis memberikan semacam bobot pada atribut yang digunakan, dan peneliti secara implisit dapat mengenali atribut yang paling efektif pada target yang diprediksi (Al Radaideh dan Al Nagi, 2012). Pohon Keputusan adalah metodologi klasifikasi, di mana proses klasifikasi dimodelkan dengan menggunakan satu set keputusan hierarkis pada variabel fitur, diatur dalam struktur seperti pohon (Aggarwal, 2015). Penggunaan Algoritma pohon keputusan harus memenuhi 3 syarat, yaitu memerlukan variabel target yang telah diklasifikasi sebelumnya karena pohon keputusan termasuk metode *supervised*, karena belajar dari contoh sebelumnya maka data pelatihan yang digunakan harus kaya dan beragam, dan yang terakhir kelas atribut dari target harus diskrit.

Pohon keputusan merupakan salah satu metode klasifikasi yang menarik, dilakukan pembentukan kumpulan simpul keputusan yang dihubungkan oleh cabang-cabang, meluas ke bawah dari simpul akar hingga berakhir di simpul daun. Dimulai dari simpul akar,

yang menurut konvensi ditempatkan di bagian atas diagram pohon keputusan, atribut diuji pada simpul keputusan, dengan setiap kemungkinan hasil yang dihasilkan di cabang. Setiap cabang kemudian mengarah ke simpul keputusan dalam struktur pohon yaitu daun yang terakhir. Pohon keputusan adalah struktur pohon seperti diagram alur, di mana setiap simpul menunjukkan pengujian pada nilai atribut, setiap cabang mewakili hasil pengujian, dan daun pohon mewakili kelas atau distribusi kelas (Han et.al., 2012). Gambar 2.3 merupakan contoh pohon keputusan untuk permainan golf, dimana model tersebut digunakan untuk menentukan apakah bermain golf atau tidak berdasarkan kondisi cuaca.



Gambar 2.3
Contoh pohon keputusan untuk permainan Golf
Sumber : (Bramer, 2016)

Kriteria penghentian untuk pertumbuhan pohon keputusan terkait erat dengan strategi pemangkasan yang mendasarinya (Aggarwal C, 2015). Aturan pemangkasan itu yang biasa disebut *pruning*. Pemangkasan dilakukan ketika setiap simpul daun hanya berisi satu kelas, jika tidak dilakukan pemangkasan maka hasilnya tanpa simpul dan tidak terdapat cabang keputusan. Metode *pruning* dijalankan untuk meningkatkan akurasi, ketika akurasi klasifikasi meningkat jika dilakukan pemangkasan maka simpul berhenti dan harus dilakukan pemangkasan. Ketika pohon keputusan dibangun, beberapa cabang yang dibentuk mungkin saja mempunyai sifat yang tidak relevan dalam menentukan kelas data atau memiliki *noise* cabang dalam data *training*. Cabang-cabang yang membingungkan dan tak menentukan itulah yang perlu dipangkas agar tingkat akurasi meningkat. Bramer M (2016) menyebutkan bahwa terdapat 2 *pruning* yaitu pertama disebut *prepruning*, dimana sebuah aturan rencana pembangunan pohon keputusan dengan jumlah cabang lebih sedikit daripada yang akan terjadi. Jenis *pruning* kedua adalah *post pruning* (biasa

disebut *pruning* saja), berbeda dengan *prepruning* metode ini menghapus bagian pohon yang telah dihasilkan dengan harapan memiliki pohon lebih pendek dan lebih sederhana. Keduanya bertujuan untuk meningkatkan akurasi dari prediksi data dari model klasifikasi yang dibuat. *Pruning* mencoba untuk mengidentifikasi dan menghapus beberapa cabang, dengan tujuan meningkatkan akurasi klasifikasi pada data yang tidak terlihat (Han et.al., 2012).

Persyaratan tertentu harus dipenuhi sebelum algoritma pohon keputusan dapat diterapkan:

1. Algoritma pohon keputusan mewakili pembelajaran yang diawasi, dan dengan demikian membutuhkan variabel target yang telah diklasifikasikan sebelumnya. Kumpulan data pelatihan harus disediakan yang memberikan algoritma dengan nilai-nilai variabel target.
2. Kumpulan data pelatihan ini harus kaya dan beragam, menyediakan algoritma dengan penampang yang sehat dari jenis rekaman klasifikasi yang mungkin diperlukan di masa mendatang. Pohon keputusan belajar dengan contoh, dan jika contoh secara sistematis tidak ada untuk *subset* rekaman yang dapat ditentukan, klasifikasi dan prediksi untuk *subset* ini akan bermasalah atau tidak mungkin terjadi.
3. Kelas atribut target harus diskrit. Artinya, seseorang tidak dapat menerapkan analisis pohon keputusan ke variabel target kontinu. Sebaliknya, variabel target harus menggunakan nilai yang dengan jelas dibatasi sebagai milik kelas tertentu atau bukan milik kelas.

Pohon keputusan memiliki keunggulan interpretasi dan pemahaman yang mudah bagi pengambil keputusan untuk dibandingkan dengan pengetahuan domain mereka untuk validasi dan pembenaran keputusan mereka (Al Radaideh dan Al Nagi, 2012). Klasifikasi menggunakan pohon keputusan merupakan teknik yang cukup populer, hal ini disebabkan konstruksi pohon tidak memerlukan pengetahuan seorang pakar atau pengaturan parameter, dan cocok untuk penemuan pengetahuan yang sifatnya eksplorasi data. Pohon keputusan dapat menghasilkan model dengan aturan yang dapat dibaca dan mudah

diinterpretasikan oleh manusia biasa. Contoh algoritma pohon keputusan antara lain ID3, C4.5/J4.8, CART, Naïve Bayes Tree dan lainnya.

2.4 Algoritma C4.5

Algoritma C4.5 atau J48 adalah salah satu algoritma pohon keputusan untuk permasalahan klasifikasi dalam *data mining* yang paling populer dan kuat serta memiliki akurasi cukup baik. Sangat penting untuk membandingkan algoritma klasifikasi ketika akan mengimplementasikan *data mining* sehingga dapat memilih algoritma yang mempunyai akurasi paling baik agar bantuan yang diberikan oleh data analitik SDM ini dapat menunjang pengelolaan SDM yang lebih baik dan mengatasi retensi karyawan dengan baik.

Tabel 2.2 menjelaskan berdasarkan hasil riset Z Jaffar, N Waheed, Z Kanwal di tahun 2019 Algoritma J48 mempunyai validitas dan ketepatan yang lebih baik, untuk itu desain *data mining* ini akan memilih algoritma J48 sebagai algoritma yang akan digunakan di tahap pemodelan. J48 adalah algoritma turunan dari C4.5 yang berjalan di *platform* Java. Fitur algoritma J48 ditemukan jika aplikasi yang digunakan WEKA. Penelitian ini menggunakan Rapidminer, dimana hanya ada algoritma C4.5 yang alur algoritmanya sama dengan J48.

Tabel 2.2 Performansi algoritma Klasifikasi analisis data SDM

Classifier Algorithm	Training Set (%)	Testing Set (%)
J48	98.84	97.67
Naïve Bayes	85.35	84.53
Bayes Net	93.83	93.18
OneR	89.35	89.45
Logistic Regressions	79.75	80.21

Sumber : (Jaffar, et.al., 2019)

Teknik C4.5 adalah salah satu algoritma pohon keputusan yang dapat menghasilkan baik pohon keputusan maupun kumpulan aturan; dan membangun pohon untuk tujuan meningkatkan akurasi prediksi (Al Radaideh dan Al Nagi, 2012). Menurut Kusriani & Luthfi

secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam *data mining* adalah (Dhika & Destiwati, 2015):

1. Pilih atribut sebagai simpul akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang
4. Ulang proses untuk setiap cabang sampai seluruh kasus-kasus telah memiliki kelas yang sama.

C4.5 menghasilkan pohon dengan bentuk yang lebih bervariasi. Algoritma C4.5 menggunakan konsep perolehan informasi atau pengurangan entropi untuk memilih pembagian yang optimal. Misalkan ketika memiliki variabel X yang nilai k yang memungkinkan memiliki probabilitas p_1, p_2, \dots, p_k . Berapa jumlah bit terkecil, rata-rata per simbol, yang diperlukan untuk mengirimkan aliran simbol yang mewakili nilai X maka yang paling optimal diprioritaskan. Pada algoritma C4.5, semua aturan dalam sebuah kelas yang kumpulan aturannya memiliki panjang deskripsi terkecil diprioritaskan daripada aturan lainnya (Aggarwal, 2015). Simpul yang direlasikan dengan atribut, baik simpul akar (*root*) atau simpul internal, pemilihannya didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Penghitungan nilai *Gain* menggunakan rumus persamaan :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Rumus 2.1 Menghitung nilai Gain

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i | : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Sebelum melakukan perhitungan *Gain*, pembentukan model pohon keputusan perlu menghitung nilai *Entropy* terlebih dahulu. Tingkat keragaman setiap atribut pada pohon

keputusan yang berperan dalam penentuan hasil keputusan ditentukan oleh parameter *Entropy* ini. Penghitungan tinggi rendahnya nilai *Entropy* dapat dilihat pada persamaan :

$$Entropy(s) = \sum_{i=1}^n -p_i * \log_2(p_i)$$

Rumus 2.2 Menghitung nilai *Entropy*

n : Jumlah partisi S

pi : Proporsi dari Si terhadap S

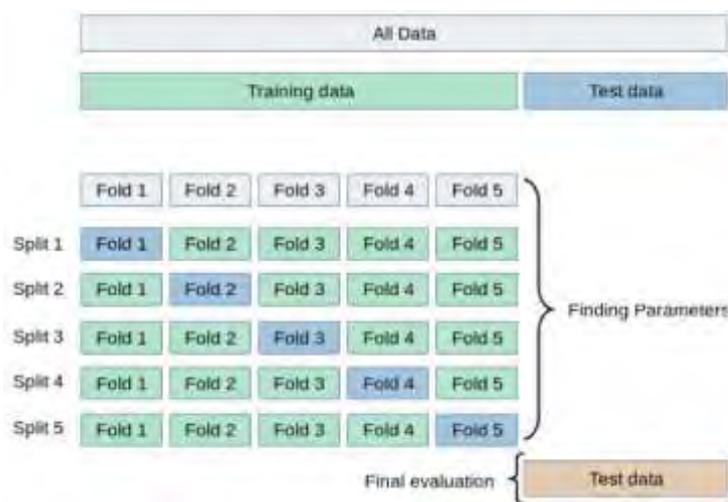
C4.5 dapat mengatasi atribut diskrit dan kontinu (Singh dan Gupta, 2014). Hal ini merupakan kelebihan dari C4.5 dibandingkan ID3 dan CART. Ketiga algoritma tersebut mampu mengatasi atribut kategorikal, namun untuk hanya C4.5 yang mampu mengatasi dengan baik untuk atribut diskrit dan kontinu. Singh dan Gupta (2014) dalam penelitiannya juga mengatakan bahwa C4.5 mampu mengatasi *missing values* dan *data noise*.

2.5 Cross Validation

Cross validation adalah metode statistik untuk mengevaluasi dan atau membandingkan algoritma pembelajaran dengan membagi data menjadi dua bagian: satu digunakan untuk mempelajari atau melatih model dan yang lainnya digunakan untuk memvalidasi model (El-Habi dan El-Ghareeb, 2014). Pengukur tingkat akurasi prediksi dari proses klasifikasi biasanya harus memisahkan data menjadi dua bagian antara lain data *training* dan data *testing*. Data *training* digunakan untuk membentuk klasifikasi, sedangkan data *testing* digunakan untuk melakukan prediksi menggunakan klasifikasi yang telah dibentuk sehingga dapat dilakukan pengukuran tingkat akurasi prediksi. Terdapat tiga strategi umum yang digunakan untuk melakukan hal tersebut yaitu membagi data secara manual menjadi data *training* dan data *testing*, *k-fold cross-validation*, dan *N-fold cross-validation* (Bramer M, 2016). Penggunaan data *training* dan data *testing* yang sama biasanya akan menimbulkan masalah, untuk itulah *cross validation* muncul menjadi solusi.

Karami dan kawan-kawan (2012) menyebut bahwa ciri *cross validation*, membagi data *training* dan data *validasi* secara menyilang kemudian digilir secara berputar secara

berurutan sehingga setiap data memiliki peluang untuk divalidasi (El-Habi dan El-Ghareeb, 2014). Aturan *cross validation* membagi data *training* menjadi beberapa bagian kecil yang disebut *k-fold*, k ini merupakan jumlah bagian kecil tersebut. Gambaran pembagian tersebut ditunjukkan pada Gambar 2.4. Prosedur pada setiap *k-fold* tersebut adalah pertama kali, model yang dibuat akan dilatih menggunakan data sejumlah k-1, lalu model yang dihasilkan akan divalidasi menggunakan sebagian data yang disebut sebagai data *testing* untuk menghitung tingkat akurasi. Perhitungan rata-rata dari sejumlah pengulangan merupakan nilai yang digunakan untuk mengukur performansi.



Gambar 2.4
Proses pembagian data pada *Cross Validation*
Sumber : (scikit-learn.org)

Penggunaan *cross validation* dalam Rapidminer, menggunakan *operator cross validation* yang merupakan *nested operator*. Disebut *nested operator* karena di dalamnya terdiri dari dua sub proses, yaitu proses *training* dan *testing*. Pada *testing* proses ini dilakukan pengukuran performansi model. *Cross validation* akan melakukan proses sebanyak k kali, dengan setiap subset k dipastikan digunakan sekali sebagai data tes. Nilai k ini dapat diatur yang merupakan jumlah kelipatan yang diinginkan, hasil akhir akurasi performansi didapatkan dari hasil iterasi k yang dibuat rata-rata.

2.6 Confusion Matrix

Tahap evaluasi dan pengukuran kinerja pada klasifikasi dapat menggunakan *confusion matrix* untuk menghitung akurasi proses pemodelan. *Confusion matrix* adalah

alat yang berguna untuk menganalisis seberapa baik klasifikasi yang dilakukan dengan mengenali tupel dari kelas yang berbeda (Han J, et.al., 2012). Cara tabel menyajikan informasi kinerja pengklasifikasi yang disebut *confusion matrix* diperkenalkan, bersama-sama dengan gagasan klasifikasi positif dan negatif benar dan salah (Bramer M, 2016). *Confusion matrix* biasa digambarkan sebagai bentuk 4 kombinasi yang masing-masing kombinasi tersebut mempunyai arti dari nilai prediksi dan nilai aktual yaitu *true positive*, *true negative*, *false positive* dan *false negative*.

Tabel 2.3 *True dan false positives dan negatives*

<i>Correct Classification</i>	<i>Classified as</i>	
	+	-
+	<i>true positives</i>	<i>false negatives</i>
-	<i>false positives</i>	<i>true negatives</i>

Sumber: (Bramer M, 2016)

Confusion matrix cocok untuk proses klasifikasi yang mempunyai klasifikasi 2 label. Tabel 2.3 menjelaskan 4 istilah untuk mengukur proses evaluasi klasifikasi yang dapat dijelaskan sebagai berikut :

1. *True positives* (TP), ini berarti prediksi positif dan hasil sesuai
2. *True negatives* (TN), mempunyai arti prediksi negatif dan hasil klasifikasi sama negatif.
3. *False positives* (FP), berarti prediksi positif namun ternyata hasil klasifikasi bukan positif, artinya salah prediksi.
4. *False negatives* (FN), mempunyai arti salah prediksi dimana prediksi negatif tetapi hasil dari klasifikasi tidak sesuai dengan prediksi.

Pada penjelasan 4 kombinasi tersebut bisa disimpulkan secara sederhana bahwa terdapat 2 hasil dari klasifikasi yaitu TP dan TN merupakan proses klasifikasi benar dan sesuai sedangkan hasil FP dan FN menginformasikan bahwa hasil dari proses klasifikasi tidak sesuai atau terjadi kesalahan. Selain TP, TN, FP dan FN, terdapat simbol P dan N dalam melakukan pengukuran evaluasi klasifikasi. P merupakan simbol jumlah seluruh hasil aktual yang bernilai positif sedangkan N adalah simbol jumlah seluruh hasil aktual yang bernilai negatif. Kemudian juga terdapat simbol P' dan N' yang berarti bahwa jika P' merupakan total jumlah prediksi yang bernilai positif dan N' merupakan total jumlah prediksi yang bernilai negatif. Dengan model tersebut kita bisa mengukur tingkat akurasi, kesalahan

klasifikasi atau tingkat *error*, tingkat sensitifitas dan tingkat spesifik serta tingkat presisi dari hasil klasifikasi. Rumus-rumus dari pengukuran tersebut dapat dilihat pada Tabel 2.4.

Tabel 2.4 Rumus pengukuran hasil proses klasifikasi

Pengukuran Klasifikasi	Rumus
Tingkat akurasi, <i>recognition rate</i>	$\frac{TP + TN}{P + N}$
Tingkat kesalahan, <i>misclassification rate</i>	$\frac{FP + FN}{P + N}$
Tingkat sensitivitas, <i>true positive rate, recall</i>	$\frac{TP}{P}$
Tingkat spesifik, <i>true negative rate</i>	$\frac{TN}{N}$
Presisi	$\frac{TP}{TP + FP}$

Sumber: (Han J, et.al., 2012)

2.7 Manajemen Sumber Daya Manusia

Manajemen sumber daya manusia adalah proses perencanaan, pengorganisasian, pengarahan, dan pengendalian atas pengadaan tenaga kerja, pengembangan, kompensasi, integrasi, pemeliharaan, dan pemutusan hubungan kerja dengan sumber daya manusia untuk mencapai sasaran perorangan, organisasi, dan masyarakat (Almassri, 2016). Manajemen sumber daya manusia mengacu pada kebijakan dan praktik yang terlibat dalam melaksanakan aspek sumber daya manusia dari posisi manajemen termasuk perencanaan sumber daya manusia, analisis pekerjaan, rekrutasi, seleksi, orientasi, kompensasi, penilaian kinerja, pelatihan dan pembangunan, dan hubungan kerja (Jerome, 2013). Selain itu, manajemen sumber daya manusia dapat dilihat sebagai pendekatan khusus untuk pekerjaan manajemen yang berupaya untuk mencapai keunggulan kompetitif melalui penyebaran strategi yang berkomitmen tinggi dan tenaga kerja yang mampu. Armstrong mengusulkan manajemen sumber daya manusia sebagai sesuatu yang strategis dan rasional pendekatan untuk mengelola aset organisasi yang paling berharga, karyawan yang berkontribusi pada pencapaian tujuan dan sasaran mereka (Jerome, 2013).

Praktik manajemen sumber daya manusia menawarkan kontribusi positif terhadap kepuasan kerja sehingga kepercayaan di tempat kerja, komitmen, keterikatan, kinerja akan meningkat, oleh karena itu organisasi akan mencapai tujuannya (Mira, et.al., 2019). Sedangkan fungsi dari manajemen sumber daya manusia adalah untuk memotivasi karyawan dan meningkatkan efektivitas tenaga kerja (Sikaroudi dan Ghousi, 2015). Fungsi-fungsi manajemen sumber daya manusia menurut Sutrisno di tahun 2009 antara lain (Nurhidayati, 2016):

1. Perencanaan, merupakan kegiatan perhitungan kebutuhan tenaga kerja di masa akan datang sesuai keperluan perusahaan untuk mencapai tujuan secara efektif dan efisien.
2. Pengorganisasian, yaitu pembagian tenaga kerja beserta tugas dan tanggung jawab serta hubungan dan koordinasi antar karyawan atau bagian dalam sebuah organisasi.
3. Pengadaan dan Pengarahan, yaitu proses penyediaan tenaga kerja mulai dari pembukaan lowongan kerja, penyeleksian dan penerimaan dalam bentuk kontrak kerja serta penempatannya. Setelah dilakukan pengadaan maka perlu adanya pengarahan yang merupakan pemberian petunjuk, cara atau maksud maupun tujuan organisasi agar tenaga kerja dapat saling bekerja sama dan bekerja dengan baik.
4. Pengendalian, proses mengarahkan dan menggerakkan tenaga kerja dalam mengikuti arahan dari pemimpin, rencana perusahaan, dan menaati peraturan-peraturan yang telah ditetapkan untuk mencapai tujuan perusahaan.
5. Pengembangan, kegiatan yang bertujuan untuk meningkatkan produktifitas dan performansi tenaga kerja dengan cara mengadakan pelatihan, seminar, atau pendidikan baik secara teoritis ataupun praktek.
6. Kompensasi merupakan pemberian hak kepada tenaga kerja sebagai imbal jasa yang diberikan oleh perusahaan sebagai kewajiban perusahaan yang tertuang di peraturan pemerintah tentang ketenagakerjaan.
7. Pengintegrasian, kegiatan yang bertujuan untuk menyatukan kepentingan perusahaan dan kebutuhan tenaga kerja dalam proses pemenuhan kontrak kerja untuk memenuhi visi misi perusahaan agar tercipta situasi yang kondusif dan solid dalam lingkungan kerja.

8. Pemeliharaan merupakan proses pengelolaan tenaga kerja dalam suatu perusahaan agar tenaga kerja tersebut dapat bekerja secara optimal sehingga dapat membantu pencapaian performansi perusahaan yang baik, loyal dan mampu mempertahankan performansinya sampai masa pensiun baik secara fisik, mental dan skill serta pengetahuan yang dimiliki. Komitmen tenaga kerja terhadap perusahaan dan keinginan untuk memberikan kontribusi positif perlu dijaga oleh perusahaan sehingga berpengaruh terhadap kesinambungan proses bisnis perusahaan, biasanya juga bisa disebut pengelolaan retensi.
9. Kedisiplinan, fungsi ini berkaitan dengan upaya untuk membentuk seluruh entitas dalam sebuah perusahaan untuk mengikuti peraturan yang berlaku seperti tenaga kerja terhadap peraturan perusahaan, peraturan pemerintah atau undang-undang ketenagakerjaan.
10. Pemberhentian, fungsi terakhir dari manajemen sumber daya manusia adalah proses pemutusan hubungan kerja antara perusahaan dan tenaga kerja, baik pada masa tengah proses kerja ataupun karena datangnya masa pensiun.

Mustiko (2012) menyimpulkan pada penelitiannya berdasarkan beberapa referensi bahwa jika dikelompokkan terdapat 2 fungsi dari manajemen sumber daya manusia yaitu fungsi manajemen dan fungsi operasional. Fungsi manajemen terdiri dari perencanaan, pengorganisasian, pengarahan dan pengendalian. Sedangkan untuk fungsi operasional terdiri dari perekrutan dan seleksi, pelatihan dan pengembangan, kompensasi balas jasa dan jaminan sosial, pengembangan karir, integritas, pemeliharaan tenaga kerja, dan pemutusan hubungan kerja. Pengelolaan retensi karyawan baik menurut Sutrisno (2009) dalam penelitian Nurhidayati (2016) maupun menurut Mustiko (2012), pengelolaan retensi karyawan merupakan salah satu fungsi dari manajemen sumber daya manusia yang dapat dimasukkan dalam fungsi bagian pemeliharaan tenaga kerja.

2.8 Retensi Karyawan

Retensi karyawan adalah suatu cara dan upaya perusahaan untuk mempertahankan sumber daya manusia yang potensial dalam perusahaan tersebut agar dapat mencapai

tujuannya (Sa'diyah, et.al., 2017). Al Jarradi menyebutkan bahwa retensi adalah kemampuan perusahaan untuk menjaga nilai karyawan yang berkontribusi terhadap keberhasilan organisasi selama hubungan tersebut saling menguntungkan. Sedangkan Okioaga menyatakan retensi karyawan merupakan salah satu elemen yang mempengaruhi mobilitas kerja karyawan secara umum dalam suatu organisasi, juga mempunyai arti pergantian karyawan yang dapat dianggap sebagai sisi lain yang sama dari mata uang. Retensi karyawan saat ini merupakan salah satu masalah penting dalam organisasi (Lahida, et.al., 2017). Retensi karyawan merupakan salah satu fungsi manajemen sumber daya manusia yaitu termasuk dalam fungsi pemeliharaan.

Tabel 2.5 Faktor retensi karyawan

Dimensi	Indikator
Komponen Organisasional	<ol style="list-style-type: none"> 1. Budaya dan nilai organisasi 2. Perubahan strategi, peluang dan manajemen organisasi 3. Kontinuitas dan keamanan kerja
Peluang Karier	<ol style="list-style-type: none"> 1. Adanya kontinuitas pelatihan dan pengembangan diri 2. Kesempatan mendapatkan pengembangan diri dan bimbingan 3. Adanya perencanaan karier
Penghargaan	<ol style="list-style-type: none"> 1. Gaji dan tunjangan kompetitif berdasarkan kapabilitas, pengalaman dan kinerja 2. Adanya perbedaan penghargaan berdasarkan kinerja 3. Adanya pengakuan dari <i>supervisor</i> atas kinerja 4. Pemberian tunjangan dan bonus spesial
Rancangan dan Tugas Pekerjaan	<ol style="list-style-type: none"> 1. Tanggung jawab dan otonomi kerja jelas 2. Adanya fleksibilitas jam dan lokasi kerja 3. Adanya keseimbangan antara pekerjaan dan kehidupan
Hubungan Kerja	<ol style="list-style-type: none"> 1. Perlakuan yang adil dan tidak diskriminatif 2. Adanya dukungan dari manajer 3. Hubungan Kerja yang baik

Sumber : (Marliani dan Sakapurnama. 2014)

Marliani dan Sakapurnama (2014) menyebutkan dalam penelitiannya bahwa Mathis dan Jackson di tahun 2006 menjelaskan retensi merupakan usaha untuk mempertahankan karyawan agar tetap berada dalam organisasi guna mencapai tujuannya. Mathis dan Jackson (2006) menyebutkan faktor-faktor yang menyebabkan karyawan bertahan di perusahaan, faktor-faktor tersebut diolah oleh Marliani dan Sakapurna (2014) menjadi seperti pada Tabel 2.5, yaitu komponen organisasional, peluang karier, penghargaan, rancangan tugas dan pekerjaan serta hubungan karyawan. Sedangkan Adzka (2017), melakukan penelitian-penelitian terdahulu dan menyimpulkan bahwa ada 14 faktor yang

mempengaruhi retensi karyawan antara lain kompensasi, pelatihan dan pengembangan, persepsi dukungan organisasi, pengembangan karier, penghargaan, kepuasan kerja, lingkungan kerja, kepemimpinan, faktor demografi, stres kerja, keamanan kerja, budaya organisasi, keseimbangan kehidupan kerja, serta desain pekerjaan.

Strategi dan praktik motivasi dapat dilihat sebagai nutrisi tambahan yang diperlukan untuk meningkatkan keterlibatan, kepuasan, komitmen, dan kinerja karyawan di tempat kerja (Vito et.al., 2016). Pimpinan perusahaan atau manajer memiliki tanggung jawab untuk menyediakan lingkungan kerja dan perangkat kerja yang dapat memotivasi semua karyawannya karena hal tersebut adalah kebutuhan hidup di lingkungan kerja yang harus dipenuhi yang semakin kompetitif di era persaingan seperti ini. Pimpinan perusahaan harus memperhatikan faktor-faktor yang mempengaruhi motivasi karyawan misalnya seperti yang disampaikan oleh Heidarian, Kelarijani dan timnya pada tahun 2015 dari hasil penelitiannya menyatakan bahwa variabel penting yang mempengaruhi faktor motivasi adalah gelar akademik, status pekerjaan, status perkawinan, jenis kelamin, usia dan masa kerja. Hoppock (1935) mendefinisikan kepuasan kerja sebagai setiap kombinasi dari keadaan psikologis, fisiologis, dan lingkungan yang menyebabkan seseorang dengan jujur mengatakan, '*Saya puas dengan pekerjaan saya*'. Ketika mempertimbangkan kepuasan kerja, variabel demografis harus dipertimbangkan untuk memahami secara menyeluruh faktor-faktor yang mungkin menyebabkan kepuasan dan ketidakpuasan kerja (Scott, et.al., 2005). Faktor-faktor yang berpengaruh dalam retensi karyawan perlu diperhatikan dan menyediakan lingkungan kerja serta perangkat kerja yang dapat memotivasi, ketika pimpinan perusahaan gagal memperhatikan hal tersebut maka bisa jadi karyawan tidak akan berkomitmen pada perusahaan dan meninggalkan perusahaan tersebut untuk berpindah ke pesaingnya. Literatur penelitian menunjukkan bahwa kepuasan kerja, seperti motivasi karyawan, dipengaruhi oleh karakteristik pribadi karyawan, seperti usia, jenis kelamin, masa kerja, tingkat pendidikan dan senioritas, dengan sejumlah penelitian menunjukkan hubungan tersebut dalam beberapa tahun terakhir (Roos W, 2005).

2.9 Teori Hierarki Maslow

Teori motivasi manusia dari Maslow (1943) adalah teori pertama yang menganggap kebutuhan akan pertumbuhan, pemenuhan, dan perkembangan sebagai hal utama bagi karyawan dalam hal ini sisi psikologi mereka, oleh karena itu, dia dihargai sebagai ayah dari psikologi kemanusiaan (Mira, et.al., 2019). Teori Maslow digunakan sebagai alat khususnya dalam aktivitas terkait SDM di berbagai perusahaan untuk pengelolaannya (Lahida, et.al., 2017). Lahida dan kawan-kawan (2017) juga menyebutkan bahwa retensi karyawan sangat penting untuk kebaikan proses bisnis jangka panjang sebuah perusahaan. Oleh karena itu, Pengelolaan retensi karyawan dapat dilakukan dengan implementasi teori hierarki kebutuhan Maslow yang meliputi kompensasi, pengkondisian lingkungan dan prosedur dalam sebuah perusahaan. Teori hierarki kebutuhan Abraham Maslow masih penting dan relevan dalam organisasi bisnis saat ini (Jerome, 2013). Jika mencermati hasil penelitian-penelitian mengenai retensi karyawan yang telah diulas pada sub bab 2.8, terlihat faktor-faktor yang mempengaruhi retensi karyawan berkorelasi dengan teori hierarki kebutuhan Maslow.



Gambar 2.5
Teori hierarki kebutuhan Maslow
Sumber : (Jerome, 2013)

Maslow mengklasifikasikan kebutuhan manusia ke dalam struktur piramida yang terdiri dari lima tingkat, dari tingkat paling bawah ke tingkat paling atas antara lain kebutuhan fisiologis, rasa aman dan ketenangan, hubungan sosial dan cinta, penghargaan

dan atau pengakuan serta aktualisasi diri yang merupakan tingkat paling atas. Perusahaan harus memperhatikan implementasi teori tersebut yang meliputi pemenuhan kebutuhan fisiologis, kebutuhan akan rasa aman dan ketenangan, hubungan sosial dan rasa cinta yang baik antar sesama karyawan, adanya sistem penghargaan dan pengakuan kepada karyawan serta mekanisme aktualisasi diri. Hierarki kebutuhan Maslow dibangun untuk menunjukkan motivasi manusia pada umumnya, namun teori Maslow juga cocok dalam masalah dunia kerja untuk menjelaskan kepuasan kerja (Mira, et.al., 2019). Hierarki kebutuhan Maslow diperlukan untuk mengukur seberapa puas karyawan dan membantu mempertahankan karyawan (Lahida, et.al., 2017).

2.9.1 Kebutuhan Fisiologis

Kebutuhan fisiologis merupakan tingkat paling dasar dari hierarki kebutuhan Maslow, ini berarti bahwa kebutuhan fisiologis ini adalah kebutuhan yang paling utama dan pertama yang harus terpenuhi agar motivasi seseorang dapat dijaga. Kebutuhan fisiologis ini bisa direlasikan dengan kebutuhan primer manusia seperti, pangan, sandang dan papan. Makanan, tempat tinggal, mendapatkan gaji dan sebagainya dianggap sebagai kebutuhan dasar (S Ghatak, S Singh. 2019). Berkaitan dengan dunia kerja maka kebutuhan fisiologis ini akan terpenuhi dengan adanya gaji.

2.9.2 Rasa Aman dan Ketenangan

Setelah kebutuhan dasar terpenuhi untuk meningkatkan tingkat motivasi maka kebutuhan di tingkat kedua juga harus terpenuhi. Tingkat ini tujuannya adalah memberikan rasa aman dan stabilitas di kehidupan seseorang. Misalnya, pekerjaan tetap memberikan stabilitas dan kehidupan yang terjamin bagi seseorang (Ghatak dan Singh, 2019). Kebutuhan akan keselamatan diwujudkan dengan jaminan kerja, tabungan dan jaminan kesehatan, kesehatan mental, hari tua dan disabilitas (Lahida, et.al., 2017). Keamanan fisik seperti keamanan dan perlindungan dari bahaya kecelakaan kerja dengan memberikan asuransi dan penerapan prosedur K3 (Keselamatan dan Kesehatan Kerja), serta penyediaan transportasi bagi karyawan (Sari dan Dwiarti, 2018). Sedangkan Siagian

dan Sondang (2012) di dalam Jurnal Perilaku dan Strategi Bisnis oleh Sari dan Dwiarti pada tahun 2018 juga menyebutkan tentang keamanan secara psikologis juga penting mendapatkan perhatian seperti perlakuan yang manusiawi dan adil, jaminan akan kelangsungan pekerjaannya, jaminan akan hari tuanya pada saat mereka tidak ada lagi.

2.9.3 Hubungan Sosial dan Cinta

Dalam organisasi, kebutuhan ini mempengaruhi keinginan menjalin hubungan kelompok dengan rekan kerja, partisipasi dalam kelompok kerja, dan hubungan positif dengan supervisor (Lahida, et.al., 2017). Sari dan Dwiarti pada tahun 2018 juga mengartikan kebutuhan sosial ini meliputi kebutuhan untuk persahabatan, afiliasi (hubungan antar pribadi yang ramah dan akrab), dan interaksi yang lebih erat dengan orang lain. Manusia sebagai makhluk sosial diberikan sebuah naluri akan kebutuhan berhubungan dengan orang lain baik untuk berbagi cerita, berkeluh kesah atau berbagi pengalaman untuk menambah pengetahuan. Jalinan hubungan antar manusia ini memberikan keuntungan buat sesama manusia baik dalam hal bertukar informasi sampai kebutuhan tolong menolong jika dibutuhkan. Ketika seseorang tidak mempunyai hubungan sosial yang baik maka seorang manusia akan merasa terpinggirkan, tidak ada tempat berbagi keluh kesah, cerita atau berbagi informasi dan itu akan menyebabkan kesepian yang bisa sampai ke depresi.

Kebutuhan akan hubungan ini menjadi penting bagi manusia, begitupun untuk seorang karyawan dimana di dunia kerja membutuhkan kerja sama yang baik antar karyawan agar performansi kerja yang baik dapat tercapai. Proses bisnis di sebuah perusahaan yang mengalir dari satu bagian ke bagian lainnya membutuhkan proses koordinasi dan komunikasi yang baik, untuk itulah karyawan juga membutuhkan hubungan sosial yang baik dengan sesama karyawan. Dalam organisasi kebutuhan sosial akan berkaitan dengan kebutuhan akan adanya kelompok kerja yang kompak, supervisi yang baik, rekreasi bersama (Sari dan Dwiarti, 2018). Karyawan harus berada pada tingkat kebutuhan ini untuk termotivasi untuk memiliki rasa kepemilikan terhadap misi dan visi organisasi dan mengintegrasikan ambisi pribadi dengan ambisi organisasi, Antara

pengembangan pribadi dan organisasi memiliki timbal balik hubungan yang hasilnya dirasakan secara timbal balik. (Suyono dan Mudjanarko. 2017).

2.9.4 Penghargaan dan Pengakuan

Tingkat keempat dari teori hierarki kebutuhan Maslow adalah kebutuhan akan penghargaan atau pengakuan, yaitu seperti penerimaan penghargaan atau promosi untuk menunjukkan pengakuan atas kerja kerasnya (Vito, et.al., 2016). Kebutuhan akan penghargaan dan pengakuan ini sejatinya merupakan ego untuk berprestasi dan rasa ingin dihargai oleh orang lain. Maslow membagi kebutuhan akan rasa harga diri atau penghargaan ke dalam dua sub, yakni penghormatan dari diri sendiri dan penghargaan dari orang lain (Sari dan Dwiarti, 2018). Pembagian yang pertama adalah keinginan untuk dihormati orang lain, ketenaran, pengakuan, perhatian, reputasi, dan kepemilikan status. Selain kebutuhan eksternal, manusia juga membutuhkan akan keyakinan pada diri sendiri, kompetensi, prestasi, kemandirian, kebebasan, dan penguasaan. Pengakuan pekerjaan-pekerjaan yang karyawan kerjakan; saat mereka melakukan sesuatu yang bagus atau saat mereka berhasil, meyakinkan mereka bahwa perusahaan, pimpinan dan orang sekitarnya mengakui pencapaian itu. Selain itu, pemberian *reward* dengan pengumuman, rekomendasi kenaikan pangkat, pengakuan status, atau hanya sekedar pujian di depan karyawannya yang lain juga bisa dijadikan parameter untuk melakukan penghargaan dan pengakuan. Koeswara (1995) menyatakan bahwa pengakuan atau penghargaan tersebut dapat berupa pujian, pengakuan, piagam, tanda jasa, hadiah, kompensasi, insentif, prestise (wibawa), status, reputasi dll.

2.9.5 Aktualisasi Diri

Tingkat ini mengacu pada kebutuhan pemenuhan diri dan kecenderungan untuk diaktualisasikan dalam potensi seseorang, untuk menjadi yang terbaik yang mampu meningkatkan kesejahteraan (Lahida, et.al., 2017). Tingkatan ini menggambarkan 'kebutuhan pengembangan' misalnya, setiap karyawan sangat antusias ketika mereka mendapat kesempatan untuk menjalani kursus atau studi tertentu, dan kemudian mereka

mendapatkan banyak penghargaan atas pekerjaan dan usaha mereka (Ghatak dan Singh, 2019). Hal ini memungkinkan organisasi untuk sepenuhnya memanfaatkan kemampuan dan potensi karyawan, yang dengan cara ini meningkatkan produktivitas dan efektivitas bisnis secara keseluruhan (Jerome, 2013). Pemenuhan kebutuhan ini dapat dilakukan oleh para pimpinan perusahaan dengan menyelenggarakan pendidikan dan pelatihan, memberikan otonomi untuk berkreasi, memberikan pekerjaan yang menantang, dan lain sebagainya (Sari dan Dwiarti, 2018).

2.10 Rapidminer



Gambar 2.6
Magic quadrant data science and machine learning
 Sumber : (Gartner, November 2019)

Rapidminer adalah *platform data science* dengan menghadirkan kecerdasan buatan yang terbuka. Rapidminer menyatukan seluruh siklus *data science* mulai dari *data preparation* ke *machine learning* sampai ke penerapan model prediktifnya (rapidminer.com). Pengguna Rapidminer diuntungkan dengan adanya *platform* ini karena gratis dan bersifat *open source*, sehingga dapat digunakan tanpa mengeluarkan biaya. Rapidminer dibawah lisensi AGPL (*GNU Affero General Public License*) versi 3.

Rapidminer memiliki antarmuka berbasis GUI yang mudah digunakan dan interaktif, memungkinkan untuk pengguna mudah dalam melakukan proses *data mining* secara aktif.

Rapidminer dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di University of Dortmund yang versi awalnya diluncurkan pada awal 2001. Saat ini, Rapidminer digunakan lebih dari 40 ribu organisasi di semua bidang industri untuk mendapatkan pendapatan, mengurangi biaya produksi serta operasional dan mengendalikan risiko-risiko yang ada yang mungkin muncul dalam proses kegiatan organisasi-organisasi tersebut. RapidMiner tetap menjadi *leader* dengan mencapai keseimbangan yang baik antara kemudahan penggunaan dan kecanggihan ilmu data. Kemampuan *platform*-nya untuk didekati dipuji oleh para ilmuwan data, sementara kekayaan fungsi ilmu data, termasuk keterbukaannya terhadap *source code* dan fungsionalitas, membuatnya juga menarik bagi ilmuwan data yang berpengalaman (Gartner).

2.11 Hasil Penelitian Terdahulu

Penelitian yang berkaitan dengan *data mining* dan pengelolaan SDM yang terdahulu teridentifikasi sebagaimana terlihat pada Tabel 2.6.

Tabel 2.6 Penelitian terkait *data mining* SDM terdahulu

No	Penulis	Judul	Metode	Bidang	Hasil Penelitian
1	Dwi Wanda Bayu S, Aswan Supriyadi	Penerapan <i>Data Mining</i> dalam Menentukan Karyawan Berprestasi dengan Metode Algoritma C4.5	C4.5	-	1. Algoritma C4.5 menghasilkan prediksi karyawan berprestasi dengan mengacu pada nilai dari masing-masing atribut yang muncul pada pohon keputusan. 2. Algoritma C4.5 dapat menghasilkan tingkat akurasi berdasarkan <i>confusion matrix</i> sebesar 75.67%
2	Rendragraha Kumara, Catur Supriyanto	Klasifikasi <i>Data Mining</i> Untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma <i>Decision Tree</i> C4.5	<i>Decision Tree</i> C4.5	Badan Kepegawaian Daerah Kabupaten Banjarnegara	1. Metode <i>decision tree</i> memiliki tingkat akurasi yang baik dalam menyelesaikan solusi. 2. Hasil dari tiga kali pengujian, didapatkan akurasi terbaik dari pengujian

No	Penulis	Judul	Metode	Bidang	Hasil Penelitian
					pertama sebanyak 75% dari <i>dataset</i> . Sebanyak 250 pelamar yang dibagi menjadi 225 pelamar / 90% menjadi data <i>training</i> dan 25 pelamar / 10% menjadi <i>data testing</i>
3	Rino Guphita	Penerapan Klasifikasi Status Pegawai menggunakan Metode Naïve Bayes di RSU H Syaiful Anwar	Naïve Bayes	RSU H Syaiful Anwar	<ol style="list-style-type: none"> 1. Sistem klasifikasi status pegawai ini dapat digunakan untuk mengetahui status pegawai baru. 2. Variabel penentu yang digunakan dalam penelitian ini adalah <i>level</i> pendidikan, pengalaman kerja dan gaji pokok seorang pegawai.
4	Muhammad Septian Fauzi	<i>Data Mining</i> Klasifikasi Untuk Penyeleksian Calon Karyawan Baru Menggunakan Algoritma ID3	ID3	PT. Infomedia Solusi Humanika	Algoritma <i>Decision Tree</i> ID3 akurat diterapkan untuk penentuan diterimanya pelamar dengan tingkat akurasi terbaik yaitu 95,45% dari tiga kali percobaan
5	Budanis Dwi Mulyani	Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode <i>Decision Tree</i>	<i>Decision Tree</i>	Pendidikan	<ol style="list-style-type: none"> 1. Aplikasi membuat jadwal kerja baru berdasarkan jadwal kerja lama dengan menggunakan metode <i>Decision Tree</i> 2. Rule yang digunakan adalah <i>rule</i> januari 2011 karena <i>rule</i> tersebut memiliki nilai akurasi paling tinggi yaitu sebesar 87% 3. Atribut hasil yang digunakan adalah atribut jadwal.
6	Fitria Rachmawati	Analisis Algoritma C4.5 untuk Pengangkatan Karyawan Tetap Studi Kasus PT Citra Abadi Sejati Bogor	C4.5	Perusahaan <i>Garment</i> produk tekstil	<ol style="list-style-type: none"> 1. Variabel yang digunakan : nilai, pendidikan, dan masa kerja. 2. Atribut penilaian dengan nilai tinggi masa kerja dengan nilai masa promosi 11 karyawan, dan dari atribut pendidikan dengan nilai SMA mendapatkan 2 karyawan. 3. Akurasi 99,49 %

No	Penulis	Judul	Metode	Bidang	Hasil Penelitian
7	Amir M Esmmaieeli, Sikaroudi, Rouzbeh Ghousi, Ali Esmmaieeli, Sikaroudi	<i>A Data Mining Approach to Employee Turnover Prediction (Case Study : Arak Automotive Part Manufacturing)</i>	Naïve Bayes, SVN, PNN, KNN	Arak Automotive Part Manufacturing	<ol style="list-style-type: none"> 1. Naïve Bayes - merupakan metode paling <i>user friendly</i> dan memiliki performansi yang bagus dibandingkan dengan SVN, PNN dan KNN. 2. Sedangkan PNN dengan bertambahnya data mempunyai peningkatan akurasi yang lebih baik sejalan dengan bertambahnya waktu dan jumlah data.
8	Qasem A. Al Radaideh, Eman Al Nagi	<i>Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance</i>	<i>Decision Tree</i>	Perusahaan TI	<ol style="list-style-type: none"> 1. Judul Pekerjaan merupakan atribut terkuat kemudian lulusan tipe perguruan tinggi berasal dengan tingkat kelulusan (<i>grade</i>). 2. Umur tidak menunjukkan dampak yang signifikan sedangkan status perkawinan dan jenis kelamin memiliki dampak untuk memprediksi performansi. 3. Rekomendasi untuk mengumpulkan beberapa parameter yang lebih cocok dan tepat lainnya. 4. Rekomendasi untuk menggunakan data karyawan yang telah resign agar dapat digunakan untuk mendapatkan performansi yang lebih baik

Beberapa penelitian pada Tabel 2.6 dijadikan sebagai bahan pembandingan dan bahan referensi dalam penelitian ini. Perbedaan antara penelitian ini dan penelitian-penelitian sebelumnya tersebut pada Tabel 2.6 adalah perihal maksud dari klasifikasi. Di penelitian-penelitian sebelumnya, di Indonesia belum ada penelitian yang berkaitan dengan pengelolaan retensi dan menjaga *turn over* karyawan. Sedangkan untuk penelitian

pengelolaan retensi yang ada di luar negeri, data yang digunakan menggunakan data yang yang diambil dari sistem informasi SDM dan tidak dihubungkan dengan pemenuhan teori hierarki kebutuhan Maslow. Selain itu lokasi penelitian yang terkait dengan ketersediaan data juga berbeda yaitu data yang digunakan data yang berasal dari PT Neuronworks Indonesia. Untuk itulah di penelitian ini akan mencoba menguji berdasarkan data-data yang berelasi dengan teori hierarki kebutuhan Maslow dimana teori tersebut merupakan teori yang dipakai di PT. Neuronworks Indonesia dalam pengelolaan pegawai dan merupakan teori dasar dalam mengelola motivasi karyawan.

BAB III

OBJEK DAN METODOLOGI PENELITIAN

3.1 Profil PT Neuronworks Indonesia

PT Neuronworks Indonesia (Neuron) adalah perusahaan yang bergerak di bidang Teknologi informasi yang berdiri dari tahun 2008 dengan menyediakan solusi dari hulu sampai hilir yang dapat membantu perusahaan dalam menghadapi tantangan bisnis yang berkembang begitu cepat & dinamis. Neuron didirikan berdasarkan akta No 1 tanggal 13 Februari 2008 dari Notaris Dewi Hannawati, SH dan telah mendapat pengesahan dari Menteri Kehakiman dan Hak Asasi Manusia Republik Indonesia dengan surat keputusan Nomor : AHU – 20350.AH.01.01 tahun 2008 tanggal 23 April. Berdasarkan pasal 3 dari Akta Perusahaan, perusahaan bergerak di bidang jasa, kegiatan usaha tersebut termasuk diantaranya menjalankan usaha-usaha Bidang Jasa, Jasa Teknologi Informasi (TI) dan Jasa Pengembangan Peranti Lunak.

Saat ini lebih dari 100 proyek TI skala nasional maupun regional yang telah dikerjakan. Salah satu konsumen dari luar negeri yaitu Telkomcel Timor Leste & Telin Malaysia. Dengan total karyawan 110 orang, Neuron siap memberikan kemudahan untuk menyelesaikan permasalahan sehari-hari melalui solusi TI.



Gambar 3.1
Logo PT Neuronworks Indonesia

Gambar 3.1 menunjukkan Logo Neuron. Logo terdiri dari 2 bagian yaitu 3 garis dan tulisan Neuron beserta dengan *tagline*-nya. Bagian garis yang terdiri dari 3 buah garis berurutan dengan warna merah, merah tua dan hitam mempunyai arti filosofi bagaimana Neuron membangun keunggulan usahanya. Garis-garis condong miring ke kanan menunjukkan kecepatan dalam menjalankan usahanya. Tiap garis dimulai dari bawah kecil lalu membesar ke atas menggambarkan bahwa Neuron berusaha untuk tumbuh dan

membesar dengan berkembangnya jaman. Warna-warna pada garis tersebut mempunyai arti:

1. Merah Cerah

Semangat dan bergairah dalam menemukan solusi untuk pelanggan. Ceria melayani pelanggan.

1. Merah Hati

Menggunakan hati dalam melayani pelanggan, jujur, tulus, ikhlas dan selalu mengutamakan *win-win solution*, mengusahakan agar pelanggan selalu puas terhadap layanan yang diberikan.

2. Hitam Murni

Fokus, tegas dan disiplin mengikuti prinsip-prinsip keilmuan dalam mengembangkan *software* agar tercipta *reliable* dan *zero bug software*. Warna tersebut juga bisa berarti disiplin dalam bekerja.

3.1.1 Visi Misi dan Produk Layanan

Visi Neuron adalah meningkatkan kinerja pelanggan dengan cara melakukan perbaikan, penyempurnaan dan kemajuan di segala sudut proses bisnis pelanggan melalui solusi TI yang unggul dan terbaik. Sedangkan Misinya adalah

1. Membantu mempersiapkan pondasi kokoh bagi pelanggan agar tercapai potensi bisnis yang penuh melalui Teknologi Informasi.
2. Melakukan inovasi secara terus menerus terhadap produk dan pelayanan untuk pencapaian tertinggi dalam setiap solusi.
3. Memberikan solusi Teknologi Informasi yang handal sehingga pelanggan akan lebih mudah dalam menyelesaikan pekerjaan sehari-hari.

Keunggulan usaha perusahaan terjemahan dari visi misi yang ada di logo jika disimpulkan tertera pada Gambar 3.2

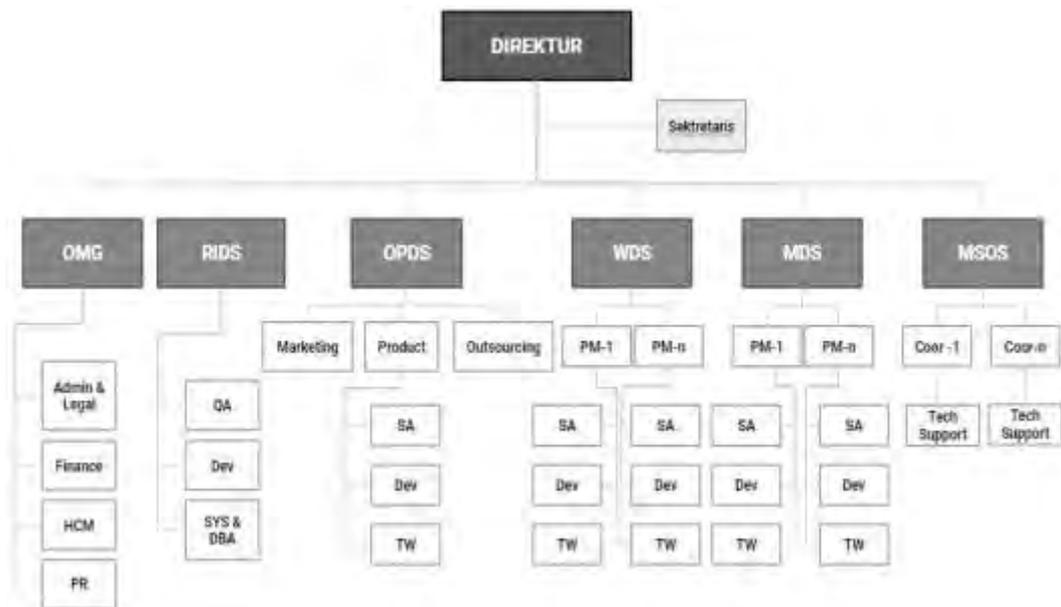


Gambar 3.2
Intepretasi logo dan keunggulan perusahaan

Neuron melayani pembuatan perangkat lunak dengan berbagai jenisnya seperti *Customer Relationship Management System (CRMS)*, *Integration System Solution (ISS)*, *Workflow Management System (WMS)*, *Business Intelligent (BI)* dan *Mobile Application Development*. Produk-produk Neuron antara lain Sistem Informasi Kepegawaian, Sistem Antrian, *SMS Blast System* dan Aplikasi *Survey*. CRMS dihadirkan agar perusahaan-perusahaan dapat memberikan layanan terbaik di setiap waktu. Solusi ISS untuk menyelesaikan tugas integrasi antar sistem yang berbeda dengan tujuan memberikan kemudahan pada perusahaan untuk meningkatkan efisiensi dan efektifitas dalam proses bisnisnya dalam hal komunikasi atau koordinasi antar bagian dan antar sistem. WMS menghadirkan agar perusahaan dapat mengefisienkan dan mengefektifkan proses komunikasi dan koordinasi antar karyawan, bagian dan jabatan. Solusi BI menghadirkan data-data perusahaan secara grafis, yang ditampilkan secara informatif dan mudah dipahami untuk mengevaluasi kinerja sehingga pengambil keputusan dapat dengan cepat mengambil keputusan yang tepat berdasarkan data yang ada.

3.1.2 Struktur Organisasi

Neuron memiliki 6 Divisi antara lain *Operational Management (OMG)*, *Web Development Services (WDS)*, *Mobile Application Development Services (MDS)*, *Maintenance Support Operational Services (MSOS)*, *Research Innovation Development Services (RIDS)* dan *Outsourcing and Product Development Services (OPDS)*. Pembagian struktur organisasi tersebut berdasarkan fungsi operasional, fungsi inovasi dan peningkatan teknologi serta fungsi penjualan berdasarkan bidang penjualan. Struktur organisasi secara lengkap terlihat pada Gambar 3.3.



Gambar 3.3
Struktur Organisasi PT. Neuronworks Indonesia

1. Direktur
 - a. Menentukan kebijakan-kebijakan yang berlaku di perusahaan.
 - b. Memonitoring dan mengontrol kerja produksi (teknis) dan operasional (keuangan,HRD, dan lain-lain) perusahaan.
 - c. Mencari peluang, menjaring, membentuk dan membina kerja sama dengan pelanggan.
 - d. Mengkoordinasikan *Technical Officer* dan *Operasional Officer* agar selaras dan dapat bekerjasama dengan baik.
 - e. Mencari solusi permasalahan perusahaan secara keseluruhan.
 - f. Memimpin Perusahaan.
2. Divisi *Operational Management* (OMG)
 - a. Menyusun rekomendasi kebijakan-kebijakan perusahaan (strategi organisasi dan strategi bisnis).
 - b. Membuat surat menyurat dan melakukan monitor serta mengotrol administrasi dan legalitas perusahaan.
 - c. Mencatat, mengelola dan melaporkan keuangan perusahaan.

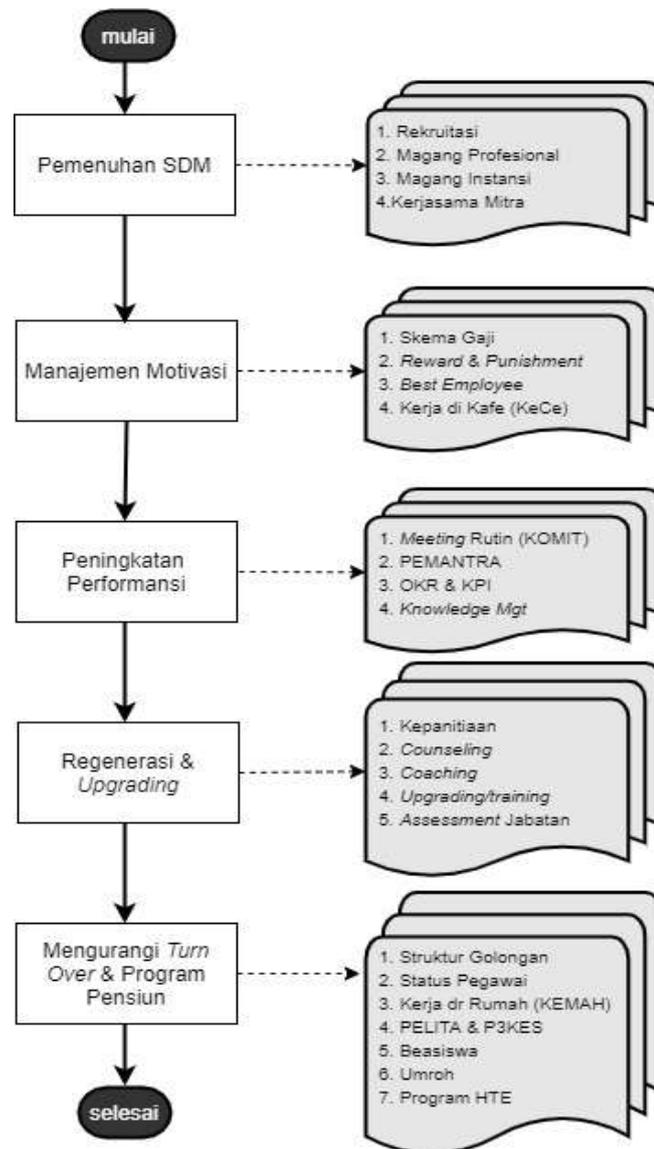
- d. Mengelola SDM secara keseluruhan dan menyusun kebijakan-kebijakan di bidang SDM untuk direkomendasikan.
 - e. Memonitor, mengontrol dan memastikan pengelolaan aset-aset dan alat produksi perusahaan dikelola dengan baik.
 - f. Mengelola hubungan perusahaan dengan *stakeholder-stakeholder* yang ada.
 - g. Melaporkan kepada Direktur perihal kondisi organisasi meliputi kondisi internal, aset-aset, kondisi SDM dan kinerja financial perusahaan.
3. Divisi *Web Development Services* (WDS)
- a. Menyelesaikan seluruh proyek berbasis *web* dan non *mobile app*.
 - b. Meningkatkan kualitas aplikasi yang diserahkan terimakan meliputi aspek *security*, fitur/fungsional solusi, jumlah *bug*, performansi, dan penerapan UX.
 - c. Mengimplementasikan standar sistem *web based development*.
 - d. Membuat SLG *web based development*.
 - e. Membuat parameter penilaian *web programmer*.
 - f. Mengumpulkan *librarys code* sebagai bahan untuk meningkatkan percepatan *time development* dan atau pembuatan *add on* dan *library* untuk memperkaya NAF.
 - g. Mengimplementasikan NAF dalam implementasi proyek.
 - h. Memberikan rekomendasi *script*, *library*, *tools*, *framework* kepada RIDS ketika menemukan kesulitan dalam sehari-hari.
4. Divisi *Mobile Application Development Services* (MDS)
- a. Menyelesaikan seluruh proyek berbasis *mobile apps*.
 - b. Meningkatkan kualitas aplikasi yang diserahkan terimakan meliputi aspek *security*, fitur/fungsional solusi, jumlah *bug*, performansi, dan penerapan *user mobile experience*.
 - c. Mengimplementasikan standar sistem *mobile apps dev*.
 - d. Membuat SLG *mobile apps development*.
 - e. Membuat parameter penilaian *mobile apps programmer*.
 - f. Mengumpulkan *librarys code* sebagai bahan untuk meningkatkan percepatan *time development*.

- g. Mempelajari, memilih dan memakai *frameworks* terbaik untuk pengembangan *mobile apps*.
 - h. Memberikan rekomendasi *script, library, tools, framework* kepada RIDS ketika menemukan permasalahan/solusi sehari-hari.
 - i. Membantu tim produk dalam mengembangkan aplikasi produk versi *mobile*.
5. Divisi *Research Innovation Development Services (RIDS)*
- a. Menyediakan *platform development* Neuron.
 - b. Merumuskan standarisasi dan sosialisasi *platform* Neuron.
 - c. Melakukan penilaian dalam tingkat keberhasilan implementasi *platform* yang sudah dikembangkan.
 - d. Melakukan riset teknologi terbaru untuk diimplementasikan di Neuron.
 - e. Mengukur kualitas dalam kesesuaian standar *platform*.
 - f. Membuat standarisasi *security, scalability, UI* yang digunakan untuk *tools* semua divisi
 - g. Melaporkan hasil pemahaman dan implementasi pada setiap divisi.
 - h. Bersama dengan tim *web based* implementasi proyek menggunakan NAF.
 - i. Menjaga standarisasi NAF terimplementasi dalam implementasi proyek.
6. Divisi *Maintenance and Support Operational Services (MSOS)*
- a. Menyelesaikan seluruh proyek yang sifatnya *Maintenance* dan *Support Operational* dari perangkat lunak yang telah diserahkan.
 - b. Menjaga kualitas support kepada pelanggan dengan menerapkan SLA sesuai kesepakatan dalam kontrak kerja sama dengan pelanggan.
 - c. Mengimplementasikan standar *maintenance* dan *support operational* yang berlangsung selama 24/7.
 - d. Membuat dan mengembangkan *tools* pembantu dalam melakukan *support operational*.
7. Divisi *Outsourcing and Product Development Services (OPDS)*
- a. Menemukan ide dan konsep produk baru melalui riset.
 - b. Memastikan *final* produk sesuai dengan kebutuhan konsumen.

- c. Melakukan update produk berdasarkan permintaan perubahan bisnis.
- d. Memperbaiki proses *development* agar produktif dan menghasilkan produk yang berkualitas
- e. Mengikuti *best practices* agar proses development produk berjalan lancar dan berkesinambungan.
- f. Mengelola dokumentasi kebutuhan, fungsi dan teknik untuk referensi di kemudian hari.
- g. Melakukan riset tren pengembangan teknologi aplikasi terbaru.
- h. Menganalisis, memecahkan masalah dan menemukan solusi isu-isu permasalahan produk dengan tepat waktu.
- i. Memasarkan dan menjual produk yang sudah dirilis
- j. Mengelola proyek-proyek yang tidak bisa dikerjakan oleh tim internal dan memungkinkan untuk dilakukan *outsourcing* baik tenaga individu atau perusahaan TI lainnya.

3.1.3 Proses Retensi Karyawan

Karyawan merupakan aset paling berharga bagi Neuron. Sebagai perusahaan yang bergerak di bidang Jasa Konsultan TI maka karyawan merupakan kekuatan utama yang mampu menggerakkan performansi kinerja perusahaan menjadi baik. Proses pengelolaan karyawan di Neuron dimulai dari rekrutasi sampai ke pengelolaan di masa pensiun. Proses tersebut terlihat pada Gambar 3.4, setiap proses atau tahapan pada gambar tersebut berelasi dengan program kerja, kegiatan atau mekanisme perusahaan yang berelasi dengan konsep teori hierarki kebutuhan Maslow. Beberapa program dibuat agar karyawan sebagai aset berharga perusahaan dapat menyelesaikan pekerjaannya dengan baik dan loyal terhadap perusahaan. Kendala paling utama dari sebuah perusahaan jasa konsultan TI adalah *turn over* yang tinggi, sehingga pengerjaan proyek menjadi terhambat dan perusahaan tersebut tidak mampu untuk menyelesaikan proyek dengan mulus dan baik. Untuk itulah Neuron berusaha untuk mengelola karyawannya dengan baik.



Gambar 3.4
Proses bisnis pengelolaan retensi di Neuron

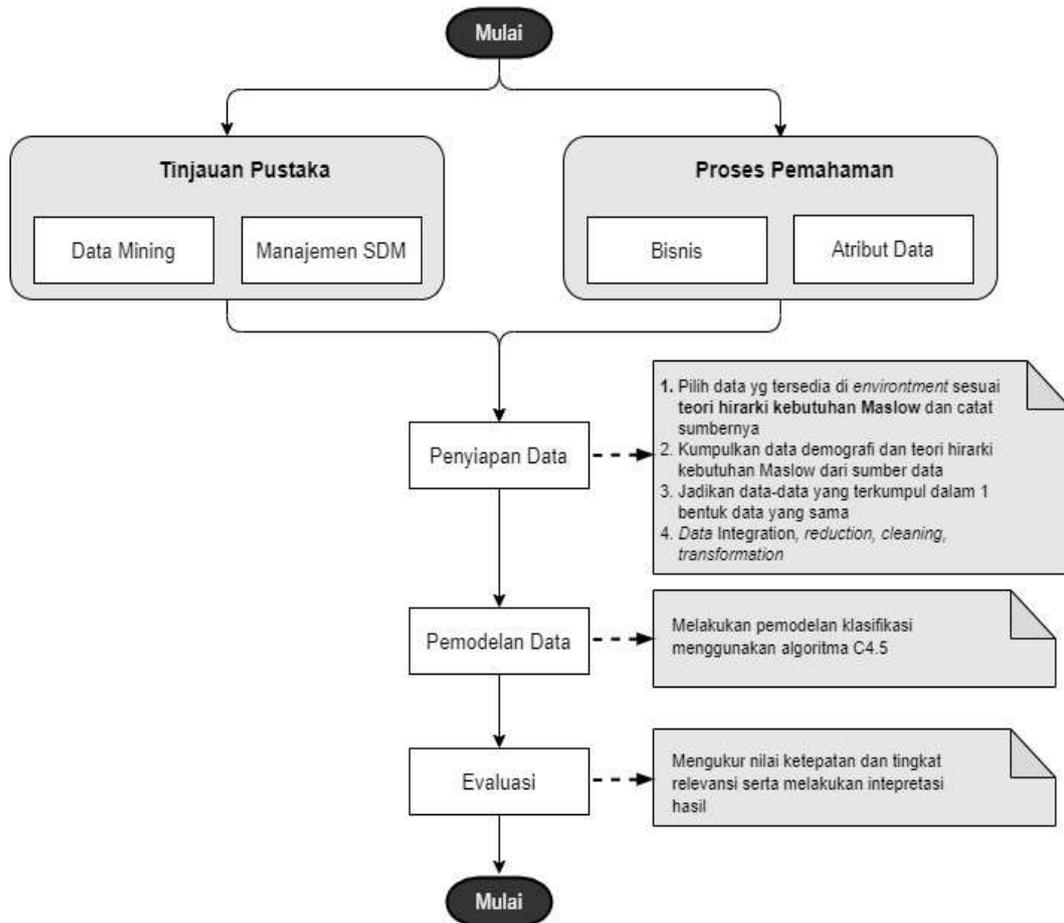
Neuron mencoba untuk mengimplementasikan teori hierarki kebutuhan Maslow sebagai teori awal motivasi manusia untuk mengelola retensi karyawan. Pengelolaan karyawan menggunakan teori hierarki kebutuhan Maslow di Neuron diimplementasikan dengan memenuhi aspek-aspek di setiap tingkat hierarkinya. Penjelasan keterhubungan tiap tingkat pada teori hierarki kebutuhan Maslow dan kegiatan pengelolaan karyawan ditunjukkan pada Tabel 3.1, setiap kegiatan yang ada pada tingkatan tersebut merupakan gambaran usaha Neuron dalam menerapkan konsep teori hierarki kebutuhan Maslow untuk mengelola retensi karyawan di Neuron.

Tabel 3.1 Program kerja Neuron berdasarkan Hierarki Maslow

Tingkat Hierarki Maslow	Kegiatan Pengelolaan Karyawan
Aktualisasi Diri	Neuron mengadakan kegiatan ekstrakurikuler seperti kegiatan futsal, bulutangkis dan yoga untuk karyawan perempuan. Juga terdapat kegiatan belajar mengaji Tahsin. Selain itu ada program beasiswa untuk karyawan sehingga setiap karyawan memiliki kesempatan untuk meningkatkan Pendidikan dan pengetahuan untuk melanjutkan studi. Selain itu terdapat <i>upgrading knowledge</i> dan <i>skill</i> yang rutin dilakukan setiap tahunnya.
Pengakuan dan Penghargaan	Pemantra adalah salah satu program penghargaan yang dilakukan setiap bulannya. Penghargaannya berupa penambahan bonus untuk gaji bagi karyawan yang disiplin dan produktif tiap bulannya. Penghargaan lainnya adalah di tengah tahun dan akhir tahun berdasarkan penilaian tengah tahun dan akhir tahun. Pengakuan terhadap karyawan dilakukan dengan <i>open assessment</i> untuk lelang jabatan, sehingga karyawan memungkinkan untuk naik jabatan sebagai bentuk apresiasi atas performansinya.
Hubungan Sosial dan Cinta	Neuron meningkatkan hubungan antar karyawan agar tercipta soliditas dan hubungan yang baik dengan melakukan <i>company gathering</i> dua kali dalam setahun serta kegiatan lainnya yang sifatnya kebersamaan untuk mempererat hubungan antar anggota tim seperti kegiatan kerja di kafe, perlombaan ulang tahun Neuron dan acara tujuh belasan.
Rasa Aman dan Ketenangan	Sesuai peraturan pemerintah, maka Neuron membayarkan Jaminan Pensiun, Jaminan Hari Tua. Untuk menambah tingkat ketenangan dan rasa aman akan masa depan Neuron membuat program Hari Tua Eksklusif yang merupakan tambahan uang pensiun bagi karyawan tetap yang diperoleh dari penyisihan laba tiap tahunnya. Neuron memperhatikan tingkat lembur karyawan dan kesehatan karyawan dengan program P3Kes.
Kebutuhan Fisiologis	Pemenuhan penggajian sesuai dengan peraturan pemerintah (UMR) dan terdapat pemberian bonus tengah tahun. Jika Neuron mengalami performansi keuangan yang baik dengan menunjukkan peningkatan laba perusahaan, maka karyawan diberikan bonus akhir tahun.

3.2 Metodologi Penelitian

Metodologi Penelitian yang dilakukan terdefinisi seperti pada Gambar 3.5, merupakan urutan proses penelitian yang dimulai dari Perumusan Masalah, Tinjauan Pustaka, Pemahaman Bisnis dan Atribut Data, Penyiapan Data, Pemodelan Data dan Evaluasi.



Gambar 3.5
Alur tahapan penelitian

Alur tahapan penelitian akan digunakan sebagai acuan untuk melakukan detail penelitian dan pembahasan yang akan dipaparkan pada Bab IV, penjelasannya sebagai berikut :

1. Perumusan Masalah

Perumusan Masalah merupakan tahap pertama pada penelitian, dimana proses menentukan masalah-masalah yang ada baik dalam kehidupan sehari-hari atau berdasarkan sumber bacaan sebelumnya berkaitan dengan tema *data mining*. Tema *data mining* merupakan tema yang menarik saat ini dimana istilah analitikal data sedang *booming* dengan berjalannya pertambahan data dan informasi di dunia teknologi informasi dan internet. Masalah sehari-hari ditemukan di lingkungan kerja yang berkaitan dengan pengelolaan karyawan yang seringkali menghadapi kendala. Masalah tersebut dirumuskan untuk dicari penyelesaiannya dan tujuan akhir dari penyelesaian masalah

tersebut. Tahap ini melakukan proses penggalian perumusan masalah dan tujuan *data mining* di Neuron.

2. Tinjauan Pustaka

Pada tahap ini dilakukan tinjauan pustaka baik berkaitan dengan *data mining* beserta metode-metode yang ada, algoritma-algoritma yang dapat dipakai di *data mining* sampai pada penggunaan *tools data mining*. Selain mempelajari *data mining*, juga dilakukan kajian terhadap teori pengelolaan sumber daya manusia perihal teori hierarki kebutuhan Maslow dan berkaitan dengan retensi karyawan.

3. Pemahaman Bisnis dan Atribut Data

Sebelum menyiapkan data, agar data-data yang disiapkan tepat dan sesuai maka harus melakukan pemahaman bisnis yang berlaku di Neuron. Setiap perusahaan tentu saja memiliki kekhasan sendiri-sendiri, untuk itulah perlu dilakukan pemahaman bagaimana Neuron mengelola SDM-nya bagaimana pemenuhan hierarki maslow di Neuron. Data-data seperti apa yang sesuai dengan penelitian ini yang dapat digunakan untuk analisa yang tepat dan dimodelkan. Apa maksud dari setiap kegiatan yang dipunyai dan data yang ada pada setiap kegiatan dapat digunakan untuk apa, bagaimana karakteristiknya dan perlu diperlakukan seperti apa.

4. Penyiapan Data

Data penelitian didapatkan dari beberapa sumber seperti aplikasi *Human Resource Management Information System* (HRMIS), data kegiatan ekstrakurikuler, data penilaian tengah tahun dan akhir tahun, kuisisioner dan data media sosial. Data yang di dalam HRMIS dikumpulkan seperti data karyawan, demografi, data gaji, absensi, data lembur data jam masuk dan lain sebagainya. Sedangkan data penilaian tengah tahun dan akhir tahun diambil dari hasil evaluasi tiap tahunnya yaitu parameter yang berkesesuaian seperti data *teamwork* dan perilaku. Data media sosial dilakukan *grab* data dari media sosial *Instagram* yang meliputi, jumlah *follower* sesama karyawan di Neuron dan jumlah *like post* oleh rekan di Neuron. Dengan mengumpulkan data dari beberapa sumber data diharapkan atribut-atribut data yang didapatkan dapat memenuhi setiap tingkat dari hierarki kebutuhan Maslow. Setelah dilakukan

pengambilan data dan diintegrasikan menjadi satu kesatuan *dataset*, maka dilakukan proses *data pre-processing* seperti *data cleaning*, *data integration*, *data reduction*, dan *data transformation*.

5. Pemodelan Data

Pada tahap ini dilakukan pemrosesan data lanjutan sesuai input data yang telah ditentukan berdasarkan atribut-atribut yang berkesesuaian dengan hierarki kebutuhan Maslow. Pemrosesan tersebut dilakukan menggunakan algoritma *data mining* yang sesuai dengan tipe *data mining* berdasarkan perumusan dan tujuan di awal, dalam hal ini adalah tipe *data mining* klasifikasi dan algoritma C4.5. Pada tahap ini dilakukan beberapa percobaan untuk menemukan model terbaik yang akan digunakan sebagai bahan rekomendasi perbaikan pengelolaan retensi karyawan di Neuron.

6. Evaluasi

Setelah dilakukan pemodelan dan pemrosesan menggunakan algoritma *data mining*, maka tahap selanjutnya adalah melakukan evaluasi dari hasil pemrosesan tersebut untuk menilai sejauh mana model yang diterapkan relevan dengan proses klasifikasi karyawan untuk pengelolaan retensi karyawan. Evaluasi tersebut meliputi seberapa tepat pemodelan yang dilakukan dan tingkat relevansinya. Jika hasilnya kurang tepat dan tingkat relevansinya rendah maka dilakukan pengulangan proses dari *Data Preparation* sampai pemodelan sampai ditemukan nilai yang lebih akurat dan angka relevansinya tinggi. Pada tahap evaluasi ini dipilih model terbaik yang selanjutnya akan dilakukan interpretasi data sehingga didapatkan rekomendasi perbaikan pengelolaan retensi di Neuron berdasarkan teori hierarki kebutuhan Maslow.

3.3 Data Penelitian

Data penelitian diambil dari data karyawan dan program kerja yang berkaitan dengan pengelolaan retensi di Neuron dari tahun 2016 sampai tahun 2021. Penelitian ini akan menggunakan 2 macam data yaitu data demografi karyawan dan data atribut karyawan sesuai teori hierarki kebutuhan Maslow.

1. Data Demografi Karyawan

Data pertama ini diambil dari *Human Resource Management Information System* (HRMIS) yang dipunyai oleh Neuron. Data tersebut diambil dari tabel profil karyawan yang mempunyai atribut antara lain nik, umur, jenis kelamin, status (menikah/tidak), tingkat pendidikan, jumlah anak dan asal sekolah. Data Demografi ini dibutuhkan untuk mengetahui profil dasar karyawan dan hubungannya dengan standar nilai di setiap tingkat dari teori hierarki kebutuhan Maslow.

2. Data Berbasis Teori Hierarki Maslow

Sekumpulan data yang masuk dalam kategori data kedua ini adalah data yang diidentifikasi berkaitan dengan penerapan teori hierarki kebutuhan Maslow. Neuron berusaha menerapkan teori hierarki kebutuhan Maslow dalam mengelola dan memotivasi karyawannya.

Tabel 3.2 Atribut-atribut data dan sumber data

Hierarki Maslow	Atribut <i>Dataset</i>	Sumber Data
Aktualisasi Diri	<ol style="list-style-type: none"> 1. Keikutsertaan Kegiatan Ekstrakurikuler 2. Keikutsertaan upgrading atau pelatihan 3. Beasiswa Sekolah 	<ol style="list-style-type: none"> 1. Data keikutsertaan sertifikasi 2. Laporan Kegiatan ekstrakurikuler 3. Data Penerima Beasiswa
Penghargaan dan Pengakuan	<ol style="list-style-type: none"> 1. Jumlah <i>monthly best employee</i> 2. Penghargaan Tahunan 3. Pernah Rekomendasi <i>Assessment</i> 4. Bonus Umroh 	<ol style="list-style-type: none"> 1. HRMIS <i>Gamification Performance Mgt</i> 2. HRMIS data 3. Data Program Umroh 4. Laporan <i>Assessment</i> kenaikan jabatan
Hubungan Sosial & Cinta	<ol style="list-style-type: none"> 1. <i>Status follow</i> medsos Perusahaan 2. Jumlah <i>like Post</i> Perusahaan 3. Jumlah <i>follower</i> Medsos se-Neuron 4. Evaluasi 360 parameter <i>teamwork</i> 5. Jumlah kehadiran 	<ol style="list-style-type: none"> 1. Kuisisioner 2. Laporan Penilaian dan Evaluasi Karyawan 3. HRMIS <i>Performance Mgt</i> 4. HRMIS <i>Attendance</i>
Rasa Aman & Ketenangan	<ol style="list-style-type: none"> 1. Status Kepegawaian 2. Lama Bekerja 3. Jumlah Lembur 4. Jumlah Cuti 5. Jumlah Sakit 6. Perolehan Dana Pensiun Tambahan 	<ol style="list-style-type: none"> 1. HRMIS <i>Profil</i> 2. HRMIS <i>Attendance</i> 3. Data perhitungan pensiun tambahan
Kebutuhan Fisiologis	<ol style="list-style-type: none"> 1. Jumlah Gaji 2. Jumlah Bonus Tahunan 	HRMIS <i>Payroll</i>

Pada Tabel 3.2 terlihat bahwa data-data seperti gaji, jumlah lembur, perolehan dana pensiun tambahan (program HTE), penghargaan-penghargaan seperti pemberian *best employee*, keikutsertaan ekstrakurikuler dan pelatihan sampai data keterikatan karyawan dengan sesama karyawan lainnya yang didapatkan dari media sosial dikumpulkan di dalam kategori ini. Data jenis inilah yang menjadi perbedaan penelitian ini dengan penelitian lainnya. Cara mengumpulkan data jenis ini adalah melakukan penelusuran macam-macam program kerja atau kegiatan untuk pengelolaan karyawan lalu mencatat data-data yang terlibat di dalam kegiatan tersebut. Selanjutnya setelah dilakukan pencatatan data yang terlibat dilakukan pemetaan dengan teori hierarki kebutuhan Maslow. Setelah seluruh tingkat dari hierarki kebutuhan Maslow terpenuhi baik kegiatan dan data yang terlibat maka dicatat sumber data untuk mendapatkan data tersebut yang nantinya akan berguna ketika tahap penyiapan data dalam proses ekstraksi data.

BAB IV

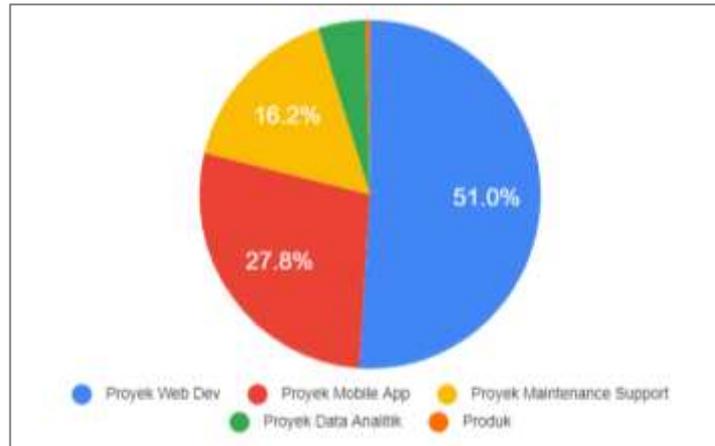
HASIL DAN PEMBAHASAN

4.1 Tahapan Pemahaman

Sebelum melakukan pemodelan menggunakan metode klasifikasi dan algoritma pohon keputusan yaitu C4.5 maka perlu dilakukan penyiapan data. Penyiapan data ini membutuhkan pengetahuan tentang kebutuhan bisnis sesuai objek penelitian yaitu Neuron. Kebutuhan bisnis Neuron apa dan bagian yang mana sehingga diperlukan klasifikasi menggunakan *data mining*. Sebagaimana ditentukan dalam perumusan masalah dan tujuan pada Bab I, perlu digali lagi lebih detail latar belakang dan proses bisnis di Neuron yang akan dijelaskan pada sub bab pemahaman proses bisnis. Kemudian setelah memahami latar belakang dan proses bisnis Neuron maka perlu untuk memilih atribut-atribut yang dibutuhkan dan memahami karakteristik data serta pentingnya mengapa atribut tersebut digunakan dalam pemodelan. Pemahaman data ini sangat penting sebelum dilakukan proses penyiapan data misalnya saja pemahaman akan data yang terlibat dalam pengelolaan retensi di Neuron yang menggunakan konsep teori hierarki kebutuhan Maslow. Pemahaman data ini akan dibahas pada sub bab pemahaman data.

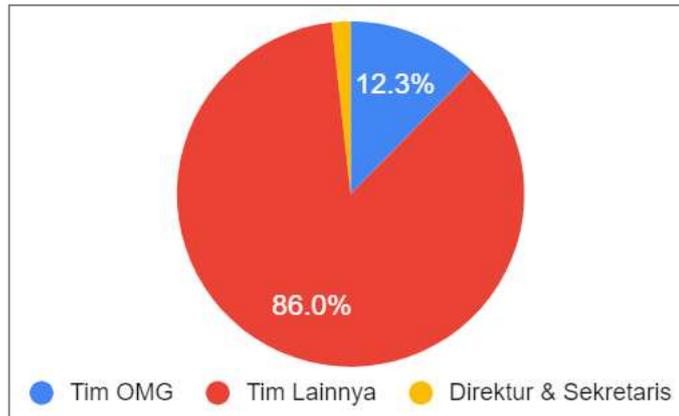
4.1.1 Pemahaman Bisnis

Neuron merupakan perusahaan konsultan TI yang proses penyelesaian pekerjaan pengembangan perangkat lunak atau pembuatan produk perangkat lunak bergantung pada karyawannya. Oleh karena itu Neuron menganggap bahwa karyawan merupakan aset terpenting mereka. Sesuai portofolio yang telah dijelaskan pada Bab III dan berdasarkan hasil wawancara dengan pihak Neuron, pendapatan terbesar mereka didapatkan dari pekerjaan proyek pengembangan perangkat lunak. Komposisi pendapatan tersebut dapat dilihat pada Gambar 4.1, yang menunjukkan bahwa mayoritas pendapatan didapatkan dari pekerjaan proyek antara lain *web development*, *mobile application*, data analitik dan *maintenance support operational*. Pekerjaan berbasis proyek tersebut menyebabkan Neuron tergantung pada kondisi karyawannya.



Gambar 4.1
Komposisi pendapatan Neuron

Pada bagan struktur organisasi di Bab III terlihat komposisi bagian dan anggota didominasi oleh karyawan bagian produksi dalam hal ini pekerja proyek pengembangan perangkat lunak. Komposisi tersebut disimpulkan dalam gambar yang terlihat pada Gambar 4.2. Pekerjaan proyek pengembangan perangkat lunak ini membutuhkan *programmer* untuk proses penyelesaiannya. Jangka waktu pengerjaan pekerjaan tersebut biasanya lebih dari 6 bulan, maka sangat penting bagi Neuron untuk menjaga aset terpenting yaitu karyawannya. Hal ini juga berlaku bagi Neuron untuk memenangkan persaingan dengan pesaingnya, tentu saja dengan menjaga karyawannya tetap bertahan dan berkerja dengan baik sampai proyek pengembangan perangkat lunak selesai. Keberhasilan untuk dapat menjaga karyawannya bertahan dan bekerja dengan baik menyebabkan kewajiban Neuron dapat terpenuhi dan keberhasilan proyek itu akan memberikan nama baik sehingga permintaan proyek tetap berdatangan di kemudian hari. Berbeda jika Neuron gagal mempertahankan karyawan terbaiknya, misalnya karyawan terbaik memilih keluar dari perusahaan apalagi jika pindah ke pesaing, hal ini akan menyebabkan proses penyelesaian proyek terhambat.



Gambar 4.2
Komposisi karyawan operasional dan non operasional

Proses perekrutan karyawan baru untuk membantu penyelesaian proyek tidak semudah membalikkan telapak tangan, butuh tahapan-tahapan yang harus dilalui agar karyawan baru tersebut dapat membantu penyelesaian proyek. Tahapan tersebut dimulai dengan proses rekrutasi kemudian proses beradaptasi melalui program pengenalan, pelatihan dan *transfer knowledge* kemudian baru meneruskan pekerjaan yang ditinggalkan oleh karyawan sebelumnya. Jadi Neuron memilih untuk mengelola retensi karyawan dengan sebaik mungkin sehingga hal tersebut tidak mengganggu jalannya proses penyelesaian pekerjaan proyek pelanggannya.

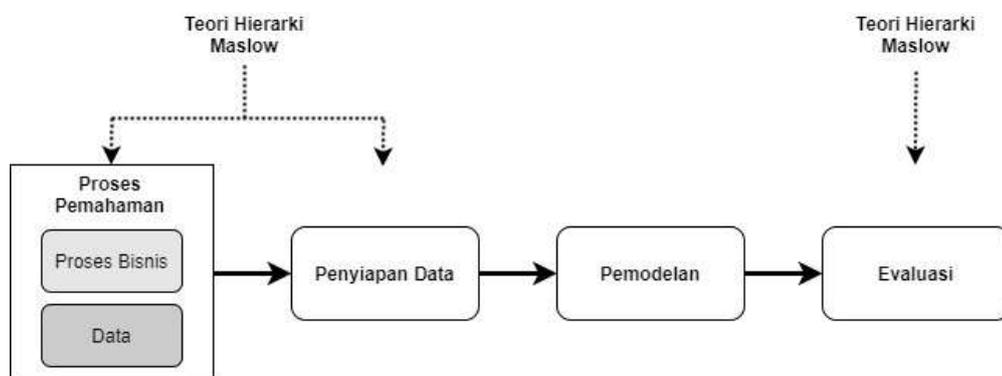
Dari deskripsi profil dan proses bisnis pada Bab III terlihat bahwa Neuron mempunyai program-program pengelolaan retensi karyawan. Neuron menentukan konsep dan strategi berdasarkan teori hierarki kebutuhan Maslow. Proses pengelolaan retensi di Neuron tergambar pada Gambar 3.4 sedangkan program-programnya terpaparkan dengan lebih detail pada Tabel 3.1. Program-program tersebut telah dijalankan namun tiap tahun tetap ada karyawan yang mengundurkan diri. Hal ini tentunya menjadi sebuah landasan pemikiran, apakah pelaksanaan program-program yang telah didesain untuk mengelola retensi karyawan cukup berhasil. Apakah ada program yang belum optimal dalam pelaksanaannya sehingga menjadikan faktor karyawan untuk mengundurkan diri. Menjadi sangat penting, pelaksanaan program-program pengelolaan retensi karyawan di Neuron dievaluasi. Proses evaluasi tersebut harus dimulai dengan melakukan pengumpulan data-data yang melekat pada program-program tersebut kemudian

dilakukan proses klasifikasi menggunakan algoritma pohon keputusan C4.5. Klasifikasi tersebut untuk mendapatkan pengetahuan program pada tingkat hierarki kebutuhan Maslow mana yang mempunyai pengaruh terhadap keputusan *resign* karyawan Neuron.

Kesimpulan dari pemahaman bisnis tersebut adalah dibutuhkannya klasifikasi menggunakan *data mining* untuk menemukan hal-hal dan dengan tujuan sebagai berikut :

1. Apa saja yang berpengaruh terhadap keputusan karyawan untuk *resign* berdasarkan program-program yang telah dilaksanakan oleh Neuron.
2. Rekomendasi program-program apa yang perlu ditingkatkan untuk membantu pengelolaan retensi karyawan.

Berdasarkan pemahaman latar belakang dan kebutuhan bisnis tersebut maka diperlukan penyesuaian proses *data mining* dengan teori hierarki kebutuhan Maslow untuk menemukan pengetahuan yang tepat. Pengetahuan yang dihasilkan harus memiliki akurasi yang baik agar tepat guna mengelola retensi karyawan di Neuron. Tahapan *data mining* yang terpengaruh untuk menyesuaikan terhadap kebutuhan ini adalah tahapan proses pemahaman bisnis dan pemahaman data, proses penyiapan data termasuk di dalamnya ekstraksi data, *data cleaning* dan *data transformation*, serta yang paling penting pada tahapan evaluasi. Tahap evaluasi dilakukan untuk menentukan model mana yang memiliki akurasi tinggi, dan bagaimana menerjemahkan hasil model yang selaras dengan pengelolaan retensi karyawan menggunakan teori hierarki kebutuhan Maslow. Hal tersebut dapat digambarkan seperti pada Gambar 4.3.



Gambar 4.3
Penyelarasan proses *data mining* dan teori hierarki Maslow

4.1.2 Pemahaman Data

Neuron mempunyai konsep pengelolaan retensi menggunakan teori hierarki kebutuhan Maslow. Program-program yang ada pada Neuron telah terpetakan pada setiap hierarki kebutuhan Maslow seperti pada Tabel 3.1 dan hasil dari program-program tersebut terdapat data dan laporan yang tersimpan pada tempat dan bentuk yang berbeda. Hal tersebut tergambar pada Tabel 3.2. Dari tabel tersebut terlihat bahwa untuk melakukan klasifikasi harus mengumpulkan data-data dari beberapa sumber. Ada 9 sumber data yang berbeda untuk memenuhi data-data sesuai teori hierarki kebutuhan Maslow yang dibutuhkan dalam proses klasifikasi dengan *data mining*. Setiap data pada sumber yang berbeda tersebut harus direlasikan dengan data primer karyawan, sehingga terjamin integrasi datanya. Data primer yang dipilih pada penelitian ini adalah Nomor Induk Kependegawaian (NIK).

Tabel 4.1 Contoh variasi data tiap tahun

NIK	Jenis Kelamin	Umur	Tingkat Pendidikan	Lama Bekerja	Status Menikah
1931312054	Laki-laki	23	D3	3	Belum Menikah
1931312054	Laki-laki	24	S1	4	Sudah Menikah
95140155	Wanita	23	SMA Sederajat	4	Sudah Menikah
95140155	Wanita	24	S1	5	Sudah Menikah
1911401056	Wanita	26	D3	3	Belum Menikah
1911401056	Wanita	28	S1	5	Sudah Menikah
91140961	Laki-laki	25	S1	2	Belum Menikah
91140961	Laki-laki	26	S2	3	Sudah Menikah

Sumber: HRMIS Neuron

Setiap tahun, data primer karyawan tetap namun beberapa data demografinya berubah seperti umur, status menikah, jumlah anak dan tingkat pendidikan. Contoh variasi data tiap tahun ditunjukkan pada Tabel 4.1, terlihat data karyawan yang sama namun di tahun berbeda memiliki data yang berbeda untuk masing-masing atribut. Data tingkat pendidikan berubah karena ditemukan bahwa beberapa karyawan di Neuron melanjutkan pendidikan formal selama bekerja, sehingga ketika telah selesai pendidikan formal tersebut dan dinyatakan lulus oleh instansi pendidikannya maka status pendidikan karyawan tersebut berubah. Hal lain yang berubah tiap tahun adalah parameter lama bekerja, lama bekerja setiap tahunnya bertambah. Hal sama ditemukan untuk data dari hasil pengelolaan retensi

menggunakan teori hierarki Maslow misalnya data gaji, jumlah kehadiran, jumlah lembur, keikutsertaan kegiatan ekstrakurikuler dan rekomendasi *assessment* untuk kenaikan jabatan. Status karyawan juga bisa berubah untuk karyawan yang *resign* dan belum *resign*. Oleh karena itu data yang akan digunakan pada proses klasifikasi pada penelitian ini adalah data tiap karyawan tiap tahun baik data demografi dan data berdasarkan pengelolaan retensi menggunakan teori hierarki kebutuhan Maslow. Untuk itu data primer yang awalnya NIK perlu untuk dirubah menjadi NIK ditambah tahun kejadian atau tahun data.

Sesuai pemahaman proses bisnis, dapat disimpulkan bahwa untuk membangun klasifikasi agar mengetahui apa saja yang berpengaruh terhadap keputusan karyawan untuk *resign* maka label data untuk proses klasifikasi tersebut adalah atribut *resign*. Label ini berisi data yang menjelaskan status *resign* tiap karyawan Neuron. Data label dapat ditemukan di sumber data aplikasi *Human Resource Information System* (HRMIS) Neuron. Pada aplikasi HRMIS, data status karyawan tercatat aktif dan tidak aktif, serta tercatat kapan status tersebut berubah dari aktif menjadi tidak aktif. Karyawan yang mempunyai status tidak aktif maka dianggap karyawan tersebut *resign*, dan waktu yang tercatat untuk perubahan status tersebut merupakan kapan karyawan tersebut *resign*. Data ini yang akan diekstraksi pada proses penyiapan data untuk direlasikan dengan data primer karyawan

Data dari berbagai sumber yang berbeda biasanya mempunyai format berbeda, jenis dan bentuk yang berbeda dan jumlah yang berbeda. Hal ini menjadi tantangan ketika akan melakukan tahap pemodelan. Oleh karena itu menjadi sangat penting proses penyiapan data. Data dari berbagai sumber biasanya memiliki nilai yang berbeda untuk satu jenis atribut data misalnya data identitas primer karyawan bisa jadi dari sumber yang berbeda mencantumkan nama sedangkan sumber lainnya mencantumkan NIK sebagai data primer.

Data dari berbagai sumber juga tidak sama jumlahnya misalnya laporan kegiatan umroh. Tidak semua karyawan tercatat pada laporan sedangkan untuk data yang diambil dari sumber aplikasi HRMIS data seluruh karyawan terekam seluruhnya, baik karyawan aktif maupun yang sudah tidak aktif atau karyawan tetap dan karyawan kontrak. Data yang

diambil dari laporan umroh hanya karyawan yang telah melakukan umroh yang Namanya tertera di laporan. Data jenis seperti ini ditemukan juga untuk data laporan *assessment* kenaikan jabatan. Data dana pensiun tambahan yang disebut Hari Tua Eksklusif (THE) juga hanya menyimpan data karyawan-karyawan tetap karena program tersebut dikhususkan hanya kepada karyawan dengan status tetap saja. Keterangan detail mengenai karakteristik data tersebut dapat ditemukan pada Tabel 4.2.

Hal ini menandakan bahwa proses penyiapan data bisa dibilang merupakan tahapan penting dan berat pada penelitian ini. Proses pembersihan data karena nilai yang berbeda dan format berbeda harus teliti. Proses integrasi data dikarenakan data-data yang ada dari beberapa sumber berbeda dengan format berbeda dan bentuk yang berbeda harus berjalan dengan baik. Proses Integrasi dapat dilakukan dengan cara menyamakan format, bentuk dan jenis *file* terakhir sebelum dilakukan pemodelan. Oleh karena itu di dalam tahap penyiapan data terdapat proses ekstraksi data dari berbagai sumber, menyamakan jenis dan format data, *data integration*, *data cleaning*, *data reduction* dan *data transformation*. Status data pada tiap *dataset* mengenai perlu tidaknya dilakukan data *cleaning* dan data *transformation* terdapat pada Tabel 4.2.

Tabel 4.2 Pemahaman *dataset* pengelolaan retensi karyawan

Dataset	Atribut	Keterangan	Data Cleaning	Data Transformation
Demografi	1. NIK 2. Umur 3. Jenis Kelamin 4. Status Menikah 5. Jumlah anak 6. Alamat asal	Tersedia untuk seluruh karyawan baik yang masih aktif atau sudah tidak aktif. NIK unik karena merupakan <i>primary key</i> dalam data profil karyawan. Terjamin integrasi datanya.	Tidak	Perlu
Aktualisasi Diri	1. Keikutsertaan Kegiatan Ekstrakurikuler 2. Keikutsertaan Pelatihan/Seminar 3. Beasiswa Sekolah	Bisa saja ada data yang hilang dan hanya tercatat yang mengikuti kegiatan ekstrakurikuler, pelatihan atau mendapatkan beasiswa sekolah. Data tersimpan dalam catatan dokumen bukan aplikasi. Karakteristik data rawan inkonsistensi dan terdapat <i>missing data</i> .	Perlu	Perlu
Penghargaan & Pengakuan	1. Jumlah <i>monthly best employee</i>	Jumlah <i>monthly best employe</i> karena didapatkan dari HRMIS, akan terdapat data seluruh karyawan.	Perlu	Perlu

Dataset	Atribut	Keterangan	Data Cleaning	Data Transformation
	2. Penghargaan Tahunan 3. Pernah Rekomendasi <i>Assessment</i> 4. Bonus Umroh	Sedangkan untuk penghargaan tahunan atau tengah tahun, rekomendasi <i>assessment</i> dan bonus umroh hanya karyawan yang mendapatkan saja yang tercatat.		
Hubungan Sosial & Cinta	1. <i>Status follow</i> medsos Perusahaan 2. Jumlah <i>like Post</i> Perusahaan 3. Jumlah <i>follower</i> Medsos se-Neuron 4. Evaluasi 360 parameter <i>teamwork</i> 5. Jumlah kehadiran	Pada <i>dataset</i> ini untuk data dari media sosial, Instagram, tidak semua karyawan memiliki, sehingga dimungkinkan ada data yang tak terdapat nilai. Sedangkan untuk Evaluasi 360 parameter <i>teamwork</i> , jumlah kehadiran dan tingkat kedisiplinan data seluruh karyawan dapat ditemukan.	Perlu	Perlu
Rasa Aman & Ketenangan	1. Status Kepegawaian 2. Lama Bekerja 3. Jumlah Lembur 4. Jumlah Cuti 5. Jumlah Sakit 6. Perolehan Dana Pensiun Tambahan	Semua data pada <i>dataset</i> ini diambil dari aplikasi HRMIS jadi dipastikan data seluruh karyawan ditemukan, namun untuk data perolehan dana pensiun tambahan, karena dana tersebut hanya diperuntukan karyawan dengan status tetap maka data yang akan ditemukan hanya karyawan tetap, data karyawan kontrak tidak ditemukan pada data ini.	Perlu	Perlu
Kebutuhan Fisiologis	1. Jumlah Gaji 2. Jumlah Bonus	Semua data pada <i>dataset</i> ini diambil dari aplikasi HRMIS sehingga seluruh data karyawan yang berkaitan dengan Gaji dan Bonus ditemukan. Data gaji dan bonus variasi data sangat banyak, perlu dilakukan transformasi untuk menyederhanakan agar dapat dilakukan klasifikasi.	Tidak	Perlu

4.2 Penyiapan Data

Pada proses penyiapan data, dilakukan pengumpulan data dari berbagai sumber seperti yang telah dijelaskan di pembahasan Bab III. Data-data tersebut dikumpulkan lalu

dikelompokkan menjadi 6 *dataset* berdasarkan teori hierarki kebutuhan Maslow. Enam *dataset* tersebut antara lain *dataset* demografi, *dataset* hierarki kebutuhan Maslow fisiologis, *dataset* hierarki kebutuhan Maslow rasa aman dan ketenangan, *dataset* hierarki kebutuhan Maslow hubungan sosial dan cinta, *dataset* hierarki kebutuhan Maslow penghargaan dan pengakuan serta yang terakhir yaitu *dataset* hierarki kebutuhan Maslow aktualisasi diri. Tahapan penyiapan data ini terdiri dari fitur ekstraksi dan *select* dari sumber-sumber data, kemudian dilanjutkan proses *data integration*, *data reduction*, *data cleaning* dan *data transformation*. Proses *data reduction* dan *data cleaning* dilakukan secara bersamaan dan gantian, sedangkan untuk *data integration* merupakan proses awal yang dibutuhkan sebelum proses *data reduction* untuk menjamin integrasi data karyawan, *Data transformation* dilakukan setelah tidak ditemukan *noise* data dan nilai yang hilang dari satu atau beberapa karyawan, data sudah bersih. Semua proses tersebut diselaraskan dengan pemahaman bisnis dan data yang berkaitan dengan penerapan teori hierarki kebutuhan Maslow di Neuron.

4.2.1 Ekstraksi Data

Setelah dipahami data-data yang digunakan dalam pemodelan maka berdasarkan sumber-sumber data yang dicatat dilakukan ekstraksi data-data sesuai pada pembahasan Bab III, dan Sub Bab 4.1. Pada tahap ini seluruh data hasil ekstraksi akan dilakukan penyimpanan dalam bentuk data tabel dalam format *.xlsx (excel workbook)*.

Dataset pertama yang akan diekstraksi dan dikumpulkan adalah data demografi karyawan. Data demografi ini dapat diekstraksi dari HRMIS Neuron. Untuk mendapatkan data demografi karyawan maka melibatkan 7 tabel *database* antara lain tabel *employee*, *person*, *emp_education*, *emp_family*, *address*, *contact_relation*, dan *contact_detail*. Dari 7 tabel *database* tersebut terdapat 62 atribut, maka perlu dilakukan seleksi atribut-atribut yang dibutuhkan saja berdasarkan pemahaman data di Bab III dan sub bab 4.1.2 yaitu data demografi terdiri dari NIK untuk mengidentifikasi personil karyawan, umur, tingkat pendidikan, jenis kelamin, status menikah, jumlah anak, dan asal sekolah. Setelah

dilakukan seleksi dan ekstraksi maka telah dikumpulkan data demografi dengan data seperti ditunjukkan pada Tabel 4.3.

Tabel 4.3 *Dataset* demografi

No	NIK	Umur	Tingkat Pendidikan	...	Jumlah Anak
1	18008020012016	36	S1	...	2
2	18008020012017	37	S1	...	2
3	18008020012018	38	S1	...	3
4	18008020012019	39	S1	...	3
5	18008020012020	40	S1	...	3
6	18008020012021	41	S1	...	3
7	18008020022016	36	D3	...	1
8	18008020022017	37	D3	...	2
9	18008020022018	38	D3	...	2
10	18008020022019	39	D3	...	2
...
558	09920121532021	22	SMA Sederajat	...	0
559	09821021542021	34	S1	...	0
560	09621021552021	23	SMA Sederajat	...	0
561	09621021552021	25	S1	...	0
562	9721031562021	24	S2	...	0
563	00021031572021	21	S1	...	0
564	00021031572021	21	SMA Sederajat	...	0

Sumber: HRMIS Neuron

Dataset kedua yaitu data berdasarkan relasi dengan tingkat teori hierarki kebutuhan Maslow yang dalam hal ini adalah hierarki yang paling bawah yaitu kebutuhan fisiologis. Sesuai pembahasan di Bab III maka sumber data dari *dataset* ini adalah HRMIS modul *payroll*. Modul *payroll* mempunyai 14 atribut yang harus diseleksi untuk mendapatkan data yang hanya diperlukan. Dari 14 atribut tersebut didapatkan 3 atribut yang akan diekstraksi antara lain NIK, gaji rata-rata dengan melakukan agregasi dan jumlah bonus tahunan juga didapatkan dari proses *summary* data dalam 1 tahun. Maka setelah proses seleksi dan ekstraksi tersebut didapatkan data seperti pada Tabel 4.4.

Tabel 4.4 *Dataset* hierarki Maslow kebutuhan fisiologis

No	NIK	Gaji Rata-Rata	Jumlah Bonus Tahunan
1	18008020012016	10198492	67903055,92
2	18008020012017	11972294	37813049,73
3	18008020012018	12863942	62164777,1
4	18008020012019	20442796	76384891,95
5	18008020012020	20477227	114829249,2
6	18008020022016	8706390	60960982
7	18008020022017	10412366	33929149,81
8	18008020022018	11291524	54575544,67
9	18008020022019	17439366	62888180,5
10	18008020022020	17619026	96186199,33

No	NIK	Gaji Rata-Rata	Jumlah Bonus Tahunan
...
558	09920121532021	4324539,6	0
559	09821021542021	6102789,75	0
560	09621021552021	4587252,59	0
561	09621021552021	4389100,59	0
562	9721031562021	4487411,53	0
563	00021031572021	3928789,5	0
564	00021031572021	4280456	0

Sumber: Modul *Payroll* HRMIS Neuron

Dataset ketiga adalah sekumpulan data yang mencerminkan kesesuaian dengan kebutuhan pada teori hierarki kebutuhan Maslow pada tingkat rasa aman dan ketenangan. Data-data ini dapat diekstraksi dari beberapa sumber antara lain HRMIS modul profil, HRMIS modul *attendance*, dan laporan dana pensiun tambahan yaitu Hari Tua Eksklusif (HTE). Data HTE merupakan data yang didapatkan dari laporan dengan bentuk *file excel* dan terpisah dari aplikasi HRMIS. Untuk itu proses ekstraksi ini dibagi menjadi 2 tahap yaitu pertama perlu mengekstraksi data dari HRMIS selanjutnya mengkonversi data laporan dana pensiun HTE ke dalam format sesuai dengan format *dataset* dan menggabungkannya dengan data hasil ekstraksi dari HRMIS. Dari beberapa atribut yang ada pada modul profil dan *attendance*, atribut yang dipilih untuk diekstraksi antara lain :

1. Status Pegawai
2. Lama Bekerja
3. Jam Lembur
4. Jumlah Cuti
5. Jumlah Sakit

Sedangkan untuk data perolehan dana pensiun tambahan didapatkan dari laporan bagian HCM tentang laporan perolehan dana HTE yang sudah dalam bentuk *file excel*. Setelah dilakukan ekstraksi terhadap data perolehan dana HTE tersebut, maka data tersebut digabungkan dengan data yang sudah diseleksi pada modul profil dan *attendance* berdasarkan nama dan NIK. Perlu mencocokkan nama dan NIK dikarenakan data primer keduanya berbeda. Hasil dari ekstraksi dan penggabungan data-data tersebut terlihat pada

Tabel 4.5.

Tabel 4.5 *Dataset* hierarki Maslow rasa aman dan ketenangan

No	NIK	Status Pegawai	Lama Bekerja	Jumlah Lembur	...	Dana Pensiun Tambahan
1	18008020012016	Tetap	8	2,83	...	41601230
2	18008020012017	Tetap	9	130,95	...	43010378
3	18008020012018	Tetap	10	121,67	...	60222047
4	18008020012019	Tetap	11	117,18	...	108621741,39
5	18008020012020	Tetap	12	56,55	...	145054371,84
6	18008020022016	Tetap	13	3,1	...	41601230
7	18008020022017	Tetap	8	132,78	...	43010378
8	18008020022018	Tetap	9	100,03	...	60222047
9	18008020022019	Tetap	10	110,33	...	108621741,39
10	18008020022020	Tetap	11	66,65	...	145054371,84
...
558	09920121532021	Kontrak	0	0	...	-
559	09821021542021	Kontrak	0	0	...	-
560	09621021552021	Kontrak	0	0	...	-
561	09621021552021	Kontrak	0	0	...	-
562	9721031562021	Kontrak	0	0	...	-
563	00021031572021	Kontrak	0	0	...	-
564	00021031572021	Kontrak	0	0	...	-

Sumber: HRMIS Neuron

Dataset hierarki Maslow hubungan sosial dan cinta terdiri dari beberapa atribut yang datanya diambil dari beberapa sumber data antara lain data dari media sosial dalam hal ini Instagram dan HRMIS modul *attendance*. Ekstraksi data dari sumber data media sosial menggunakan metode kuisisioner dengan mengirimkan kepada karyawan. Metode ini dilakukan karena sejak 29 Juni 2020, ijin dasar untuk pengaksesan API Instagram dibatasi, dan pada 31 Desember 2020 penggunaan API *basic* dibatasi. Data dari sosial media ini dibutuhkan mengingat data yang di dapatkan dari kegiatan sedikit, maka membutuhkan parameter lain yang menunjukkan bagaimana hubungan karyawan satu dengan lainnya atau hubungan karyawan dengan perusahaan. Sedangkan untuk Evaluasi 360 derajat parameter *teamwork*, data didapatkan dari hasil penilaian oleh tim HCM yang dilakukan pada tengah tahun dan akhir tahun berupa laporan penilaian dengan tipe *file* excel. Atribut jumlah kehadiran dan tingkat kedisiplinan, data diekstrasi dari HRMIS. Setelah masing-masing data didapatkan maka dilakukan penggabungan ke dalam tipe *file* excel dengan format yang sama dengan *dataset* lainnya. Hasil dari ekstrasi *dataset* hierarki Maslow hubungan sosial dan cinta ini terlihat pada Tabel 4.6.

Tabel 4.6 *Dataset* hierarki Maslow hubungan sosial dan cinta

No	NIK	<i>Follow Instagram Neuron</i>	<i>Like Post Neuron</i>	<i>Follower se-Neuron</i>	...	Jumlah Kehadiran
1	18008020012016	Ya	19	19	...	230
2	18008020012017	Ya	9	19	...	217
3	18008020012018	Ya	10	19	...	219
4	18008020012019	Ya	11	19	...	227
5	18008020012020	Ya	12	19	...	237
6	18008020022016				...	240
7	18008020022017				...	234
8	18008020022018				...	228
9	18008020022019				...	235
10	18008020022020				...	259
...
558	09920121532021	Ya	0	0	...	98
559	09821021542021	Ya	0	0	...	76
560	09621021552021	Tidak	0	0	...	76
561	09621021552021	Tidak	0	0	...	68
562	9721031562021	Ya	0	0	...	57
563	00021031572021	Tidak	0	0	...	37
564	00021031572021	Tidak	0	0	...	38

Ekstraksi data-data yang masuk ke dalam *dataset* hierarki kebutuhan Maslow penghargaan atau pengakuan dilakukan terhadap sumber data yang telah didefinisikan yaitu dari HRMIS modul *gamification*, data *assessment* lelang jabatan, laporan penilaian tengah tahun dan akhir tahun, dan data peserta umroh yang diberikan oleh bagian HCM. Setelah dilakukan ekstraksi, maka data-data yang ada digabungkan dan dijadikan satu *dataset* dengan format yang sama. Tabel 4.7 merupakan hasil dari ekstraksi dan penggabungan pengumpulan data-data tersebut.

Tabel 4.7 *Dataset* hierarki Maslow penghargaan atau pengakuan

No	NIK	<i>Monthly Best Employee</i>	Penghargaan Tahunan	Rekomendasi <i>Assessment</i> naik Jabatan	Bonus Umroh
1	18008020012016	0	tidak	belum	sudah
2	18008020012017	0	tidak	belum	sudah
3	18008020012018	0	tidak	belum	sudah
4	18008020012019	0	tidak	belum	sudah
5	18008020012020	0	tidak	belum	sudah
6	18008020022016	0	tidak	belum	sudah
7	18008020022016	0	tidak	belum	-
8	18008020022017	0	tidak	belum	-
9	18008020022018	0	tidak	belum	sudah
10	18008020022019	1	tidak	belum	sudah
...
558	09920121532021	1	tidak	-	-
559	09821021542021	0	tidak	-	-

No	NIK	Monthly Best Employee	Penghargaan Tahunan	Rekomendasi Assessment naik Jabatan	Bonus Umroh
560	09621021552021	0	tidak	-	-
561	09621021552021	0	tidak	-	-
562	9721031562021	0	tidak	-	-
563	00021031572021	0	tidak	-	-
564	00021031572021	0	tidak	-	-

Dataset hierarki Maslow aktualisasi diri didapatkan berdasarkan proses ekstraksi dan pengumpulan data dari laporan tim HCM, baik berkaitan dengan laporan keikutsertaan ekstrakurikuler, pelatihan dan seminar, serta data penerimaan beasiswa sekolah. Data-data tersebut didapatkan secara terpisah dengan format *file* yang berbeda. Setelah mendapatkan data-data tersebut, maka data tersebut disatukan dalam satu buah *dataset* dengan format yang sama yaitu *file* format .xlsx.

Tabel 4.8 *Dataset* hierarki Maslow aktualisasi diri

No	NIK	Keikutsertaan Ekstrakurikuler	Upgrading /Pelatihan	Beasiswa Sekolah
1	18008020012016	2	1 kali	tidak
2	18008020012017	2	1 kali	tidak
3	18008020012018	2	1 kali	tidak
4	18008020012019	2	1 kali	tidak
5	18008020012020	2	1 kali	tidak
6	18008020022016	2	1 kali	tidak
7	18008020022016	2	1 kali	tidak
8	18008020022017	2	1 kali	tidak
9	18008020022018	2	1 kali	tidak
10	18008020022019	2	1 kali	tidak
...
558	09920121532021	3	2 kali	tidak
559	09821021542021	1	3 kali	tidak
560	09621021552021	1	3 kali	tidak
561	09621021552021	1	3 kali	tidak
562	9721031562021	1	3 kali	tidak
563	00021031572021	0	3 kali	tidak
564	00021031572021	0	-	tidak

Setelah data dilakukan ekstraksi untuk setiap *dataset* yang ada, maka langkah selanjutnya dilakukan *data preprocessing* yaitu *data integration*, *data reduction*, *data cleaning* dan *data transformation*. Proses tersebut dilakukan menggunakan Microsoft 365 Pro Plus Excel dan Rapidminer Studio Educational v 9.8. Microsoft 365 Pro Plus Excel digunakan karena hasil data ekstraksi berupa *file format* .xlsx sehingga proses integrasi data

lebih mudah. Selain itu proses data *cleaning* sangat spesifik dan tidak bisa diberlakukan untuk seluruh karyawan, namun tergantung dari karakteristik karyawan tersebut atau kejadian di satu tahun. Kemudian sekaligus mengintegrasikan *dataset* dan dilakukan *data reduction* sesuai pemahaman data serta *data cleaning* menggunakan fitur yang ada. Rapidminer digunakan untuk memastikan apakah proses *data preprocessing* yang telah dilakukan berjalan dengan baik, proses ini bisa sebut *data preprocessing* lanjutan.

4.2.2 Data Integration dan Data Reduction

Hasil proses ekstraksi data-data terlihat bahwa karyawan yang telah *resign* datanya masih tersimpan, baik di sumber data HRMIS atau di catatan laporan secara digital dan laporan berbentuk *hardcopy* bagian HCM. Beberapa baris data karyawan yang telah *resign* untuk data tahun setelah karyawan tersebut *resign* harus dihapus karena beberapa data berkaitan dengan atribut yang ada tidak mempunyai nilai karena tidak ada kegiatan. Pengurangan data secara dimensi horizontal juga dilakukan terhadap atribut NIK karena data tersebut terlalu unik yang tidak memungkinkan untuk diikuti dalam proses pemodelan klasifikasi. Sehingga dari jumlah atribut tiap *dataset* berkurang 1 atribut. Secara kesimpulan pengurangan data dilakukan terhadap data :

1. Atribut NIK
2. Atribut asal sekolah
3. Data karyawan telah *resign* sebelum 2016
4. Data karyawan *resign* untuk baris data tahun-tahun setelah *resign*

Dari penjelasan perlunya pengurangan data secara jumlah khususnya untuk data dengan ciri data karyawan *resign* sebelum 2016 dan data karyawan *resign* untuk tahun setelah *resign* maka untuk menjamin konsistensi dan integrasi data, perlu dilakukan *data integration* terlebih dahulu. Penghapusan data karyawan tersebut harus sesuai dengan data primer, yaitu data yang terdapat pada atribut NIK. Oleh karena penghapusan data juga melibatkan atribut NIK, yang mana seluruh data tersebut akan hilang maka akan berpengaruh hubungan data antara *dataset* satu dengan *dataset* lainnya. *Data Integration* ini dilakukan sebelum dilakukan penghapusan data baik atribut NIK ataupun data karyawan

yang telah *resign* sebelum tahun 2016. *Data Integration* ini dilakukan dengan cara menggabungkan seluruh *dataset-dataset* yang ada sehingga terjamin integrasi datanya dan keterhubungan yang mencerminkan satu kesatuan dengan data primer karyawan, yaitu NIK. Hasil ekstraksi yang telah dilakukan mendapatkan jumlah data karyawan sebanyak 564, kemudian setelah dilakukan pengurangan dengan karakteristik yang telah disebutkan jumlah data menjadi 489. Data ini dapat dilihat pada Lampiran 1. Atribut asal sekolah dilakukan penghapusan karena setelah dilakukan *data cleaning* variasinya masih terlalu tinggi sehingga akan mengurangi tingkat akurasi. Penghapusan data atribut NIK dilakukan pada tahap *data reduction* lanjutan setelah dilakukan *data cleaning* dan *data transformation* selesai.

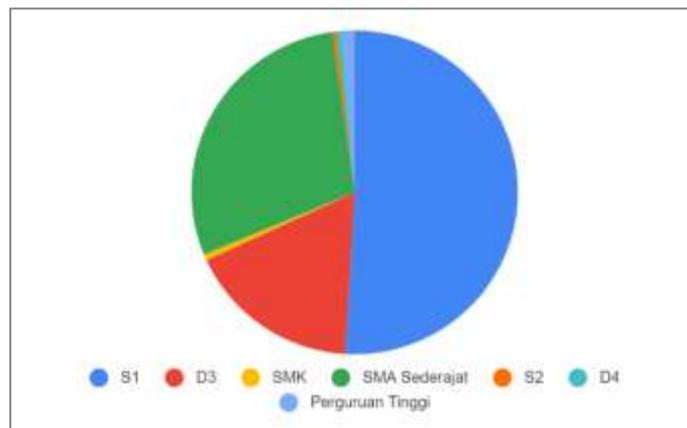
4.2.3 Data Cleaning

Ketika melakukan penyeleksian data yang digunakan dan berdasarkan pemahaman data maka terdapat data-data yang perlu dibersihkan dan dikonsistensikan. Salah satu yang perlu dibersihkan adalah tingkat pendidikan, data ini perlu dikonsistensikan untuk menyamakan antara SMA Sederajat dan SMK. Selain itu juga menyamakan antara pendidikan D4 dengan S1. Data Gaji karena menggunakan rata-rata dan data bonus tahunan karena didapatkan berdasarkan nilai penilaian akhir tahun dimana hasil kedua data tersebut tidak bulat maka perlu dilakukan pembulatan. Data HTE dikarenakan didapatkan dari laporan bagian HCM perihal perolehan HTE dimana yang dicatat di dalam laporan adalah nama-nama karyawan tetap maka beberapa data tidak ditemukan untuk karyawan yang berstatus kontrak. Oleh karena itu data karyawan kontrak untuk atribut dana pensiun tambahan diisi nilainya dengan nilai 0. Selain itu beberapa data yang perlu dibersihkan terdapat pada atribut-atribut berikut:

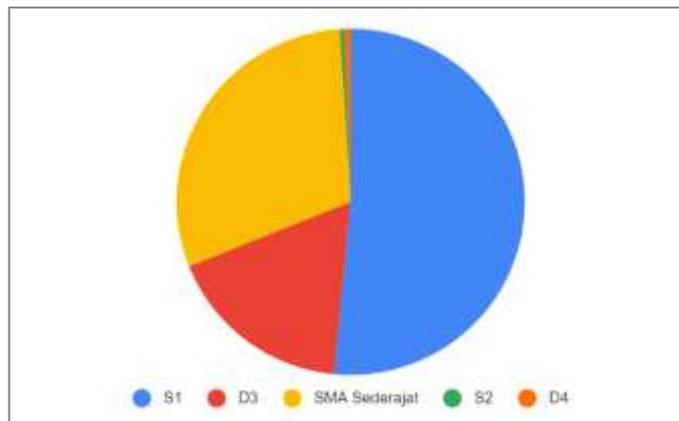
1. Tingkat pendidikan

Data pada atribut ini beberapa mempunyai arti sama namun secara data yang tertulis tidak konsisten, untuk itu perlu dilakukan *data cleaning* dengan menyamakan agar data konsisten. Sebaran data tingkat pendidikan terlihat pada Gambar 4.4. SMK disamakan dengan SMA sederajat karena mempunyai arti yang sama. SMA sederajat

untuk mengakomodir lulusan SMA atau persamaan SMA. Selain itu tercatat nilai perguruan tinggi pada tingkat pendidikan, tidak jelas apakah S2, S1 atau D3. Maka dikonfirmasi ke bagian HCM untuk memperjelas data pada karyawan tersebut. Setelah dilakukan penelusuran dan dilakukan *data cleaning* sebaran data menjadi dibagi 5 kelompok atau kategori tingkat pendidikan dari 7 kelompok sebelumnya seperti Gambar 4.5.



Gambar 4.4
Sebaran data tingkat pendidikan sebelum *data cleaning*

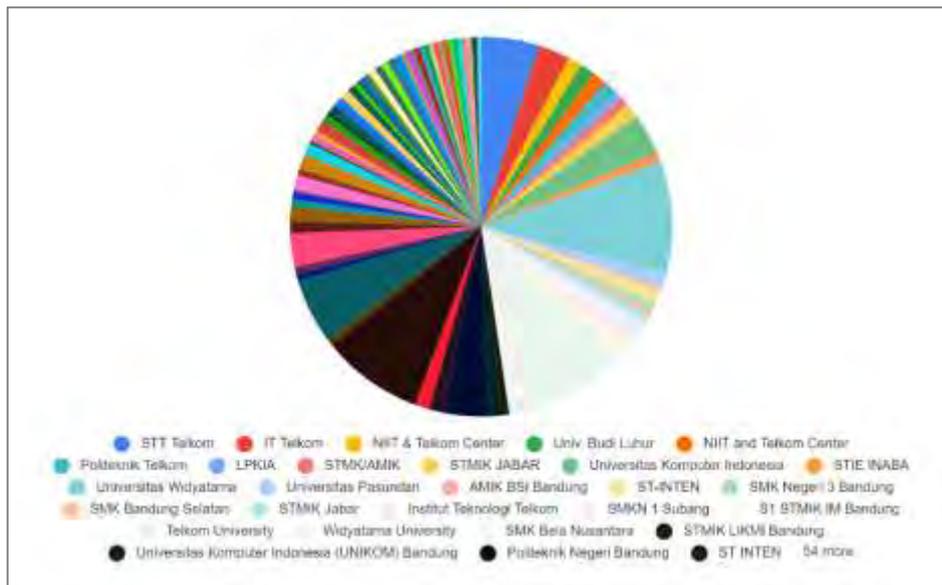


Gambar 4.5
Sebaran data tingkat pendidikan setelah *data cleaning*

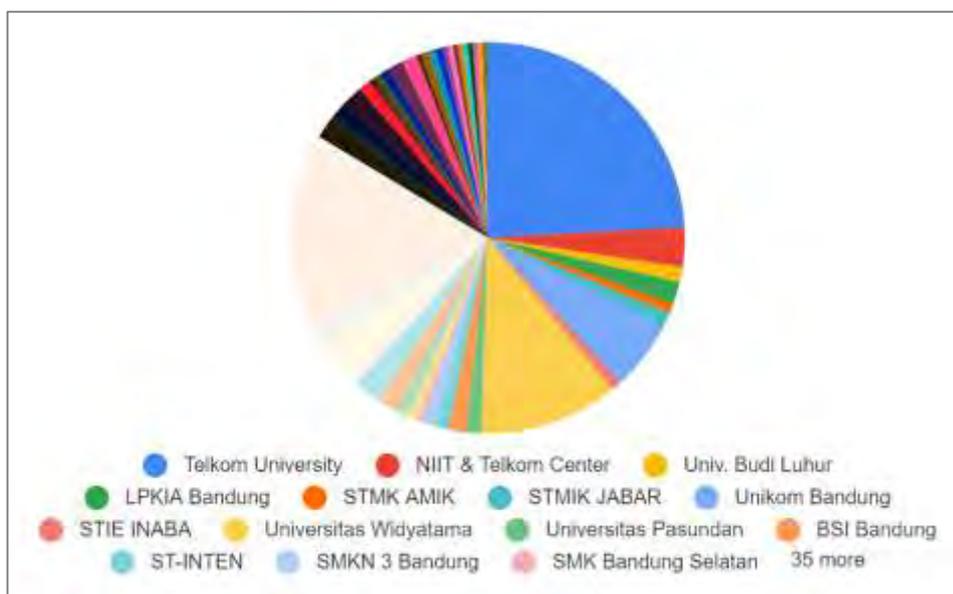
2. Asal sekolah

Data pada atribut ini beberapa mempunyai arti sama, tidak konsisten seperti terlihat pada Gambar 4.6, untuk itu perlu untuk dibuat konsisten. Dengan menggunakan fitur

find and replace pada aplikasi Microsoft Excel, dilakukan penyeragaman data yang mempunyai arti sama. Misal terlihat terdapat data nilai STT Telkom, IT Telkom, Institut Teknologi Telkom, dan University Telkom yang perlu diseragamkan menjadi Telkom University. Setelah dilakukan *data cleaning* data dari jumlah 82 kelompok data atribut asal sekolah menjadi 49 kelompok, ditunjukkan pada Gambar 4.7.



Gambar 4.6
Sebaran data asal sekolah sebelum *data cleaning*



Gambar 4.7
Sebaran data asal sekolah setelah *data cleaning*

3. *Teamwork*

Data pada atribut ini kurang lengkap dimana terdapat nilai *null* untuk data pada tahun 2021 dikarenakan proses penilaian tengah tahun belum selesai dilakukan sehingga tidak diketahui nilainya apalagi nilai akhir tahun. Agar mengurangi jumlah *missing value* pada *dataset* dan agar hasil lebih baik saat pemodelan, maka nilai *null* pada data tahun tersebut diganti oleh nilai pada tahun sebelumnya yaitu pada tahun 2020. Sedangkan untuk karyawan yang baru masuk pada tahun 2021 data diganti dengan nilai rata-rata untuk yaitu 3,91.

4. Bonus tahunan

Data bonus 2021 bernilai *null* untuk seluruh karyawan dikarenakan bonus biasanya diberikan akhir tahun, oleh karena itu perlu diisi nilainya dengan bonus tahun 2020 agar tidak terjadi *missing value*. Sedangkan beberapa data bonus untuk karyawan *resign* yang nilainya *null* maka nilai diganti menjadi nilai bonus pada tahun sebelumnya. Selanjutnya bagi karyawan yang baru masuk di tahun 2021 dan tidak ada data bonus sebelumnya maka sesuai dengan bisnis proses di Neuron bahwa untuk karyawan kontrak diisi nilai bonus 2 kali gaji bulanan.

5. Dana pensiun tambahan

Terdapat data *null* bagi karyawan berstatus kontrak maka *missing value* diganti dengan nilai 0 sedangkan untuk karyawan berstatus tetap di tahun 2021 maka *missing value* diganti dengan nilai sama seperti tahun sebelumnya.

6. *Follow* Instagram Neuron

Beberapa karyawan tidak mempunyai akun Instagram dan beberapa tidak melakukan input kuisisioner Instagram. Oleh karena itu beberapa terdapat nilai *null*. Untuk nilai *null* tersebut di ganti nilainya dengan “tidak”.

7. *Like Post* Neuron

Data ini ada kaitannya dengan atribut *follow* Instagram Neuron. Beberapa nilai pada data di atribut ini bernilai *null*. Maka proses pembersihan data ini disesuaikan dengan nilai di atribut *follow* Instagram Neuron, jika tidak mem-*follow* nilai *null* tersebut diganti dengan nilai 0.

8. *Follower* se-Neuron

Data ini ada kaitannya dengan atribut *follow* Instagram Neuron. Beberapa nilai pada data di atribut ini bernilai *null*. Maka proses pembersihan data ini disesuaikan dengan nilai di atribut *follow* Instagram Neuron sama seperti data *like post* Neuron, jika tidak mem-*follow* maka nilai *null* tersebut diganti dengan nilai 0.

9. Jumlah kehadiran

Data tahun 2021 karena belum penuh 1 tahun maka untuk karyawan yang masuk sebelum 2021 jumlah kehadiran akan disamakan dengan tahun sebelumnya. Sedangkan untuk karyawan yang baru masuk di tahun 2021, kehadiran akan menggunakan jumlah rata-rata ditahun 2021 setelah proses pertama tadi. Untuk karyawan yang mempunyai nilai kurang dari 200, mempunyai arti merupakan karyawan yang baru masuk di tengah tahun atau lebih. Bisa juga berarti mungkin karyawan tersebut *resign* pada waktu kurang dari tengah tahun, maka akan diganti menggunakan jumlah rata-rata pada tahun tersebut.

10. *Upgrading* atau Pelatihan

Data pada atribut ini tidak konsisten karena terdapat data numerik dan nominal. Oleh karena itu perlu dilakukan *data cleaning* merubah yang nominal menjadi numerik yaitu dengan menghapus “ kali” pada data dan merubah sisa data menjadi numerik.

4.2.4 Data Transformation

Hasil proses pemilihan atribut dan ekstrasi memperlihatkan bahwa data-data yang didapatkan dari beberapa sumber data tersebut harus disesuaikan sebelum dilakukan pemodelan. Setelah dilakukan *data cleaning* dan *data reduction* maka perlu dilakukan perubahan bentuk data atau *data transformation*. Perubahan data pada setiap *dataset* menggunakan aplikasi Microsoft 365 Pro Plus Excel, dengan memanfaatkan fitur *find and replace* dan penggunaan *formula*. Salah satu yang perlu dilakukan perubahan bentuk data adalah *dataset* demografi karyawan. Perubahan bentuk data yang dilakukan akan menggunakan pola seperti terlihat pada Tabel 4.9.

Tabel 4.9 Transformasi *dataset* demografi karyawan

Atribut	Nilai	Nilai setelah transformasi
Umur	17 < Umur < 21	0
	20 < Umur < 26	1
	25 < Umur < 31	2
	30 < Umur < 36	3
	35 < Umur < 41	4
	40 < Umur	5
Status Menikah	Sudah Menikah	sudah
	Belum Menikah	belum
Tingkat Pendidikan	SMA sederajat	SMA
	D3	D3
	S1	S1
	S2	S2

Dataset Hierarki Maslow untuk kebutuhan fisiologis dari hasil seleksi dan ekstraksi terlihat jelas perlu dilakukan transformasi. Transformasi yang dilakukan terhadap data gaji yaitu dengan cara mengkategorisasikan gaji menjadi 5 bagian sesuai kelompok tingkatan gaji yang ada di Neuron. Sedangkan untuk data bonus dilakukan pengelompokan menjadi 4 bagian berdasarkan jumlahnya yaitu kelipatan 10 juta dengan kelompok terakhir bonus dengan lebih atau sama dengan 30 juta. Secara terperinci transformasi *dataset* hierarki Maslow Kebutuhan Fisiologis dapat dilihat pada Tabel 4.10.

Tabel 4.10 Transformasi *dataset* hierarki Maslow kebutuhan fisiologis

Atribut	Nilai	Nilai setelah transformasi
Gaji rata-rata	3.500.000 < Gaji < 5.000.000	0
	4.999.999 < Gaji < 7.000.000	1
	6.999.999 < Gaji < 10.000.000	2
	9.999.999 < Gaji < 15.000.000	3
	14.999.999 < Gaji	4
Bonus 1 tahun	Bonus < 10.000.000	0
	9.999.999 < Bonus < 20.000.000	1
	19.999.999 < Bonus < 30.000.000	2
	29.999.999 < Bonus	3

Sedangkan untuk *dataset* hierarki Maslow rasa aman dan ketenangan, perubahan yang dilakukan terhadap data disesuaikan dengan hasil wawancara bagian HCM. Untuk lama bekerja, karyawan yang baru pertama masuk akan ditawarkan kontrak dahulu selama 2 tahun. Setelah 2 tahun selanjutnya akan ditawarkan kontrak sebagai karyawan tetap tergantung dari performansi karyawan tersebut. Untuk itu transformasi pertama adalah melakukan kategorisasi untuk karyawan yang baru bekerja sampai 2 tahun. Selanjutnya

dilakukan pengkategorisasian lama bekerja per 4 tahun. Kategorisasi terakhir untuk atribut lama bekerja adalah sesuai jarak waktu berdirinya perusahaan di tahun 2008, maka lama bekerja yang lebih dari 12 tahun merupakan generasi pertama karyawan di Neuron masuk kategori ini.

Pengelompokan nilai pada atribut jumlah lembur didasarkan pada wawancara dengan bagian HCM yang menyatakan bahwa perlu dibedakan antara sedikit lemburnya atau tidak pernah lembur, yang normal dan terlalu banyak lembur yang tidak baik untuk kesehatan dan ketahanan mental karyawan. Sedikit lembur didefinisikan untuk karyawan yang lembur dari tidak pernah lembur sampai maksimal 1 jam tiap minggunya sehingga kategorisasi pertama dari atribut ini adalah dari tidak pernah lembur sampai 48 jam dalam setahun. Kategorisasi berikutnya meningkat, lebih dari 1 jam dalam seminggu sampai 2 jam tiap minggunya yaitu jika ditotal dalam 1 tahun di 96 jam. Kategorisasi berikutnya meningkat 1 jam tiap minggunya, sampai di kategorisasi terakhir yaitu melakukan lembur lebih dari 5 jam tiap minggunya.

Pengkategorisasian jumlah cuti dalam setahun dibagi menjadi 4 kategori, dari kategori paling sedikit yaitu pengelompokan cuti kurang dari 5 hari, dan yang paling banyak melebihi batas cuti 12 hari dalam 1 tahun. Atribut sakit memiliki karakteristik seperti jumlah cuti, karena dihitung dalam bentuk satuan hari. Namun berbeda dengan cuti yang terdapat batas cuti sesuai peraturan pemerintah, untuk sakit tidak ada batasnya maka untuk sakit kategorisasi atributnya dibuat per kelipatan 6 hari dengan kategorisasi yang paling banyak adalah lebih dari 24 hari dalam satu tahun.

Tabel 4.11 Transformasi *dataset* hierarki Maslow rasa aman dan ketenangan

Atribut	Nilai	Nilai setelah transformasi
Lama Bekerja	0 < Lama < 3	0
	2 < Lama < 6	1
	5 < Lama < 9	2
	8 < Lama < 13	3
	12 < Lama	4
Jumlah Lembur	Jumlah < 49	0
	48 < Jumlah < 97	1
	96 < Jumlah < 145	2
	144 < Jumlah < 193	3
	192 < Jumlah < 241	4
	240 < jumlah	5
Jumlah Cuti	Jumlah < 5	0

Atribut	Nilai	Nilai setelah transformasi
	4 < Jumlah < 9	1
	8 < Jumlah < 13	2
	12 < Jumlah	3
Jumlah Sakit	Jumlah < 7	0
	6 < Jumlah < 13	1
	12 < Jumlah < 19	2
	18 < Jumlah < 25	3
	24 < Jumlah	4
Dana Pensiun	Jumlah < 10.000.000	0
	9.999.999 < Jumlah < 30.000.000	1
	29.999.999 < Jumlah < 50.000.000	2
	49.999.999 < Jumlah < 100.000.000	3
	99.999.999 < Jumlah	4

Dataset hierarki Maslow hubungan sosial dan cinta ditransformasikan tergantung dari masing-masing atribut seperti terlihat pada Tabel 4.12. Untuk atribut *follow* Instagram, agar membuat algoritma bekerja lebih baik maka dirubah menjadi nilai 1 dan 0. Sedangkan untuk *like post* Neuron nilai terendah jika setidaknya pernah *me-like post* Neuron di media sosial kurang dari 6 kali dengan dikonversi nilainya mejadi 0. Ini untuk menunjukkan keterikatan dengan *post* Neuron yang sedikit. Kemudian rentang nilai tingkat berikutnya untuk nilai transformasi kelipatan 10 kali sampai tingkat akhir yaitu dikonversi menjadi nilai 4 untuk jumlah *like post* lebih dari 35 kali, untuk menunjukkan ikatan yang kuat dengan perusahaan. Atribut *follower* se-Neuron menunjukkan seberapa dekat antar anggota tim di Neuron. Data atribut tersebut dikelompokkan berdasarkan seberapa besar seseorang dikenal dan rekan kerja di Neuron mau mengikuti kemudian menjadi *follower* karyawan tersebut. Jumlah dari 0 sampai 2 merupakan nilai terkecil, dimana karyawan tersebut hanya mengenal yang setiap hari sering ketemu dan satu tim tapi dengan jumlah sangat kecil. Untuk diketahui 1 tim di Neuron biasanya terdiri dari 6 sampai 8 orang. Oleh karena itu jumlah 0 sampai 2 merupakan nilai terkecil dan merupakan tingkat paling bawah. Tingkat selanjutnya adalah hanya mengenal dalam satu tim saja, oleh karenanya rentang nilai dari *follower* berkisar 3 sampai 9. Sedangkan seorang karyawan akan dianggap populer jika *follower*-nya sama dengan 30 atau lebih. Ini artinya karyawan tersebut telah dikenal oleh karyawan lintas divisi.

Atribut *teamwork* mempunyai skala dari 1 sampai 5, dengan nilai tertinggi 5, namun dikarenakan nilai didapatkan dari penilaian 360 derajat maka nilai rata-rata yang diperoleh seorang karyawan terdapat bilangan desimalnya. Oleh karena itu pada atribut ini, datanya harus dibulatkan agar algoritma pemodelan berjalan dengan baik. Atribut jumlah kehadiran mempunyai nilai rata-rata tiap tahun, dimana nilai rata-rata paling rendah adalah di tahun 2018 dengan nilai rata-rata jumlah kehadiran 225. Sedangkan nilai rata-rata paling tinggi di tahun 2020 dengan nilai 251. Nilai rata-rata ini digunakan untuk acuan apakah karyawan tersebut hadir di bawah rata-rata, masuk rentang rata-rata atau di atas rata-rata. Masuk dalam rentang rata-rata didefinisikan nilai 225 sampai 235. Di bawah 225 merupakan nilai di bawah rata-rata sedangkan di atas 235 dianggap di atas rata-rata. Di bawah rata-rata disamakan dengan nilai 0, dalam rentang rata-rata disamakan dengan nilai 1, dan di atas rata-rata disamakan dengan nilai 2.

Tabel 4.12 Transformasi *dataset* hierarki Maslow hubungan sosial dan cinta

Atribut	Nilai	Nilai setelah transformasi
<i>Follow Instagram</i>	ok	1
	tidak	0
<i>Like Post Neuron</i>	Jumlah < 6	0
	5 < Jumlah < 16	1
	15 < Jumlah < 26	2
	25 < Jumlah < 36	3
	35 < Jumlah	4
<i>Follower se-Neuron</i>	Jumlah < 3	0
	2 < Jumlah < 10	1
	9 < Jumlah < 20	2
	19 < Jumlah < 30	3
	29 < Jumlah	4
<i>Teamwork</i>	0,99 < nilai < 2	1
	1,99 < nilai < 3	2
	2,99 < nilai < 4	3
	3,99 < nilai < 4,5	4
	4,49 < nilai	5
Jumlah Kehadiran	Jumlah < 225	0
	224 < Jumlah < 236	1
	235 < Jumlah	2

Dataset hierarki Maslow penghargaan dan pengakuan mempunyai 4 atribut. Dari 4 atribut tersebut hanya atribut bonus umroh yang tidak perlu dilakukan transformasi karena datanya terdiri dari 2 jenis yaitu sudah dan belum. Sedangkan untuk atribut lainnya perlu dilakukan transformasi. Atribut *monthly best employee*, nilai data bisa dari 0 sampai 12,

melihat jumlah bulan dalam 1 tahun. Oleh karena itu, perlu dibuatkan kategorisasi agar proses pemodelan lebih efektif dan memiliki akurasi yang baik.

Tabel 4.13 Transformasi *dataset* hierarki Maslow penghargaan dan pengakuan

Atribut	Nilai	Nilai setelah transformasi
Monthly Best Employee	0	tidak
	0 < Jumlah < 4	sedikit
	3 < Jumlah < 7	jarang
	6 < Jumlah < 10	sering
	10 < Jumlah	selalu
Penghargaan Tahunan	0	Tidak
	0 < Jumlah	Ya
Rekomendasi Assessment Naik Jabatan	0	Tidak
	0 < Jumlah	Ya

Pengkategorisasian atribut *monthly best employee* dibuat menjadi 5 kategori terdiri dari tidak pernah, sedikit, jarang, sering dan selalu, dimana pembagian tersebut berdasarkan nilai 0 dan kategori selanjutnya kelipatan 3. Untuk penghargaan tahunan, karena dimungkinkan 1 karyawan bisa mendapatkan lebih dari 1 penghargaan namun kejadiannya langka maka atribut ini hanya dibagi menjadi 2 kelompok yaitu kelompok yang tidak mendapatkan penghargaan dengan nilai “tidak” dan kelompok yang mendapatkan penghargaan dengan nilai “ya”. Rekomendasi assessment untuk kenaikan jabatan juga memungkinkan karyawan mendapatkan 2 rekomendasi dalam 1 tahun, oleh karena itu atribut ini diperlakukan sama seperti atribut penghargaan tahunan.

Tabel 4.14 Transformasi *dataset* hierarki Maslow aktualisasi diri

Atribut	Nilai	Nilai setelah transformasi
Keikutsertaan Ekstrakurikuler	0	Tidak
	0 < Keikutsertaan < 3	Kurang
	2 < Keikutsertaan < 6	Cukup
	5 < Keikutsertaan < 10	Aktif
	9 < Keikutsertaan	Sangat Aktif
Upgrading/Pelatihan	0	Tidak
	0 < Keikutsertaan < 3	Cukup
	2 < Keikutsertaan	Baik

Transformasi *dataset* hierarki Maslow aktualisasi diri dapat dilihat seperti pada Tabel 4.14. *Dataset* hierarki Maslow untuk tingkat aktualisasi diri, perlu dilakukan transformasi karena nilai dari atribut dalam *dataset* tersebut bernilai kuantitas kecuali atribut beasiswa.

Oleh karena itu, untuk keikutsertaan ekstrakurikuler dan *upgrading*/pelatihan perlu dikelompokkan menjadi beberapa bagian sesuai nilai kuantitasnya.

Hasil dari transformasi terlihat di lampiran 1, dengan data berjumlah 489 dengan 26 atribut dan seluruh data sudah dilakukan transformasi berdasarkan aturan-aturan yang telah dijelaskan sebelumnya. Data hasil dari transformasi ini yang akan digunakan pada tahap pemodelan menggunakan Rapidminer. Sebelum melakukan pemodelan dengan Rapidminer dilakukan pengecekan data menggunakan fitur di Rapidminer, untuk melakukan pengecekan apakah proses *data preprocessing* sebelumnya hasilnya cukup baik untuk pemodelan.

4.2.5 *Preprocessing* Lanjutan

Pengecekan *missing value dataset* menggunakan Rapidminer untuk memastikan bahwa data yang digunakan dalam pemodelan benar-benar *dataset* yang memenuhi syarat untuk pemodelan. Tujuan penggunaan Rapidminer selain melakukan pengecekan dengan mudah karena tampilan visual yang baik adalah jika ditemukan *missing value* atau data memiliki *noise* maka dilakukan otomatisasi *cleaning* menggunakan Rapidminer. Hasil dari pengecekan tersebut dapat dilihat pada Gambar 4.8. Hasil lebih detail per atribut secara statistik menggunakan Rapidminer untuk nilai rata-rata, nilai terkecil dan jumlah tiap kelompok data per atribut dapat ditemukan pada Lampiran 2.

Pada lampiran 2 terlihat bahwa tidak ditemukan *missing value* pada atribut-atribut di *dataset*. Hal tersebut membuktikan proses *data cleaning* dan *data reduction* menggunakan Microsoft Excel telah berhasil menghilangkan *missing value*. Hasil secara detail per atribut dapat dilihat di Lampiran 2 untuk melihat apakah terdapat *noisy data* pada tiap atribut. Kesimpulan dari penggunaan fitur Rapidminer tersebut yaitu *dataset* yang telah disiapkan tidak terdapat *noise data*. Dari hasil tersebut, *dataset* tersebut dapat digunakan untuk pemodelan sehingga tidak perlu dilakukan proses *preprocessing* menggunakan Rapidminer.

Name	Type	Missing	Statistics	Title (25 / 25 attributes)	Summary
umur	Integer	0	0	5	1.634
gender	Polynomial	0	F (98)	M (300)	M (300), F (98)
status	Polynomial	0	menikah (214)	belum (275)	belum (275), menikah (214)
anak	Integer	0	0	4	0.455
pendidikan	Polynomial	0	S2 (3)	S1 (252)	S1 (252), SMA (143), ... (4)
gaji	Integer	0	0	4	0.882
bonus	Integer	0	0	3	1.127

Gambar 4.8 Hasil *import dataset* dan pengecekan statistik *missing values*

4.3 Pemodelan

Pada tahap pemodelan, data yang telah dilakukan *data cleaning*, *data reduction*, *data integration* dan *data transformation* dilakukan pemodelan menggunakan Rapidminer. Pemodelan dilakukan menggunakan metode klasifikasi dan algoritma pohon keputusan yaitu C4.5. Proses yang dilakukan pada Rapidminer terlihat seperti pada gambar. Sebelum dilakukan proses pemodelan menggunakan algoritma C4.5, kita lakukan pemilihan label pada data yang telah di-*import* ke Rapidminer. Karena tujuan dari *data mining* ini adalah untuk mengetahui pola model karyawan yang *resign* maka dipilih atribut *resign* untuk diset sebagai label. Proses melakukan set label terlihat pada Gambar 4.9. Kemudian operator *set role* dihubungkan dengan proses model menggunakan operator *decision tree*. Algoritma yang dipilih adalah C4.5, oleh karena itu kriteria yang digunakan adalah *gain ratio*. *Maximal depth* diisi dengan nilai 25, karena mengingat atribut yang digunakan pada *dataset* sejumlah 25. Tangkapan layer proses pengoperasian Rapidminer terlihat pada Gambar 4.10.



Gambar 4.9
Proses pemodelan penentuan target *role* pada atribut *resign*

Pemodelan harus mendapatkan tingkat akurasi yang terbaik agar pengetahuan yang dihasilkan dipercaya dan tepat guna. Oleh karena itu pemodelan dicoba dengan 3 tipe percobaan menggunakan algoritma C4.5 dengan berbeda-beda parameter yaitu:

1. Tanpa parameter *pruning* dan *prepruning*
2. Hanya menggunakan parameter *pruning*
3. Menggunakan parameter *pruning* dan *prepruning*



Gambar 4.10
Proses pemodelan menggunakan C4.5

Pemodelan pertama tidak mengaktifkan *pruning* dan *prepruning* dengan cara tidak memilihnya di bagian *parameters* terlihat seperti pada Gambar 4.10. Setelah dilakukan penyetelan parameter sesuai skenario percobaan maka desain proses pemodelan dijalankan dengan menekan tombol *run*. Tangkapan layar proses lainnya yaitu percobaan memakai parameter *pruning* dan *prepruning* terdapat pada bagian lampiran 3. Ketiga percobaan tersebut menghasilkan model yang berbeda berturut-turut seperti ditunjukkan pada gambar 4.11, 4.12 dan 4.13.



Gambar 4.11
Hasil pemodelan C4.5 tanpa *pruning* dan *prepruning*

Hasil pemodelan secara visual berbentuk *tree* menunjukkan atribut yang paling atas yang merupakan paling utama adalah atribut *upgrading*. Atribut ini berkaitan dengan laporan keikutsertaan kegiatan *upgrading* dan *pelatihan* pada *dataset* kelompok *dataset* teori hierarki kebutuhan Maslow aktualisasi diri. Secara singkat terlihat bahwa model pohon keputusan yang terbentuk menunjukkan bahwa jika karyawan mendapatkan *upgrading* dan pelatihan dengan baik dan sering maka kemungkinan perusahaan mampu untuk menjaga karyawan untuk tidak *resign*. Itu adalah salah satu contoh rekomendasi yang bisa dilakukan dengan cara membaca hasil model pohon keputusan berdasarkan pemrosesan menggunakan algoritma C4.5.



Gambar 4.12
Hasil pemodelan C4.5 dengan *pruning*

Hasil dari pemodelan menunjukkan perbedaan dengan ketiga percobaan tersebut, perbedaan yang paling mencolok adalah ketika menggunakan parameter *pruning* dan *prepruning*. Hasil pemodelan menggunakan parameter *pruning* dan *prepruning* terlihat pada gambar 4.13 dimana pohon keputusan terlihat lebih ringkas, jumlah dahan dan cabang terlihat lebih sedikit. Terlihat pada gambar 4.13 jumlah atribut yang muncul di model hanya 11 dari 25 atribut berdasarkan dataset yang berhasil di-*import* ke Rapidminer untuk dilakukan pemodelan. Hal tersebut bisa terjadi karena ketika parameter *prepruning* dipilih maka ketika algoritma C4.5 dibuat seketika itu juga dilakukan pemangkasan pada simpul daun yang tidak mempunyai kekuatan dalam hal keputusan untuk prediksi sehingga hasil yang diharapkan lebih umum untuk meningkatkan tingkat akurasi prediksi. Detail hasil pemodelan pohon keputusan untuk ketiga percobaan secara deskripsi terdapat pada Lampiran 4. Evaluasi sangat diperlukan untuk melihat model mana yang akan digunakan karena hasil pemodelan berbeda. Penentuan penggunaan model dilihat dari tingkat akurasi dan klasifikasi *error*-nya sehingga dapat rekomendasi yang diberikan akan lebih presisi dan tepat.



Gambar 4.13
Hasil pemodelan C4.5 dengan *pruning* dan *prepruning*

4.4 Evaluasi

Pada metode *data mining* menggunakan CRISP-DM dan sesuai pada metode penelitian ini, tahap selanjutnya setelah pemodelan adalah evaluasi. Pada tahap evaluasi ini diukur sejauh mana tingkat keberhasilan dari pemodelan yang telah dilakukan. Rapidminer digunakan untuk melakukan proses evaluasi sama seperti pada tahap pemodelan. Pada Rapidminer terdapat operator *cross validation* yang bisa digunakan untuk melakukan pengujian performansi masing-masing model yang sesuai pada tahap pemodelan di sub bab 4.3. *Cross validation* akan melakukan operasi validasi silang antara proses pada *data training* dan *data testing* kemudian mencari kinerja terbaik dengan melakukan pengulangan. *Cross validation* menggunakan *folds* untuk melakukan pengulangan tersebut, *folds* merupakan pembagian *dataset* yang digunakan untuk melakukan analisa kinerja terbaik dengan melipatgandakan dari hasil pembagian tersebut secara bergantian dan berurutan. Proses pembagian *dataset* untuk dibagi menjadi *data training* dan *data testing* diserahkan pada algoritma *cross validation* dan dilakukan secara otomatis.

4.4.1 Penggunaan Cross Validation



Gambar 4.14
Evaluasi pemodelan menggunakan *Cross Validation*

Proses melakukan *cross validation* seperti ditunjukkan pada gambar 4.14. Proses penggunaan *cross validation* dilakukan setelah melakukan *set role* pada atribut dengan cara meletakkan operator *cross validation* yang dipilih pada area desain proses dan menghubungkan hasil *dataset* setelah dilakukan *set role* dengan operator tersebut. Di dalam operator *cross validation* terdapat beberapa proses operator yang dapat didefinisikan sesuai dengan maksud evaluasi dari pemodelan tertentu. Untuk memasukkan proses pemodelan yang dilakukan pengukuran performansi maka dilakukan *double click* pada operator *cross validation* kemudian Rapidminer akan menunjukkan lokasi desain proses di dalam *cross validation* yang bisa dilakukan. Penambahkan operator *decision tree* pada lokasi desain proses dan menkonfigurasi agar sesuai dengan algoritma C4.5 yaitu menggunakan *gain ratio* pada bagian parameter kriteria. Sama seperti ketika membuat pemodelan, kedalaman maksimal untuk model dibuat 25 sesuai jumlah atribut yang digunakan pada *dataset* agar mendapatkan hasil yang optimal.

Konfigurasi parameter pada operator *decision tree* dalam detail proses *cross validation* disesuaikan dengan melakukan percobaan sebanyak 3 kali sesuai proses pemodelan yang telah dijelaskan pada tahap pemodelan yaitu tanpa *pruning* dan *prepruning*, menggunakan *pruning* saja dan menggunakan sekaligus *pruning* dan

prepruning. Hal ini dilakukan untuk mengukur tingkat performansi masing-masing model yang telah dihasilkan pada tahap pemodelan. Tangkapan layar proses pembuatan desain evaluasi menggunakan *cross validation* secara detail di dalam prosesnya terlihat pada gambar 4.15.



Gambar 4.15
Detail proses evaluasi di dalam *Cross Validation*

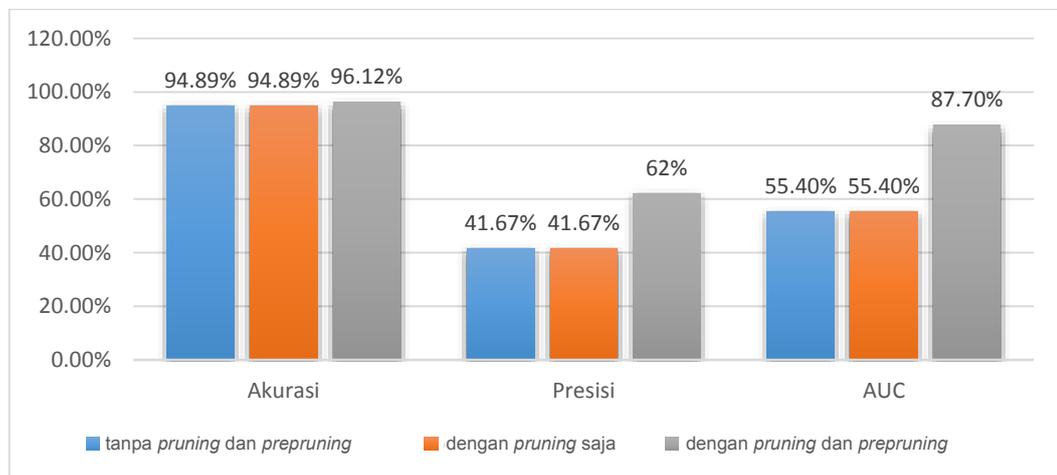
4.4.2 Hasil Pengukuran Performansi

Hasil dari proses pengukuran performansi model menggunakan *cross validation* dibagi menjadi 3 laporan hasil sesuai percobaan pemodelan yang menggunakan pohon keputusan algoritma C4.5. Dari hasil pengukuran performansi 3 model yang telah dilakukan percobaan sesuai tahap pemodelan dipilih 1 model terbaik yang dijadikan rekomendasi pengetahuan. Hasil pengukuran performansi tersebut digambarkan pada Tabel 4.15.

Tabel 4.15 Hasil performansi pohon keputusan algoritma C4.5

	Tanpa <i>Pruning</i> dan <i>Prepruning</i>	Dengan <i>Pruning</i>	Dengan <i>Pruning</i> dan <i>Prepruning</i>
Akurasi	94,89%	94,89%	96,12%
<i>Error</i> Klasifikasi	5,11%	5,11%	3,88%
Presisi	41,67%	41,67%	62%
<i>Relative Error</i>	5,28%	5,28%	4,84%
<i>AUC</i>	55,4%	55,4%	87,7%

Tabel 4.15 memberikan informasi bahwa akurasi terbaik adalah pemodelan menggunakan algoritma C4.5 dengan parameter *pruning* dan *prepruning*. Akurasi menggunakan parameter tersebut mencapai 96,12%, hanya kurang 3,88% dari nilai sempurna. Selisih tersebut merupakan nilai kesalahan klasifikasi. Tersajikan juga ternyata hasil akurasi tanpa *pruning* dan *prepruning* dibanding dengan penggunaan parameter *pruning* mempunyai nilai yang sama. Jika dilihat dari hasil pemodelan, model keduanya memang mirip yaitu sama-sama lebih panjang dan bercabang dibandingkan dengan penggunaan parameter *pruning* dan *prepruning*. Sehingga secara performansi dapat digolongkan menjadi 2 golongan nilai performansi yaitu nilai performansi model berdasarkan algoritma C4.5 menggunakan parameter *pruning* dan *prepruning* kemudian sisi lainnya yang tanpa salah satu parameter tersebut atau tanpa keduanya. Perbedaan kedua golongan tersebut di tingkat akurasi adalah sebesar 1,23%. Sedangkan secara *error* relatif kedua golongan tersebut berbeda sebesar 0,44%. Penggunaan parameter sekaligus, *pruning* dan *prepruning* mempunyai *error* relatif lebih sedikit yaitu sebesar 4,84% berbanding dengan jika tidak menggunakan salah satu atau kedua parameter tersebut yaitu 5,28%.



Gambar 4.16
Grafik hasil evaluasi 3 percobaan pemodelan

Perbandingan secara grafik terlihat pada Gambar 4.16, seberapa baik klasifikasi dilakukan berdasarkan parameter AUC (*Area Under Curve*) dimana nilai tertinggi ditunjukkan untuk model yang dihasilkan menggunakan *pruning* dan *prepruning* yaitu

sebesar 87,7%. Nilai tersebut mempunyai arti bahwa klasifikasi yang dihasilkan oleh penggunaan *pruning* dan *prepruning* bernilai baik karena masuk dalam rentang nilai AUC 80% sampai 90%. Model yang dihasilkan menggunakan *pruning* dan *prepruning* unggul cukup jauh dibandingkan 2 model lainnya. Model lainnya yang menggunakan *pruning* saja kemudian tanpa *pruning* dan *prepruning* berdasarkan nilai AUC masuk dalam kategori klasifikasi yang gagal karena termasuk dalam nilai AUC di bawah 60%.

Rapidminer dalam melakukan pengujian performansi model dari algoritma C4.5 juga dapat menghasilkan hasil *confusion matrix* dengan visual seperti tabel *confusion matrix* pada pembahasan landasan teori Bab II. Tabel hasil pengukuran performansi tersebut ditampilkan dalam bentuk *confusion matrix* seperti pada Tabel 4.16, Tabel 4.17 dan Tabel 4.18 secara berurutan untuk ketiga percobaan yaitu tanpa penggunaan parameter *pruning* dan *prepruning*, hanya menggunakan parameter *prepruning* dan menggunakan kedua parameter *pruning* dan *prepruning*. Dari ketiga tabel tersebut terlihat bahwa penggunaan parameter *pruning* dan *prepruning* memiliki *class precision* dan *class recall* lebih tinggi dibandingkan 2 percobaan lainnya. Detail hasil pengukuran performansi pemodelan secara tangkapan layar dari aplikasi Rapidminer ditampilkan pada lampiran 5.

Tabel 4.16 *Confusion matrix* algoritma C4.5 tanpa *pruning* dan *prepruning*

	True Tidak Resign	True Resign	Class Precision
Pred. Tidak Resign	454	11	97,63%
Pred. Resign	14	10	41,67%
Class Recall	97,01%	47,62%	

Tabel 4.17 *Confusion matrix* algoritma C4.5 hanya dengan *pruning*

	True Tidak Resign	True Resign	Class Precision
Pred. Tidak Resign	454	11	97,63%
Pred. Resign	14	10	41,67%
Class Recall	97,01%	47,62%	

Tabel 4.18 *Confusion matrix* algoritma C4.5 dengan *pruning* dan *prepruning*

	True Tidak Resign	True Resign	Class Precision
Pred. Tidak Resign	458	9	98,07%
Pred. Resign	10	12	54,55%
Class Recall	97,86%	57,14%	

Berdasarkan hasil evaluasi baik tingkat akurasi, tingkat presisi, kesalahan klasifikasi, AUC dan *confusion matrix* disimpulkan bahwa pemodelan menggunakan parameter

pruning dan *prepruning* merupakan pilihan terbaik untuk dijadikan model dan *rule* dalam melakukan rekomendasi untuk membantu pengelolaan retensi karyawan di Neuron. Dengan melakukan pemangkasan simpul daun yang tidak menambah kekuatan sebagai dasar pembuatan keputusan pilihan *resign* maka model dengan parameter *pruning* dan *prepruning* akan lebih memudahkan dalam memilih prioritas tingkat dari teori hierarki kebutuhan Maslow mana yang harus diperbaiki terlebih dahulu untuk menjaga karyawan agar tidak *resign*.

4.4.3 Interpretasi Model Terpilih

Setelah dilakukan evaluasi performansi untuk ketiga percobaan pemodelan menggunakan algoritma pohon keputusan C4.5, dapat disimpulkan berdasarkan hasil evaluasi pada Sub bab 4.4.2 bahwa secara umum algoritma C4.5 dapat menghasilkan model pohon keputusan yang baik untuk mendukung pengambilan keputusan dalam pengelolaan retensi karyawan. Hal ini dikarenakan hasil nilai tingkat akurasi ketiga model tersebut lebih dari 90% yang berarti cukup baik dan dapat dipercaya. Dengan dipilihnya model dengan percobaan pemodelan menggunakan parameter *pruning* dan *prepruning* bukan berarti bahwa kedua model lainnya tidak baik dan kurang tepat karena dengan nilai 94,89% dan kesalahan klasifikasi hanya sebesar 5,11% kedua model lainnya itupun sudah masuk kategori dapat dipercayai. Jika dihitung berdasarkan jumlah karyawan Neuron saat ini berjumlah 112 orang, maka jika menggunakan prediksi kedua model lainnya tersebut kesalahan prediksi *resign* hanya diangka 0,05 orang, tidak sampai 1 karyawan salah dalam prediksi untuk *resign*. Semua model hasil pemodelan dari 3 percobaan menempatkan atribut *upgrading* pada akar pohon keputusan, hal ini berarti bahwa atribut tersebut mempunyai peranan penting untuk mencegah karyawan dalam memutuskan *resign*. Pengetahuan tersebut menjadi penting untuk memprioritaskan tingkat hierarki kebutuhan Maslow mana yang perlu dilaksanakan dan diperbaiki.

Sebagaimana tujuan dari proses *data mining* pada penelitian ini, bahwa pengelolaan retensi karyawan perlu diperbaiki melalui pemrosesan data menggunakan klasifikasi dengan algoritma C4.5 maka keputusan yang tepat dalam mengambil kebijakan melalui

sebuah pengetahuan sangat penting. Oleh karena dalam mengambil kebijakan membutuhkan pertimbangan prioritas dan urgenitas maka model yang bisa menyediakan sisi prioritas dan kecepatan dalam membuat keputusan dalam mencegah karyawan *resign* harus dipilih, sehingga pengelolaan retensi dapat dilakukan secara cepat dan tepat. Hasil model dengan tingkat akurasi tinggi dan kesalahan klasifikasi terkecil yang harus dipilih untuk dijadikan pengetahuan dalam pengelolaan retensi karyawan. Pertimbangan lain yaitu berdasarkan nilai AUC maka yang mempunyai kategori klasifikasi baik, hanya model dengan algoritma C4.5 yang menggunakan parameter *pruning* dan *prepruning* sedangkan model lainnya masuk dalam kategori klasifikasi gagal atau buruk. Maka model yang dihasilkan dengan penggunaan parameter *pruning* dan *prepruning* yang dipilih untuk dijadikan model pengetahuan.

Tabel 4.19 Keterhubungan model pengetahuan dan tingkat hierarki Maslow

Tingkat Simpul	Atribut Model Pohon Keputusan	Deskripsi Atribut	Hierarki Maslow
1	<i>Upgrading</i>	Kegiatan Upgrading dan Pelatihan	Aktualisasi Diri
2	Masuk	Jumlah Kehadiran	Hubungan Sosial dan Cinta
	Lama	Lama Bekerja	Rasa Aman dan Ketenangan
3	Gaji	Rata-rata gaji sebulan	Kebutuhan Fisiologis
4	Ekstrakulikuler	Keikutsertaan kegiatan ekstrakulikuler	Aktualisasi Diri
	Like	Like Posting Neuron	Hubungan Sosial dan Cinta
5	Like	Like Posting Neuron	Hubungan Sosial dan Cinta
6	Bonus	Jumlah Bonus Tahunan	Kebutuhan Fisiologis
7	Lembur	Jumlah Jam Lembur	Rasa Aman dan Ketenangan
8	Gaji	Rata-rata gaji sebulan	Kebutuhan Fisiologis

Setelah memilih model untuk dijadikan rekomendasi pengetahuan maka perlu dihubungkan dengan teori hierarki kebutuhan Maslow agar kegiatan pengelolaan retensi karyawan di Neuron dapat dioptimalkan berdasarkan setiap tingkatan pada hierarki kebutuhan Maslow dan program kerja di Neuron. Model pohon keputusan dengan atribut-atribut sebagai simpul daun dan keterhubungannya dengan tingkat teori hierarki kebutuhan Maslow tergambar pada Tabel 4.19. Keterhubungan tersebut diinterpretasikan berdasarkan keterkaitan program kerja-program kerja yang ada di Neuron dengan teori

hierarki kebutuhan Maslow seperti yang telah dibahas pada Bab III. Bab tersebut menerangkan dengan jelas bahwa Neuron melakukan pengelolaan retensi karyawan menggunakan konsep teori hierarki kebutuhan Maslow. Oleh karena itu dilakukan penghubungan setiap kegiatan dalam program kerja yang ada disesuaikan dengan kebutuhan pada setiap tingkat teori hierarki kebutuhan Maslow dengan atribut-atribut yang menjadi simpul daun pada model klasifikasi hasil dari proses *data mining*.

Penggunaan parameter *pruning* dan *prepruning* pada penerapan pemodelan menggunakan algoritma C4.5 menyebabkan beberapa atribut terpangkas dikarenakan kurang memiliki kekuatan untuk dijadikan dasar karyawan memutuskan *resign* atau tidak. Tabel 4.20 menunjukkan terdapat 1 tingkat dari teori hierarki kebutuhan Maslow yang tidak menjadi dasar karyawan untuk memutuskan *resign* atau tidak. Hal ini menandakan pengelolaan retensi pada kegiatan pemberian penghargaan dan pengakuan kepada karyawan sudah baik dilakukan oleh Neuron. Hal tersebut berarti atribut-atribut yang masuk dalam karakteristik hierarki Maslow penghargaan dan pengakuan tidak memiliki keterkaitan kuat pada karyawan untuk memutuskan *resign* atau tidak. Sebagaimana konsep *pruning* dan *prepruning* karena atribut tersebut tidak termasuk kuat maka untuk meningkatkan akurasi dan presisi, dilakukan pemotongan sehingga untuk kasus data pada Neuron, dengan model terbaik yang dipilih atribut-atribut penghargaan dan pengakuan tidak mempunyai pengaruh.

Sedangkan tingkat dari hierarki kebutuhan Maslow yang prioritas perlu ditingkatkan kualitas kegiatannya adalah pada hierarki aktualisasi diri. Hal ini disebabkan merupakan atribut paling atas yang menjadi akar dari pohon keputusan adalah *upgrading* yang mengacu pada kegiatan *upgrading* dan pelatihan. Kegiatan *upgrading* dan pelatihan harus gencar dilakukan dan merata untuk mencegah karyawan melakukan *resign* karena kebutuhan akan aktualisasi diri terpenuhi. Aturan model yang terbaca pada Lampiran 4 adalah jika karyawan mengikuti *upgrading* dan pelatihan secara baik maka mereka memilih untuk tidak *resign*. Karyawan yang tidak mendapatkan kegiatan *upgrading* dan pelatihan cenderung untuk melakukan *resign* dengan catatan keterhubungan dengan hasil pelaksanaan program kerja lainnya dan syarat tertentu pada kondisi demografi karyawan.

Tabel 4.20 Jumlah kemunculan di model Pohon Keputusan

Hierarki Maslow	Jumlah	Keunikan Atribut	Tingkat Simpul
Aktualisasi Diri	2	2	1,4
Penghargaan dan Pengakuan	0	0	-
Hubungan Sosial dan Cinta	3	2	2,4,5
Rasa Aman dan Ketenangan	2	2	2,7
Kebutuhan Fisiologis	3	2	3,6,8

Berdasarkan kemunculan atribut-atribut *dataset* dan keterhubungannya dengan tingkat hierarki kebutuhan Maslow sebagaimana ditunjukkan pada Tabel 4.20 maka bisa disimpulkan secara urutan yang paling berpengaruh terhadap pendorong keputusan karyawan untuk *resign* adalah aktualisasi diri. Atribut yang berkaitan dengan aktualisasi diri masing-masing menempati simpul atas dari model pohon keputusan. Kemudian secara berurutan berdasarkan paling berpengaruh adalah hubungan sosial dan cinta, rasa aman dan ketenangan kemudian yang terakhir adalah kebutuhan fisiologis. Hal ini hampir sejalan dengan teori hierarki kebutuhan Maslow yang menyatakan bahwa motivasi manusia akan baik dengan memenuhi tingkatan hierarki dari bawah ke atas, jika hierarki bawah telah terpenuhi, maka perlu dipenuhi tingkatan atas selanjutnya. Urutan berpengaruh tersebut menunjukkan juga bahwa mana saja yang perlu dijadikan prioritas perbaikan pengelolaan retensi berdasarkan teori hierarki kebutuhan Maslow di Neuron.

Tabel 4.21 *Rule* hasil algoritma pohon keputusan C4.5

Rule
<p>Jika kegiatan upgrading dan pelatihan tidak dilakukan dengan baik dan merata, maka untuk karyawan yang bekerja sudah lebih dari 5 tahun hal tersebut dapat menyebabkan <i>resign</i>.</p> <p>Sedangkan untuk karyawan yang baru bekerja kurang dari 6 tahun jika tidak ada kegiatan ekstrakurikuler hal tersebut bisa menyebabkan karyawan <i>resign</i>. Maka harus diadakan kegiatan-kegiatan ekstrakurikuler yang mampu menarik minat seluruh karyawan untuk mengikutinya.</p> <p>Jika tidak bisa menyediakan kegiatan ekstrakurikuler yang menarik dan banyak untuk diikuti oleh seluruh karyawan maka harus mengadakan kegiatan yang meningkatkan hubungan sosial, keakraban dan menciptakan kesolidan antar karyawan.</p> <p>Jika kegiatan untuk meningkatkan hubungan sosial antar karyawan kurang maka perlu meningkatkan jumlah bonus tahunan dengan total lebih dari 20 juta per tahun.</p> <p>Jika tidak dapat menaikkan jumlah bonus sesuai nilai tersebut maka perlu untuk mengurangi jumlah jam lembur agar tidak lebih dari 4 jam dalam 1 bulan.</p> <p>Jika tidak bisa mengurangi jumlah jam lembur sesuai pernyataan sebelumnya maka perlu menaikkan gaji karyawan minimal 5 juta per bulannya.</p>

Pengetahuan yang didapatkan dari hasil proses *data mining* yang berupa *rule* bisa dijadikan rekomendasi untuk pengelolaan retensi karyawan berdasarkan teori hierarki kebutuhan Maslow di Neuron adalah seperti ditunjukkan pada Tabel 4.21. Selain *rule* pada tabel tersebut, perlu juga untuk mewaspadaikan dalam pengelolaan retensi karyawan dengan ciri-ciri karyawan berjenis kelamin perempuan. Karyawan berjenis kelamin perempuan perlu dilakukan perlakuan lebih untuk pengelolaannya karena walaupun masuk kategori cukup dalam mendapatkan fasilitas kegiatan *upgrading* dan pelatihan namun jika hubungan sosial dan cinta dalam perusahaan kurang terjalin dengan baik maka ada potensi untuk *resign*. Untuk itu karyawan perempuan perlu diberikan program kerja berupa kegiatan-kegiatan untuk meningkatkan keakraban, kesolidan antar karyawan dan hubungan sosial lainnya. Sedangkan untuk karyawan berjenis kelamin laki-laki dan belum memiliki anak serta memiliki gaji di bawah 5 juta rupiah perlu diberikan kegiatan untuk mengaktualisasikan dirinya seperti *upgrading* dan pelatihan atau kegiatan ekstrakurikuler. Hal tersebut juga berlaku bagi karyawan yang sudah memiliki anak baik karyawan laki-laki atau perempuan. Sedangkan karyawan yang sudah bekerja kurang dari 5 tahun dengan gaji kurang dari 5 juta rupiah perlu dinaikkan gajinya menjadi minimal 5 juta dan difasilitasi kegiatan untuk meningkatkan hubungan sosial dan cinta seperti kegiatan keakraban atau kegiatan ekstrakurikuler dan kegiatan lainnya untuk beraktualisasi diri.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan dan dijelaskan pada Bab-bab sebelumnya, bahwa penelitian untuk membantu pengelolaan retensi karyawan menggunakan klasifikasi dengan algoritma C4.5 dan berbasiskan *dataset* hierarki kebutuhan Maslow ini dapat disimpulkan sebagai berikut :

1. Pembentukan *dataset* hierarki kebutuhan Maslow dilakukan dengan cara mengumpulkan data-data kegiatan dari program kerja Neuron dan data parameter tambahan yang berelasi dengan setiap tingkat pada teori hierarki kebutuhan Maslow. Data-data ini dikumpulkan dari berbagai sumber data seperti HRMIS, laporan dari bagian HCM seperti penilaian tengah dan akhir tahun, keikutsertaan kegiatan ekstrakurikuler, pelatihan, umroh dan data perolehan dana pensiun tambahan yang didapatkan dari bagian keuangan.
2. Pengukuran tingkat akurasi model klasifikasi menggunakan algoritma C4.5 dengan *dataset* berbasiskan teori hierarki kebutuhan Maslow yang telah dilakukan *data preprocessing* berjumlah 489 *records* mendapatkan hasil tertinggi yaitu sebesar 96,12%. Tingkat akurasi ini dihasilkan dengan melakukan *correct prediction* sebanyak 470 *records* dan *wrong prediction* sebanyak 19 *records*.
3. Pemodelan klasifikasi dengan algoritma pohon keputusan C4.5 yang mendapatkan tingkat akurasi tertinggi tersebut menggunakan parameter *pruning* dan *prepruning*. Penggunaan parameter *pruning* dan *prepruning* telah mampu meningkatkan kualitas prediksi dengan mengurangi kesalahan klasifikasi sebesar 1,23% dibandingkan tanpa menggunakan parameter *pruning* dan *prepruning* atau hanya menggunakan *pruning* saja. Selain itu model tersebut juga masuk dalam kategori klasifikasi baik karena memperoleh nilai AUC sebesar 87,8%.
4. Berdasarkan pengukuran nilai performansi dari percobaan penggunaan parameter *pruning* dan *prepruning* didapatkan model pengetahuan yang mempengaruhi keputusan karyawan untuk *resign*. Atribut yang paling berpengaruh terhadap

keputusan *resign* karyawan secara berurutan dari paling penting adalah kegiatan *upgrading* dan pelatihan, kegiatan ekstrakurikuler, kegiatan keakraban dan kebersamaan, jumlah gaji, jumlah bonus dan kegiatan lembur.

5. Model pengetahuan tersebut setelah dihubungkan dengan tingkat hierarki kebutuhan Maslow didapatkan bahwa pengelolaan retensi karyawan berdasarkan teori hierarki kebutuhan Maslow di Neuron perlu diperbaiki. Secara berurutan berdasarkan paling prioritas yang perlu perbaikan meliputi aktualisasi diri, hubungan sosial dan cinta, rasa aman dan ketenangan serta yang terakhir adalah kebutuhan fisiologis. Sedangkan tingkatan hierarki kebutuhan Maslow yang tidak mempunyai pengaruh kuat terhadap keputusan *resign* di Neuron adalah pengakuan dan penghargaan.

5.2 Saran

Beberapa saran yang dapat memberikan manfaat kepada penelitian selanjutnya dan Neuron berdasarkan penelitian dan hasil kesimpulan adalah sebagai berikut :

1. Perlu dilakukan penelitian selanjutnya tentang *data mining* untuk pengelolaan retensi karyawan berdasarkan *dataset* teori hierarki kebutuhan Maslow menggunakan metode klasifikasi dengan algoritma pohon keputusan selain C4.5. Penelitian tersebut perlu membandingkan performansi model yang ditemukan untuk mengetahui algoritma yang lebih baik dari C4.5
2. Tema penelitian lainnya untuk masa akan datang yang perlu dilakukan adalah perancangan aplikasi HRMIS berbasis teori hierarki kebutuhan Maslow untuk membantu pengelolaan retensi karyawan.
3. Neuron perlu memperbaiki pengelolaan retensi karyawan dengan memperbaiki program kerja-program kerjanya sesuai pada tingkatan hierarki kebutuhan Maslow antara lain aktualisasi diri, hubungan sosial dan cinta, rasa aman dan ketenangan serta kebutuhan fisiologis. Perbaikan tersebut perlu dilakukan agar karyawan terbaik mampu dijaga untuk tetap berada di Neuron sehingga membantu meningkatkan performansi Neuron.

4. Proses pengumpulan *dataset* dari sumber data yang berbeda dan beberapa dilakukan secara manual menyebabkan proses pengumpulan tersebut membutuhkan waktu yang lama. Oleh karena itu Neuron dapat menambahkan fitur pada aplikasi HRMIS agar mampu menyimpan data-data yang dibutuhkan untuk mengelola retensi karyawan berbasis teori hierarki kebutuhan Maslow. Hal tersebut akan membantu Neuron untuk melakukan analisa pengelolaan retensi karyawan lebih baik dan lebih cepat.

DAFTAR PUSTAKA

- Adzka S A, 2017, *Analisis faktor yang mempengaruhi retensi karyawan (Studi pada Karyawan PT Bank X cabang Y)*, Skripsi, Fakultas Ekonomika dan Bisnis, Universitas Diponegoro, Semarang.
- Aggarwal C, 2015, *Data Mining : The Textbook*, Springer International Publishing Switzerland, New York, USA
- Almasri M N, 2016, *Manajemen Sumber Daya Manusia : Implementasi Dalam Pendidikan Islam*, *Jurnal Penelitian Sosial Keagamaan*, Vol. 19, No. 2.
- Al-Radaideh, Q A & Al Nagi, E, 2012, *Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance*, *International Journal of Advanced Computer Science and Applications*, Vol. 3, No. 2.
- Angrave D, et.al., 2016, *HR and analytics: why HR is set to fail the big data challenge*, *Journal of Human Resource Management*, Vol. 26, No 1, pp 1-11.
- Bhushan K, 2010, *Business Intelligence and Its Use for Human Resource Management*, *Journal of Human Resource and Adult Learning*, Vol. 6, Num 2.
- Bramer M, 2016, *Principles of Data Mining, 3rd ed*, Springer - Undergraduate Topics In Computer Science, London United Kingdom.
- Changkakati B & Chayanika D, 2020, *Data Mining Techniques In HR Analytics: A Review Of Domain Specific Concepts And Technicalities*, *International Journal of Scientific & Technology Research*, Vol. 9, Issue 3.
- Chapman P, et.al., 2000, *CRISP-DM 1.0 Step-by-Step Data Mining Guide*. Technical report, CRISP-DM.
- El-Habi A M & El-Ghareeb M, 2014, *Evaluation of Data Mining Classification Model*, *IUG Journal of Natural and Engineering Studies*, Vol. 22, No 1, pp 151-165.
- Etukudo R U, 2019, *Strategies for Using Analytics to Improve Human Resource Management*, Walden Dissertations and Doctoral Studies in Walden University.
- Fauzi M S, 2017, *Data Mining Klasifikasi untuk Penyeleksian Calon Karyawan Baru Menggunakan Algoritma ID3*, Universitas Pasundan, Bandung
- Fitz-Enz J & Mattox J R, 2014, *Predictive Analytics for Human Resources*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Ghatak S & Singh S, 2019, *Examining Maslow's Hierarchy Need Theory in the Social Media Adoption*. *FIIB Business Review* 8(4) 292–302.
- Gupitha R, 2018, *Penerapan Klasifikasi Status Pegawai menggunakan Metode Naïve Bayes di RSU H Syaiful Anwar*, *Jurnal Global*, Vol 1.
- Han J, et.al., 2012, *Data Mining Concepts and Techniques*. Morgan Koufmann Publishers, Elsevier, USA.

- Heidarian AR, et al, 2015, *The relationship between demographic characteristics and motivational factors in the employees of social security hospitals in Mazandaran*, *Caspian J Intern Med*; Vol. 6 No3, pp:170-174.
- Jaffar Z, et.al., 2019, *Predictive Human Resource Analytics Using Data Mining Classification Techniques*. *The International Journal of Computer*, Vol. 32, pp: 9–20.
- Jerome N, 2013, *Application of the Maslow's hierarchy of need theory; impacts and implications on organizational culture, human resource and employee's performance*, *International Journal of Business and Management Invention*, Vol. 2 Issue 3, pp : 39-45.
- Kana M, 2020, *How To Find Decision Tree Depth via Cross-Validation*, *Towards Data Science*, dilihat Agustus 2021, <<https://towardsdatascience.com/how-to-find-decision-tree-depth-via-cross-validation-2bf143f0f3d6>>
- Kumara R & Supriyanto C, , *Klasifikasi Data Mining Untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5*, *Skripsi,Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang*.
- Lahida N W, et.al., 2017, *Analyzing Maslow's Hierarchy of Needs, Compensation, and Work Environment on Employee Retention at PT Bank Mandiri Manado Branch Sudirman*, *Jurnal EMBA*, Vol.5 No.3, Hal. 3724-373.
- Larose D T & Larose C D, 2014, *Discovering Knowledge in Data An Introduction to Data Mining (Second Edition)*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Mariscal G, et.al., 2010, *A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies*, *The Knowledge Engineering Review*, Vol. 25:2, pp 137–166.
- Marliani S & Sakapurnama E, 2014, *Pengaruh Employer Attractiveness terhadap Retensi Karyawan Pada Direktorat Sumber Daya Manusia PT Pertamina (Persero) di Kantor Pusat*. *Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Indonesia*.
- Mayasari R, et.al., 2018, *Gambaran Lingkungan Kerja Sosial, Penghargaan Finansial dan Retensi Karyawan Pada PT Baett Mal Abadi di Cilegon*, *Journal of Business Management Education*, Vol. 3, No. 1 pp : 32-41.
- Meena M R & Parimalarani G, 2019, *Human Capital Analytics: A Game Changer for HR Professionals*, *International Journal of Recent Technology and Engineering*, Vol. 8, Issue 2S11.
- Mira M S, et.al., 2019, *The Impact of Human Resource Practices on Employees Performance through Job Satisfaction at Saudi Ports Authority BAsed on the Assumption of Maslow Theory*. *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249 – 8958, Vol. 8, Issue-5C.
- Mishra S N, et.al., 2016, *Human Resource Predictive Analytics (HRPA) For HR Management In Organizations*, *International Journal of Scientific & Technology Research*, Vol 5, Issue 5.
- Mishra S N & Lama DR, 2016, *A Decision Making Model for Human Resource Management in Organizations using Data Mining and Predictive Analytics*, *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 14, No. 5

- Muryjas P & Wawer M., 2014, *Business Intelligence As A Support In Human Resources Strategies Realization in Contemporary Organizations*. Lublin University.
- Mustiko H W, 2012, *Pengaruh Praktik Manajemen Sumber Daya Manusia Terhadap Kinerja SDM pada Meseum Pemerintah Provinsi DKI Jakarta (Museum Sejarah Jakarta, Museum Wayang, Museum Seni Rupa dan Keramik serta Museum Bahari)*, Tesis Kekhususan Administrasi dan Kebijakan Pengembangan Sumber Daya Manusia, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Indonesia.
- Nurhidayati, 2016, *Faktor-faktor yang Berhubungan dengan Retensi Karyawan di Rumah Sakit Prikasih Jakarta Tahun 2016*, Skripsi Sarjana Kesehatan Masyarakat, Universitas Islam Negeri Syarif Hidayatullah.
- Provost F & Fawcett T, 2013, *Data Science for Business*, O'Reilly Media Inc, Gravenstein Highway North, Sebastopol, CA 95472, USA.
- Rachmawati F, 2016, *Analisis Algoritma C4.5 untuk Pengangkatan Karyawan Tetap Studi Kasus PT Citra Abadi Sejati Bogor*, *Jurnal Ilmiah Teknologi dan Informasi*, Vol. 6 No. 2.
- Rapidminer Doc, 2021, *Operator Reference Guide – Cross Validation (Concurrency)*, Rapidminer Studio 9.10, dilihat 10 Agustus 2021, <https://docs.rapidminer.com/latest/studio/operators/validation/cross_validation.html>
- Roos W, 2005, *The Relationship Between Employee Motivation, Job Satisfaction, And Corporate Culture*, Thesis Master Of Science, University of South Africa.
- Ruhyati Y, 2021, *Laporan Keuangan PT Neuronworks Indonesia per 31 Desember 2020 dan Laporan Audit Independen*, Kantor Akuntan Publik Dra Yati Ruhyati, Bandung
- Sari E & Dwiarti R, 2018, *Pendekatan Hierarki Abraham Maslow Pada Prestasi Kerja Karyawan PT Madubaru (PG Madukismo) Yogyakarta*, *Jurnal Perilaku dan Strategi Bisnis*, Vol. 6, No 1, pp 58-77.
- Sa'diyah H, et.al., 2017, *Pengaruh Employee Retention dan Turnover intention terhadap kinerja karyawan melalui kepuasan kerja pada KOSPIN Jasa Cabang Jawa Timur*, *Jurnal Kompetensi Universitas Trunojoyo Madura*, Vol. 11, No 1.
- Scikit-learn developers, 2020, *Cross-validation: evaluating estimator performance*, Scikit Learn, dilihat pada 10 Agustus 2021, <https://scikit-learn.org/stable/modules/cross_validation.html>
- Segara D W B & Sunge A S, 2019, *Penerapan Data Mining dalam Menentukan Karyawan Berprestasi Dengan Metode Algoritma C4.5*, *STT Pelita Bangsa, Teknik Informatika, Jawa Barat*.
- Scott M, et.al., 2005, *The Relationships Between Selected Demographic Factors And The Level Of Job Satisfaction of Extension Agents*, *Journal of Agricultural Education*, Vol 46, No 3, pp. 2-11
- Sikaroudi A E & Ghousi, 2015, *A Data Mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing)*, *Journal of Industrial and Systems Engineering*, Vol. 8, No. 4.

- Singh S & Gupta P, 2014, *Comparative Study ID3, CART and C4.5 Decision Tree Algorithm: A Survey*, *International Journal of Advanced Information Science and Technology (IJAIST)*, Vol. 27, No 27.
- Strohmeier S & Piazza F, 2013, *Domain Driven Data Mining in Human Resource Management : A Review of Current Research*, *An International Journal Expert Systems with Applications*, Vol. 40, pp 2410-2420
- Suyono J & Mudjanarko S W, 2017, *Motivation Engineering to Employee by Employees Abraham Maslow Theory*, *Journal of Education Teaching and Learning*, e-ISSN : 2477-4878, Vol. 2, No 1, pp : 27-33.
- Vito L D, et.al., 2016, *Employee Motivation based on the Hierarchy of Needs, Expectancy and the Two-Factor Theories Applied with Higher Education Employees*, *International Journal of Advances in Management, Economics and Entrepreneurship (IJAMEE)* ISSN: 2349 – 4468, Vol. 3, Issue 1, pp: 20-32.
- Zacharski R, 2018, *Training Sets, Test Sets, and 10-fold Cross-validation*, Kdnuggets, dilihat 10 Agustus 2021, <<https://www.kdnuggets.com/2018/01/training-test-sets-cross-validation.html>>
- Zeida S & Itan N, 2020, *HR Analytics and Organizational Effectiveness*, *International Journal on Emerging Technologies*, 11(2): 683–688.

LAMPIRAN-LAMPIRAN

Lampiran 1 *Dataset siap untuk pemodelan (transforming data)*

umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hte	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
4	M	menikah	3	S1	3	3	tetap	2	0	2	0	2	1	4	4	5	2	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
4	M	menikah	3	S1	3	2	tetap	3	2	2	0	2	1	4	4	3	0	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
4	M	menikah	3	S1	3	3	tetap	3	2	0	0	3	1	4	4	4	0	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
4	M	menikah	3	S1	4	3	tetap	3	2	1	0	4	1	4	4	3	2	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
4	M	menikah	3	S1	4	3	tetap	3	1	0	0	4	1	4	4	4	3	tidak	tidak	tidak	sudah	kurang	baik	tidak	false
5	M	menikah	3	S1	4	3	tetap	4	0	0	0	4	1	0	4	4	3	tidak	tidak	tidak	sudah	tidak	cukup	tidak	false
4	M	menikah	2	D3	2	3	tetap	2	0	0	0	2	0	0	0	5	3	tidak	tidak	tidak	belum	kurang	cukup	tidak	false
3	M	menikah	2	D3	4	3	tetap	4	0	0	0	4	1	0	3	3	3	tidak	tidak	tidak	sudah	kurang	baik	tidak	false
5	M	menikah	2	S1	1	2	tetap	2	0	1	0	1	1	2	2	4	3	sedikit	tidak	tidak	sudah	cukup	cukup	beasiswa	false
...
5	M	menikah	2	S1	2	3	tetap	3	0	0	0	2	1	2	2	4	2	sering	tidak	tidak	sudah	cukup	cukup	tidak	false
5	M	menikah	3	S1	2	3	tetap	3	0	0	0	3	1	2	2	5	3	sering	ya	tidak	sudah	cukup	cukup	tidak	false
5	M	menikah	3	S1	3	3	tetap	3	0	0	0	3	1	2	2	3	3	sering	tidak	ya	sudah	cukup	cukup	tidak	false
5	M	menikah	3	S1	3	3	tetap	4	0	0	0	3	1	0	2	3	3	sedikit	tidak	tidak	sudah	kurang	baik	tidak	false
2	F	menikah	0	S1	1	2	tetap	2	0	1	0	1	1	4	4	5	3	tidak	tidak	ya	belum	aktif	cukup	beasiswa	false
2	F	menikah	1	S1	2	1	tetap	2	1	2	0	1	1	4	4	5	2	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
2	F	menikah	1	S1	2	3	tetap	3	1	0	0	2	1	4	4	4	2	tidak	tidak	ya	sudah	kurang	cukup	tidak	false
3	F	menikah	2	S1	2	3	tetap	3	1	2	0	3	1	4	4	5	0	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
3	F	menikah	2	S1	2	3	tetap	3	2	0	0	3	1	4	4	3	3	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
3	F	menikah	2	S1	3	3	tetap	3	0	0	0	1	1	0	4	3	3	tidak	tidak	tidak	sudah	tidak	baik	tidak	false

umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hnt	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
2	M	menikah	1	D3	1	2	tetap	2	0	2	0	1	1	0	3	4	2	sedikit	tidak	tidak	sudah	aktif	cukup	beasiswa	false
2	M	menikah	1	D3	2	1	tetap	2	0	2	0	1	1	0	3	4	0	tidak	tidak	tidak	sudah	aktif	cukup	tidak	false
2	M	menikah	2	D3	2	3	tetap	2	0	1	0	2	1	0	3	4	2	tidak	tidak	ya	sudah	aktif	cukup	tidak	false
2	M	menikah	2	D3	4	3	tetap	3	0	1	0	3	1	0	3	4	2	tidak	tidak	ya	sudah	aktif	cukup	tidak	false
3	M	menikah	2	D3	4	3	tetap	3	0	1	0	4	1	0	3	5	3	jarang	tidak	tidak	sudah	cukup	baik	tidak	false
3	M	menikah	2	D3	4	3	tetap	3	0	0	0	4	1	0	3	5	3	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
1	M	belum	0	SMK	0	1	tetap	2	0	0	0	1	1	1	2	5	3	tidak	tidak	tidak	sudah	aktif	cukup	beasiswa	false
1	M	menikah	0	SMK	0	1	tetap	2	0	0	0	1	1	1	2	4	3	tidak	tidak	tidak	sudah	aktif	cukup	tidak	false
2	M	menikah	0	SMK	0	1	tetap	2	0	0	0	1	1	0	2	4	0	tidak	tidak	tidak	sudah	kurang	tidak	tidak	true
2	M	menikah	0	S1	1	2	tetap	1	0	0	0	1	1	0	1	4	2	tidak	tidak	ya	sudah	kurang	cukup	beasiswa	false
2	M	menikah	1	S1	2	1	tetap	2	0	0	0	1	1	1	1	5	2	tidak	tidak	ya	sudah	kurang	cukup	tidak	false
3	M	menikah	1	S1	2	3	tetap	2	0	0	0	2	1	0	2	3	2	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
3	M	menikah	2	S1	1	3	tetap	2	0	0	0	2	1	0	2	3	2	tidak	tidak	tidak	sudah	kurang	tidak	tidak	true
...
1	M	belum	0	S1	1	1	tetap	2	1	0	0	1	1	2	3	5	2	tidak	tidak	ya	belum	aktif	cukup	tidak	false
2	M	belum	0	S1	1	2	tetap	2	1	0	0	1	1	2	4	5	2	sedikit	ya	ya	sudah	aktif	cukup	beasiswa	false
2	M	belum	0	S1	2	2	tetap	2	0	0	0	3	1	2	4	5	2	tidak	ya	tidak	sudah	sangat aktif	cukup	tidak	false
2	M	menikah	1	S1	2	1	tetap	3	0	0	0	3	1	2	4	4	3	sedikit	ya	tidak	sudah	aktif	baik	tidak	false
2	M	menikah	1	S1	3	3	tetap	3	0	0	0	3	1	0	4	4	3	tidak	tidak	tidak	sudah	tidak	cukup	tidak	false
1	M	menikah	1	S1	1	2	tetap	1	0	0	0	1	0	0	0	5	3	tidak	ya	ya	belum	kurang	cukup	tidak	false
2	F	menikah	0	S1	2	2	tetap	2	0	0	0	1	0	1	4	4	0	tidak	tidak	ya	sudah	cukup	cukup	tidak	false

umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hnt	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
2	F	menikah	0	S1	2	2	tetap	2	0	0	0	1	0	0	4	4	0	tidak	tidak	tidak	sudah	kurang	cukup	tidak	true
2	M	belum	0	S1	1	2	tetap	1	0	0	1	1	1	4	4	5	2	tidak	tidak	ya	belum	sangat aktif	cukup	beasiswa	false
2	M	menikah	0	S1	2	1	tetap	1	1	1	0	1	1	4	4	5	2	tidak	tidak	tidak	sudah	sangat aktif	cukup	tidak	false
2	M	menikah	1	S1	2	3	tetap	2	2	0	0	1	1	4	4	4	2	tidak	ya	tidak	sudah	sangat aktif	cukup	tidak	false
2	M	menikah	1	S1	3	3	tetap	2	0	0	1	3	1	4	4	4	3	tidak	tidak	ya	sudah	sangat aktif	cukup	tidak	false
2	M	menikah	1	S1	4	3	tetap	2	0	1	0	3	1	4	4	5	3	sedikit	tidak	ya	sudah	cukup	baik	tidak	false
3	M	menikah	1	S1	4	3	tetap	3	0	0	0	3	1	0	4	5	3	tidak	tidak	tidak	sudah	tidak	cukup	tidak	false
1	F	menikah	0	S1	0	1	tetap	1	0	0	1	0	1	4	4	5	2	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
2	F	menikah	1	S1	1	1	tetap	1	3	0	0	0	1	4	4	5	2	tidak	tidak	tidak	sudah	kurang	cukup	beasiswa	false
2	F	menikah	1	S1	1	2	tetap	1	3	1	0	1	1	4	4	4	2	tidak	tidak	ya	sudah	kurang	cukup	tidak	false
2	M	belum	0	SMA	0	1	tetap	1	1	0	1	1	1	2	4	3	0	tidak	tidak	tidak	sudah	sangat aktif	cukup	tidak	false
2	M	belum	0	SMA	1	1	tetap	2	0	0	0	2	1	2	4	4	2	tidak	tidak	ya	sudah	sangat aktif	cukup	tidak	false
2	M	belum	0	SMA	2	3	tetap	2	0	0	0	2	1	2	4	3	3	tidak	tidak	ya	sudah	cukup	cukup	tidak	false
2	M	belum	0	SMA	2	3	tetap	2	0	0	0	2	1	0	4	3	3	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
...
1	F	belum	0	SMA	0	0	tetap	1	0	0	0	0	1	1	4	5	3	tidak	tidak	tidak	sudah	aktif	cukup	beasiswa	false
1	F	belum	0	SMA	0	0	tetap	1	0	0	0	0	1	0	4	5	0	tidak	tidak	tidak	sudah	kurang	cukup	tidak	true
1	M	belum	0	SMA	0	0	tetap	1	0	0	0	0	1	3	4	5	3	tidak	tidak	tidak	belum	aktif	cukup	beasiswa	false
1	M	belum	0	SMA	0	0	tetap	1	0	0	0	0	1	3	4	4	2	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	belum	0	S1	0	1	tetap	1	0	0	0	1	1	3	4	5	2	tidak	ya	tidak	sudah	aktif	cukup	beasiswa	false

umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hnt	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
1	M	belum	0	S1	1	2	tetap	2	0	0	0	2	1	3	4	5	3	tidak	ya	tidak	sudah	cukup	cukup	tidak	false
1	M	belum	0	S1	1	3	tetap	2	0	0	0	2	1	3	4	3	3	tidak	tidak	ya	sudah	cukup	cukup	tidak	false
2	M	belum	0	S1	2	3	tetap	2	0	0	0	2	1	3	4	3	3	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
1	M	belum	0	S1	1	1	tetap	1	0	0	1	1	1	4	4	5	3	sedikit	tidak	ya	belum	aktif	cukup	tidak	false
2	M	menikah	0	S1	1	1	tetap	1	2	2	1	1	1	4	4	5	0	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
2	M	menikah	1	S1	1	2	tetap	1	3	0	1	1	1	4	4	5	2	tidak	tidak	tidak	sudah	kurang	cukup	beasiswa	false
3	M	menikah	4	S1	3	3	tetap	2	0	1	0	3	0	0	0	3	3	sering	tidak	ya	sudah	cukup	baik	tidak	false
3	M	menikah	4	S1	3	3	tetap	2	0	0	0	3	0	0	0	3	3	tidak	tidak	tidak	sudah	tidak	cukup	tidak	false
1	M	belum	0	D3	0	1	tetap	1	0	1	1	0	1	0	3	5	2	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	menikah	0	S1	1	1	tetap	1	0	0	0	0	1	0	3	5	2	tidak	tidak	ya	belum	sangat aktif	cukup	tidak	false
1	M	menikah	0	S1	2	3	tetap	1	0	1	0	1	1	0	3	3	2	sedikit	tidak	tidak	sudah	sangat aktif	cukup	beasiswa	false
2	M	menikah	1	S1	3	3	tetap	2	1	1	0	3	1	0	3	4	0	tidak	tidak	tidak	sudah	aktif	cukup	tidak	false
2	M	menikah	1	S1	3	3	tetap	2	1	0	0	3	1	0	3	3	3	tidak	tidak	tidak	sudah	cukup	baik	tidak	false
2	M	menikah	1	S1	3	3	tetap	2	0	0	0	3	1	0	3	3	3	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
1	F	belum	0	SMA	0	0	tetap	0	0	0	0	0	1	1	3	5	2	tidak	tidak	tidak	belum	aktif	tidak	beasiswa	false
1	F	menikah	0	SMA	0	0	tetap	1	0	0	0	0	1	1	3	5	0	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
1	F	menikah	0	SMA	0	1	tetap	1	0	0	0	0	1	1	3	3	0	sedikit	tidak	tidak	sudah	kurang	cukup	beasiswa	false
1	F	menikah	1	S1	0	1	tetap	1	0	0	0	0	1	0	4	3	0	tidak	tidak	tidak	sudah	tidak	tidak	tidak	true
1	F	belum	0	D3	0	1	tetap	0	0	1	1	0	1	0	4	5	2	jarang	tidak	tidak	belum	aktif	tidak	tidak	false
2	F	belum	0	D3	1	1	tetap	1	0	1	0	0	1	0	4	5	2	sering	ya	tidak	belum	aktif	tidak	tidak	false
...

umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hnt	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
2	F	menikah	1	S1	1	2	tetap	1	0	2	0	2	1	0	4	4	0	sedikit	ya	tidak	sudah	cukup	cukup	beasiswa	false
2	F	menikah	2	S1	1	3	tetap	2	0	0	0	2	1	0	4	4	0	jarang	tidak	tidak	sudah	kurang	tidak	tidak	false
2	F	menikah	2	S1	2	3	tetap	2	0	0	0	2	1	0	4	4	0	tidak	tidak	tidak	sudah	tidak	cukup	tidak	true
1	M	belum	0	D3	0	1	tetap	0	0	1	3	0	1	4	4	5	0	tidak	tidak	tidak	belum	sangat aktif	tidak	beasiswa	false
1	M	belum	0	D3	0	1	tetap	1	0	2	3	0	1	4	4	4	0	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
1	M	belum	0	S1	1	1	tetap	1	1	0	3	1	1	4	4	3	0	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
2	M	belum	0	S1	1	1	tetap	1	2	1	2	2	1	4	4	4	0	tidak	tidak	tidak	sudah	aktif	cukup	beasiswa	false
2	M	menikah	0	S1	1	3	tetap	2	1	1	0	2	1	4	4	3	3	tidak	tidak	tidak	sudah	cukup	cukup	tidak	false
2	M	menikah	1	S1	1	3	tetap	2	0	0	0	2	1	0	4	3	3	tidak	tidak	tidak	sudah	tidak	cukup	tidak	false
1	M	belum	0	SMA	0	0	kontrak	0	0	0	3	0	1	1	3	4	0	tidak	tidak	ya	belum	sangat aktif	tidak	tidak	false
1	M	belum	0	SMA	0	0	kontrak	1	0	0	4	0	1	1	4	4	0	sedikit	tidak	tidak	belum	sangat aktif	tidak	tidak	false
1	M	belum	0	SMA	0	1	tetap	1	0	0	1	0	1	0	4	3	0	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
1	M	belum	0	SMA	0	1	tetap	1	0	0	1	1	1	0	4	4	0	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	belum	0	SMA	0	0	tetap	2	0	0	0	1	1	0	4	3	0	sedikit	tidak	tidak	sudah	tidak	tidak	tidak	true
2	M	belum	0	S1	0	1	tetap	0	0	0	0	0	0	0	0	4	3	tidak	ya	tidak	belum	cukup	tidak	tidak	false
2	M	belum	0	S1	1	1	tetap	1	0	1	0	0	0	0	0	5	2	tidak	tidak	tidak	belum	cukup	tidak	tidak	false
2	M	menikah	1	S1	1	1	tetap	1	0	1	0	0	0	0	0	4	0	tidak	tidak	tidak	belum	cukup	tidak	tidak	false
2	M	menikah	1	S1	1	1	tetap	1	0	2	0	1	0	0	0	5	2	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
2	M	menikah	2	S1	1	3	tetap	2	0	0	0	2	0	0	0	5	3	tidak	tidak	tidak	sudah	kurang	cukup	tidak	false
3	M	menikah	2	S1	2	3	tetap	2	0	0	0	2	0	0	0	5	3	tidak	tidak	tidak	sudah	tidak	baik	tidak	false

umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hte	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
1	M	belum	0	S1	0	1	tetap	0	0	0	0	0	0	2	4	5	2	tidak	tidak	ya	belum	sangat aktif	tidak	tidak	false
1	M	belum	0	D3	1	0	kontrak	0	0	0	0	0	0	0	0	5	2	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	menikah	0	S1	1	0	kontrak	0	0	0	0	0	0	0	0	3	2	tidak	tidak	tidak	belum	aktif	tidak	tidak	false
...
2	M	menikah	1	S1	2	0	tetap	1	0	0	0	0	0	0	0	3	3	tidak	tidak	tidak	sudah	cukup	tidak	tidak	true
1	M	belum	0	D3	0	0	kontrak	0	0	0	0	0	0	0	0	5	2	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	belum	0	D3	1	0	kontrak	0	0	0	0	0	0	0	0	3	3	sedikit	tidak	tidak	belum	aktif	tidak	tidak	false
1	M	belum	0	D3	1	1	tetap	0	1	1	0	0	0	0	0	5	2	sedikit	tidak	tidak	belum	aktif	tidak	tidak	false
2	M	menikah	0	S1	1	2	tetap	1	0	1	0	1	0	0	0	3	3	jarang	ya	ya	sudah	cukup	cukup	tidak	false
2	M	menikah	0	S1	2	2	tetap	1	0	0	0	1	0	0	0	3	3	tidak	tidak	tidak	sudah	kurang	baik	tidak	false
1	M	belum	0	D3	0	0	kontrak	0	0	0	0	0	1	1	3	4	2	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
1	M	belum	0	D3	0	0	kontrak	0	1	0	0	0	1	1	3	4	2	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
2	M	belum	0	D3	1	1	tetap	0	2	1	1	0	1	1	3	5	0	tidak	tidak	tidak	belum	aktif	tidak	tidak	false
2	M	belum	0	D3	1	1	tetap	1	1	0	0	1	1	1	3	5	0	tidak	tidak	ya	belum	cukup	tidak	tidak	false
2	M	belum	0	D3	2	1	tetap	1	0	0	0	1	1	0	3	5	0	tidak	tidak	tidak	belum	kurang	cukup	tidak	false
1	F	belum	0	D3	0	0	kontrak	0	0	0	0	0	1	1	4	5	2	tidak	tidak	tidak	belum	kurang	cukup	tidak	false
1	F	menikah	0	D3	0	0	kontrak	0	0	0	0	0	1	1	4	4	2	sedikit	tidak	tidak	belum	aktif	tidak	tidak	false
1	F	menikah	0	D3	0	0	tetap	0	0	1	0	0	1	1	4	5	3	sedikit	ya	tidak	belum	sangat aktif	cukup	tidak	false
1	F	menikah	0	D3	1	3	tetap	1	1	1	0	0	1	1	4	5	3	tidak	tidak	tidak	sudah	aktif	cukup	tidak	false
2	M	belum	0	S1	0	0	kontrak	0	0	0	0	0	0	0	0	4	3	tidak	tidak	tidak	belum	aktif	tidak	tidak	false
2	M	menikah	0	S1	1	1	tetap	0	0	1	0	0	0	0	0	5	2	tidak	ya	tidak	belum	cukup	tidak	tidak	false

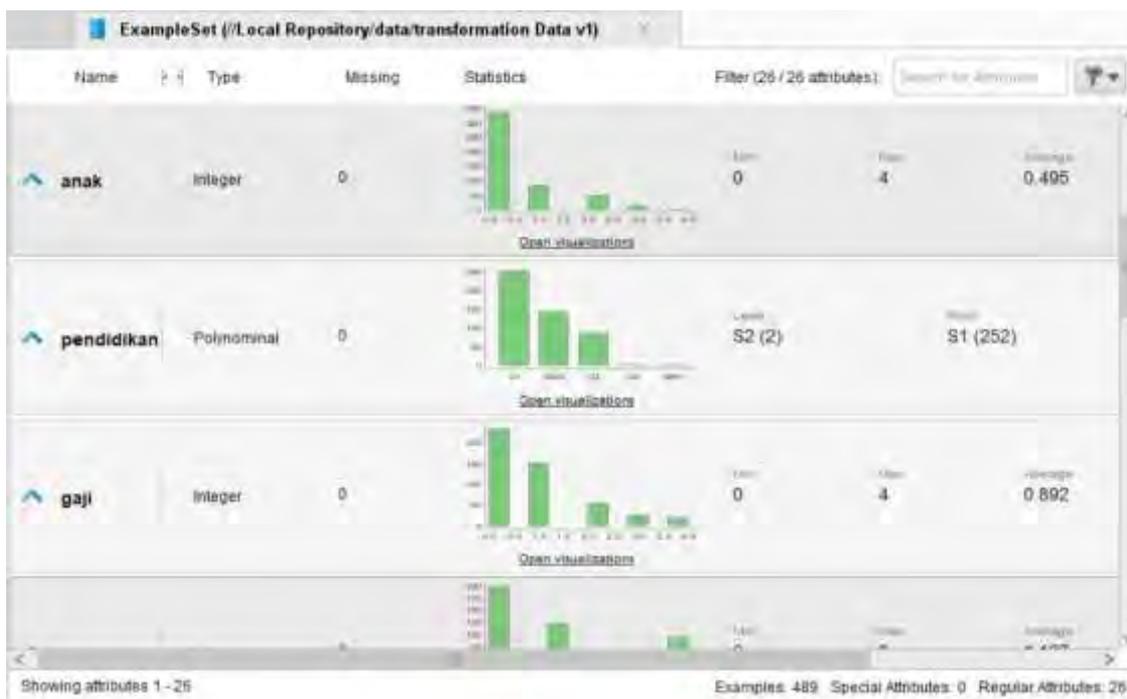
umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hnt	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
3	M	menikah	0	S1	1	2	tetap	1	0	2	0	1	0	0	0	5	3	tidak	ya	tidak	belum	cukup	tidak	tidak	false
3	M	menikah	1	S1	1	2	tetap	1	0	0	0	1	0	0	0	5	3	tidak	tidak	tidak	belum	tidak	baik	tidak	false
1	M	menikah	1	D3	0	0	kontrak	0	0	0	0	0	0	0	0	4	2	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
1	M	belum	0	D3	0	0	kontrak	0	0	0	0	0	0	0	0	4	2	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
1	M	menikah	0	D3	0	0	kontrak	0	0	0	0	0	0	0	0	5	2	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
2	M	menikah	1	D3	1	1	kontrak	1	0	0	0	0	0	0	0	4	3	tidak	tidak	tidak	belum	cukup	tidak	tidak	false
2	M	belum	0	S1	1	0	kontrak	0	1	0	0	0	0	4	3	5	3	tidak	tidak	tidak	belum	cukup	tidak	tidak	false
...
2	M	menikah	0	S1	1	2	tetap	1	0	0	0	0	0	0	3	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	D3	0	0	kontrak	0	0	0	0	0	1	0	3	3	3	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
1	M	belum	0	D3	0	0	kontrak	0	0	0	0	0	1	0	3	5	3	tidak	tidak	tidak	belum	sangat aktif	tidak	tidak	false
2	M	belum	0	D3	1	2	tetap	0	0	0	0	0	1	0	3	4	3	tidak	tidak	tidak	belum	cukup	cukup	beasiswa	false
2	M	belum	0	D3	1	2	tetap	1	0	0	0	0	1	0	3	4	3	tidak	tidak	tidak	belum	tidak	baik	tidak	false
2	M	menikah	0	D3	0	0	kontrak	0	0	0	0	0	1	0	4	3	3	tidak	tidak	tidak	belum	kurang	cukup	tidak	false
0	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	0	1	2	4	3	sedikit	tidak	tidak	belum	cukup	tidak	tidak	false
0	M	belum	0	SMA	0	1	tetap	0	0	0	0	0	0	1	2	5	3	sedikit	tidak	tidak	belum	kurang	cukup	tidak	false
1	M	belum	0	SMA	0	1	tetap	1	0	0	0	0	0	0	2	5	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
0	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	0	0	0	4	3	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
0	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	0	0	0	5	3	sedikit	tidak	tidak	belum	aktif	tidak	tidak	false
0	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	0	0	0	5	3	tidak	tidak	tidak	belum	kurang	tidak	tidak	true

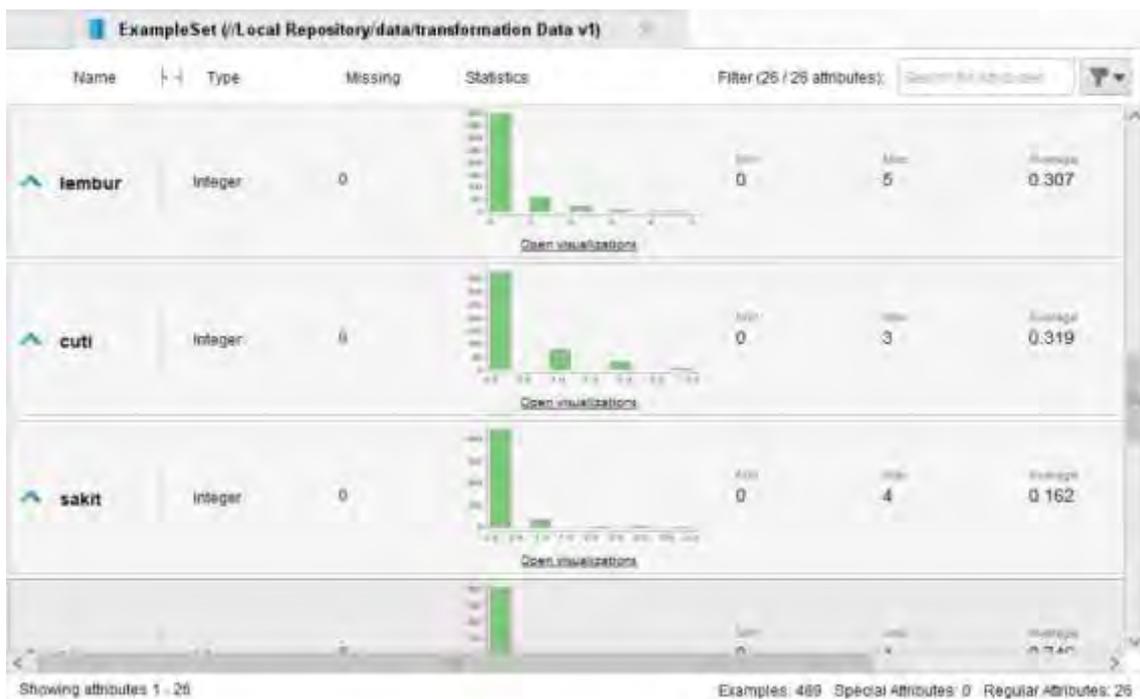
umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hte	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
1	F	belum	0	D3	1	0	kontrak	0	0	0	0	0	0	0	2	4	3	sedikit	tidak	tidak	belum	kurang	cukup	tidak	false
1	F	menikah	0	D3	1	0	kontrak	0	0	0	0	0	0	0	2	5	3	jarang	tidak	tidak	belum	kurang	tidak	tidak	false
1	F	menikah	1	D3	1	2	tetap	0	0	0	0	0	0	0	2	3	3	jarang	tidak	tidak	belum	kurang	tidak	tidak	false
1	F	menikah	1	D3	1	2	tetap	1	0	0	0	0	0	0	2	3	3	tidak	tidak	tidak	belum	tidak	baik	tidak	false
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	1	0	2	5	3	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	belum	0	SMA	1	0	kontrak	0	0	0	0	0	1	0	2	5	2	sedikit	tidak	tidak	belum	aktif	tidak	tidak	false
1	M	belum	0	S1	1	1	tetap	0	0	0	0	0	1	0	2	3	3	jarang	tidak	tidak	belum	cukup	tidak	tidak	false
1	M	belum	0	S1	1	1	tetap	1	0	0	0	0	1	0	2	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	SMA	1	0	kontrak	0	0	0	0	0	1	4	1	5	3	sedikit	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	menikah	0	SMA	1	1	kontrak	0	0	0	0	0	1	4	1	3	3	jarang	tidak	tidak	belum	cukup	tidak	tidak	false
1	M	menikah	0	S1	1	1	kontrak	0	0	0	0	0	1	0	1	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	F	belum	0	D4	0	0	kontrak	0	0	0	0	0	1	1	4	5	0	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
...
1	F	menikah	0	D4	0	1	kontrak	0	0	0	0	0	1	0	4	5	3	tidak	tidak	tidak	belum	tidak	cukup	tidak	false
1	M	belum	0	S1	0	0	kontrak	0	0	0	0	0	1	2	2	5	0	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
1	M	belum	0	S1	0	0	kontrak	0	0	0	0	0	1	2	2	4	3	sedikit	tidak	tidak	belum	cukup	cukup	tidak	false
1	M	menikah	0	S1	0	0	kontrak	0	0	0	0	0	1	0	2	4	3	tidak	tidak	tidak	belum	kurang	cukup	tidak	false
1	M	belum	0	SMA	1	0	kontrak	0	0	0	0	0	1	0	1	4	2	tidak	tidak	tidak	belum	aktif	cukup	tidak	false
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	1	0	2	3	3	sering	tidak	tidak	belum	cukup	tidak	tidak	false
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	1	0	2	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	F	belum	0	S1	0	0	kontrak	0	0	0	0	0	0	0	0	5	3	sedikit	tidak	tidak	belum	cukup	cukup	tidak	false

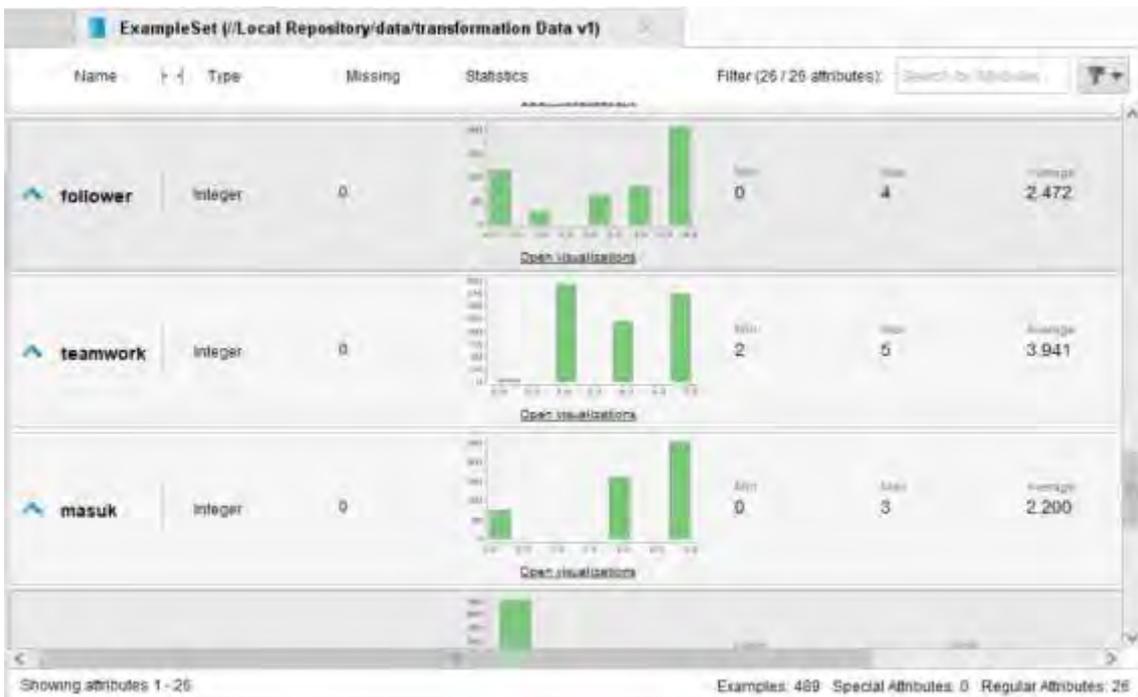
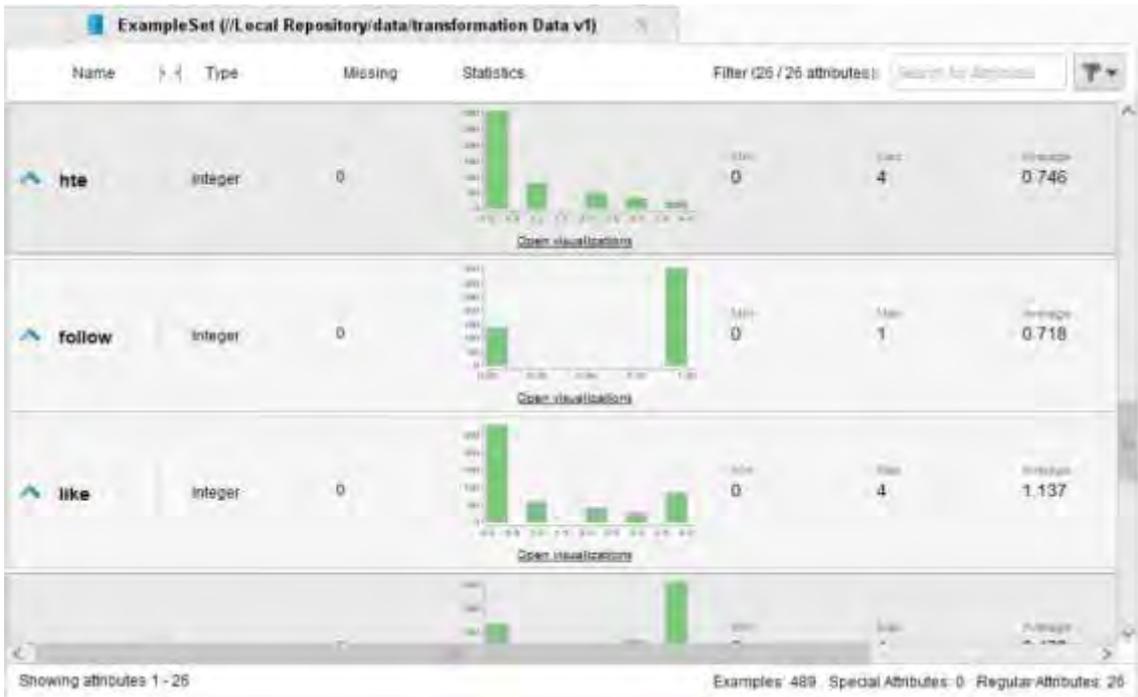
umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hnt	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
1	F	belum	0	S1	0	1	kontrak	0	0	0	0	0	0	0	0	5	3	sering	tidak	tidak	belum	aktif	tidak	tidak	false
2	F	belum	0	S1	0	1	kontrak	0	0	0	0	0	0	0	0	5	3	tidak	tidak	tidak	belum	kurang	cukup	tidak	false
1	M	belum	0	S1	0	0	kontrak	0	0	0	0	0	1	2	1	5	3	jarang	tidak	tidak	belum	sangat aktif	cukup	tidak	false
1	M	belum	0	S1	0	1	kontrak	0	0	0	0	0	1	2	2	5	3	jarang	ya	tidak	belum	cukup	tidak	tidak	false
1	M	menikah	0	S1	1	1	kontrak	0	0	0	0	0	1	0	2	3	3	sering	tidak	tidak	belum	cukup	tidak	tidak	false
1	M	menikah	0	S1	1	1	kontrak	0	0	0	0	0	1	0	2	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
3	M	belum	0	S1	1	0	kontrak	0	0	0	0	0	1	0	0	4	3	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
3	M	belum	0	S1	1	1	kontrak	0	0	0	0	0	1	0	0	3	3	jarang	tidak	tidak	belum	cukup	tidak	tidak	false
3	M	belum	0	S1	1	1	kontrak	0	0	0	0	0	1	0	0	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	D3	0	0	kontrak	0	0	0	0	0	1	0	0	3	3	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
1	M	menikah	0	D3	0	1	kontrak	0	1	0	0	0	1	0	0	3	3	tidak	tidak	tidak	belum	cukup	tidak	tidak	false
1	M	menikah	0	D3	1	1	kontrak	0	0	0	0	0	1	0	0	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	D3	0	0	kontrak	0	0	0	0	0	1	2	3	4	3	tidak	tidak	tidak	belum	sangat aktif	cukup	tidak	false
...
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	0	0	0	3	3	jarang	ya	tidak	belum	cukup	cukup	tidak	false
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	0	0	0	3	3	tidak	tidak	tidak	belum	kurang	baik	beasiswa	false
0	F	belum	0	SMA	0	0	kontrak	0	0	0	0	0	1	3	4	3	3	selalu	tidak	tidak	belum	sangat aktif	cukup	tidak	false
0	F	belum	0	SMA	0	0	kontrak	0	0	0	0	0	1	3	4	3	3	tidak	tidak	tidak	belum	tidak	cukup	beasiswa	false
1	F	belum	0	S1	0	0	kontrak	0	0	0	0	0	0	0	4	4	3	sering	tidak	tidak	belum	aktif	cukup	tidak	false
1	F	belum	0	S1	0	0	kontrak	0	0	0	0	0	0	0	4	4	3	tidak	tidak	tidak	belum	tidak	cukup	tidak	false

umur	gender	status	anak	pendidikan	gaji	bonus	pegawai	lama	lembur	cuti	sakit	hte	follow	like	follower	teamwork	masuk	monthly best	penghargaan	rekomendasi	umroh	ekstrakurikuler	upgrading	beasiswa	resign
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	1	0	0	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	S1	0	0	kontrak	0	0	0	0	0	0	0	0	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	S1	0	0	kontrak	0	0	0	0	0	1	0	1	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	1	1	0	3	3	tidak	tidak	tidak	belum	kurang	baik	tidak	false
1	M	belum	0	SMA	0	0	kontrak	0	0	0	0	0	0	0	0	3	3	tidak	tidak	tidak	belum	tidak	cukup	tidak	false

Lampiran 2 Detail pengecekan dataset secara statistik









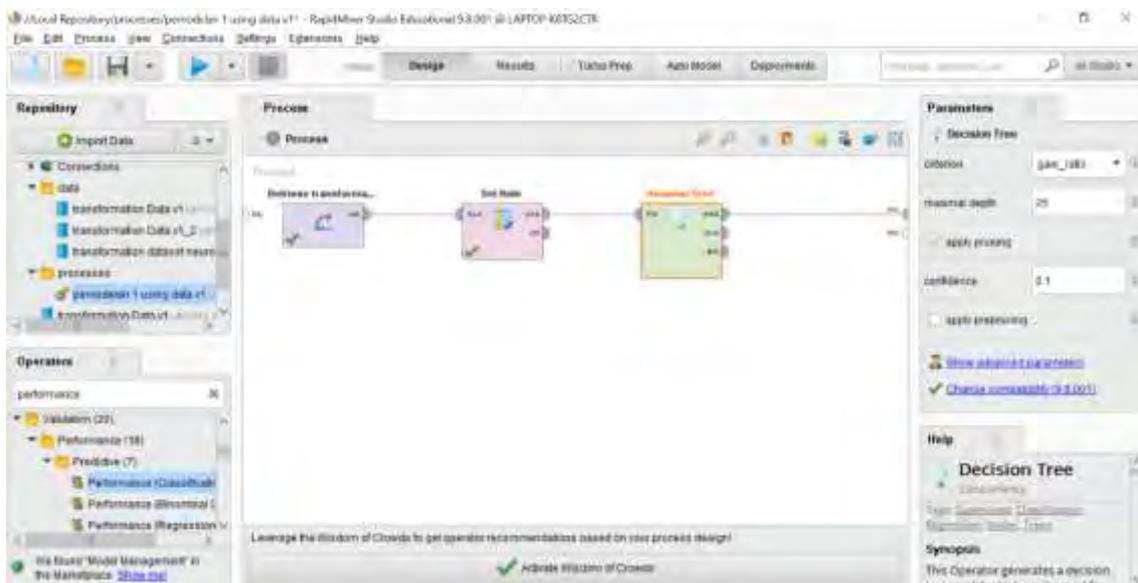


Lampiran 3Tangkapan layar percobaan pemodelan

1. Tanpa parameter *pruning* dan *prepruning*



2. Hanya menggunakan parameter *pruning* dan *prepruning*



3. Menggunakan parameter *pruning* dan *prepruning*



Lampiran 4 Hasil pemodelan berbentuk deskripsi (*rule model*)

Deskripsi model pohon keputusan tanpa menggunakan parameter *pruning* dan *prepruning* :

Tree

upgrading = baik: false {false=76, true=0}

upgrading = cukup

| masuk > 1: false {false=217, true=0}

| masuk ≤ 1

| | gender = F

| | | like > 0.500: false {false=8, true=0}

| | | like ≤ 0.500

| | | | monthly best = sedikit: false {false=1, true=0}

| | | | monthly best = tidak: true {false=0, true=3}

| | gender = M: false {false=30, true=0}

upgrading = tidak

| lama > 1.500

| | umroh = belum: false {false=1, true=0}

| | umroh = sudah

| | | monthly best = jarang: false {false=1, true=0}

| | | monthly best = sedikit: true {false=0, true=1}

| | | monthly best = tidak: true {false=0, true=3}

| lama ≤ 1.500

| | gaji > 1.500

| | | lama > 0.500: true {false=0, true=2}

| | | lama ≤ 0.500: false {false=1, true=0}

| | gaji ≤ 1.500

| | | ekstrakurikuler = aktif: false {false=40, true=0}

| | | ekstrakurikuler = cukup

| | | | follower > 3.500

| | | | | follow > 0.500: false {false=7, true=0}
 | | | | | follow ≤ 0.500: true {false=0, true=1}
 | | | | | follower ≤ 3.500: false {false=23, true=0}
 | | | ekstrakulikuler = kurang
 | | | | umur > 0.500
 | | | | | like > 0.500: false {false=8, true=0}
 | | | | | like ≤ 0.500
 | | | | | | masuk > 2.500: false {false=5, true=0}
 | | | | | | masuk ≤ 2.500
 | | | | | | | lembur > 1: false {false=2, true=0}
 | | | | | | | lembur ≤ 1
 | | | | | | | | umroh = belum
 | | | | | | | | | umur > 1.500: true {false=0, true=3}
 | | | | | | | | | umur ≤ 1.500
 | | | | | | | | | pendidikan = D3: true {false=0, true=1}
 | | | | | | | | | pendidikan = S1: false {false=2, true=0}
 | | | | | | | | | umroh = sudah: false {false=1, true=0}
 | | | | umur ≤ 0.500: true {false=0, true=1}
 | | | ekstrakulikuler = sangat aktif: false {false=43, true=0}
 | | | ekstrakulikuler = tidak
 | | | | monthly best = sedikit: false {false=1, true=0}
 | | | | monthly best = tidak
 | | | | | anak > 0.500: true {false=0, true=4}
 | | | | | anak ≤ 0.500
 | | | | | | gender = F: false {false=1, true=0}
 | | | | | | gender = M: true {false=0, true=2}

Deskripsi model pohon keputusan menggunakan parameter *pruning* saja :

Tree

upgrading = baik: false {false=76, true=0}

upgrading = cukup

| masuk > 1: false {false=217, true=0}

| masuk ≤ 1

| | gender = F

| | | like > 0.500: false {false=8, true=0}

| | | like ≤ 0.500

| | | | monthly best = sedikit: false {false=1, true=0}

| | | | monthly best = tidak: true {false=0, true=3}

| | gender = M: false {false=30, true=0}

upgrading = tidak

| lama > 1.500

| | umroh = belum: false {false=1, true=0}

| | umroh = sudah

| | | monthly best = jarang: false {false=1, true=0}

| | | monthly best = sedikit: true {false=0, true=1}

| | | monthly best = tidak: true {false=0, true=3}

| lama ≤ 1.500

| | gaji > 1.500

| | | lama > 0.500: true {false=0, true=2}

| | | lama ≤ 0.500: false {false=1, true=0}

| | gaji ≤ 1.500

| | | ekstrakurikuler = aktif: false {false=40, true=0}

| | | ekstrakulikuler = cukup
 | | | | follower > 3.500
 | | | | | follow > 0.500: false {false=7, true=0}
 | | | | | follow ≤ 0.500: true {false=0, true=1}
 | | | | follower ≤ 3.500: false {false=23, true=0}
 | | | ekstrakulikuler = kurang
 | | | | umur > 0.500
 | | | | | like > 0.500: false {false=8, true=0}
 | | | | | like ≤ 0.500
 | | | | | | masuk > 2.500: false {false=5, true=0}
 | | | | | | masuk ≤ 2.500
 | | | | | | | lembur > 1: false {false=2, true=0}
 | | | | | | | lembur ≤ 1
 | | | | | | | | umroh = belum
 | | | | | | | | | umur > 1.500: true {false=0, true=3}
 | | | | | | | | | umur ≤ 1.500
 | | | | | | | | | | pendidikan = D3: true {false=0, true=1}
 | | | | | | | | | | pendidikan = S1: false {false=2, true=0}
 | | | | | | | | | | umroh = sudah: false {false=1, true=0}
 | | | | | umur ≤ 0.500: true {false=0, true=1}
 | | | ekstrakulikuler = sangat aktif: false {false=43, true=0}
 | | | ekstrakulikuler = tidak
 | | | | monthly best = sedikit: false {false=1, true=0}
 | | | | monthly best = tidak
 | | | | | anak > 0.500: true {false=0, true=4}
 | | | | | anak ≤ 0.500
 | | | | | | gender = F: false {false=1, true=0}
 | | | | | | gender = M: true {false=0, true=2}

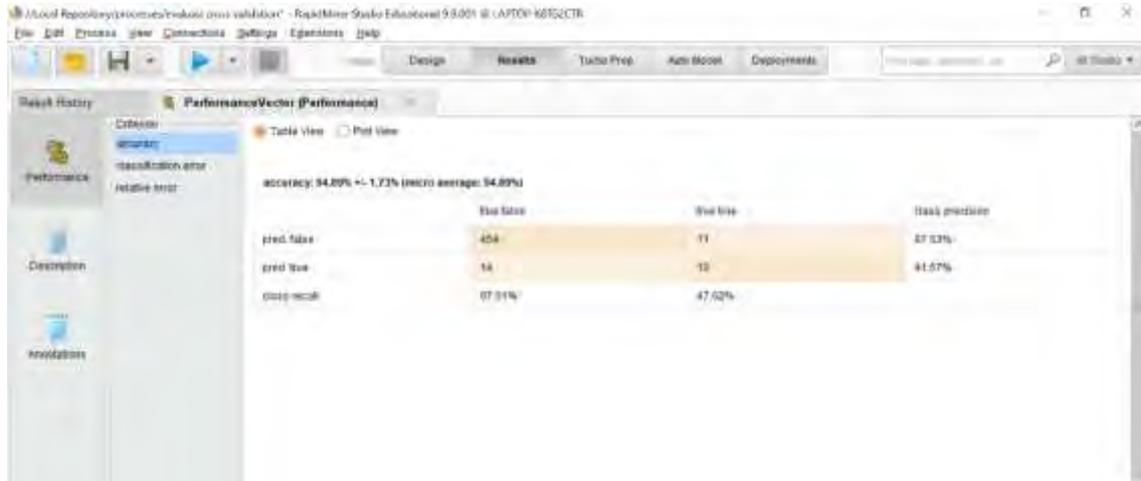
Deskripsi model pohon keputusan menggunakan parameter *pruning* dan *prepruning* :

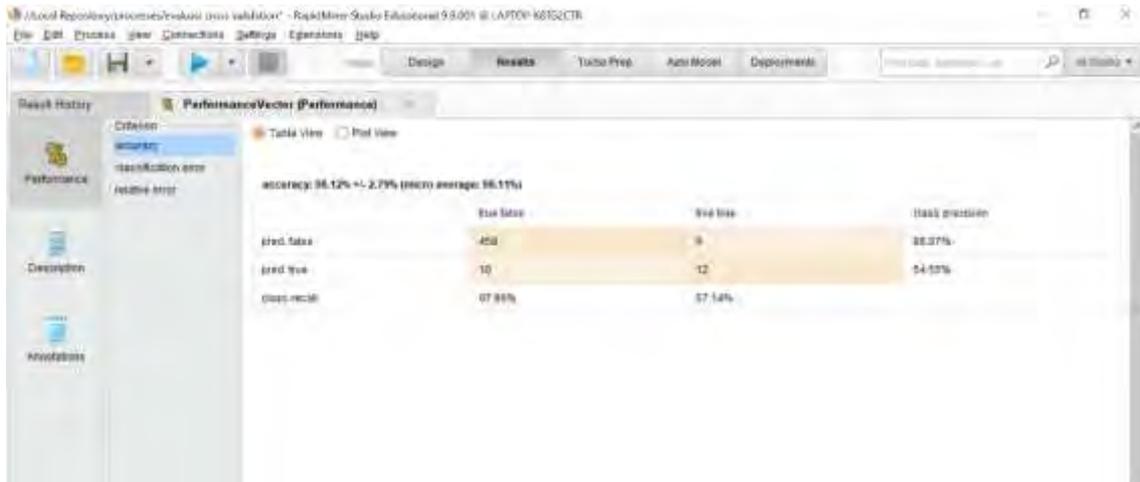
Tree

```
upgrading = baik: false {false=76, true=0}
upgrading = cukup
| masuk > 1: false {false=217, true=0}
| masuk ≤ 1
| | gender = F
| | | like > 0.500: false {false=8, true=0}
| | | like ≤ 0.500: true {false=1, true=3}
| | gender = M: false {false=30, true=0}
upgrading = tidak
| lama > 1.500: true {false=2, true=4}
| lama ≤ 1.500
| | gaji > 1.500: true {false=1, true=2}
| | gaji ≤ 1.500
| | | ekstrakulikuler = aktif: false {false=40, true=0}
| | | ekstrakulikuler = cukup: false {false=30, true=1}
| | | ekstrakulikuler = kurang
| | | | like > 0.500: false {false=8, true=0}
| | | | like ≤ 0.500
| | | | | bonus > 1.500: false {false=4, true=0}
| | | | | bonus ≤ 1.500
| | | | | lembur > 0.500: false {false=3, true=0}
| | | | | lembur ≤ 0.500
| | | | | | umur > 1.500: true {false=0, true=3}
| | | | | | umur ≤ 1.500
| | | | | | | gaji > 0.500: false {false=2, true=0}
| | | | | | | gaji ≤ 0.500: true {false=1, true=2}
| | | ekstrakulikuler = sangat aktif: false {false=43, true=0}
| | | ekstrakulikuler = tidak
| | | | anak > 0.500: true {false=0, true=4}
| | | | anak ≤ 0.500
| | | | gender = F: false {false=2, true=0}
| | | | gender = M: true {false=0, true=2}
```

Lampiran 5 Tangkapan layar hasil evaluasi performansi algoritma C4.5

Tangkapan layar hasil evaluasi secara tingkat akurasi dan *relative error*, ditampilkan berurutan tanpa parameter *pruning* dan *prepruning*, hanya *pruning* saja, dan memakai *pruning* dan *prepruning*.





Local Repository/processes/evaluate cross validation - RapidMiner Studio Educational 9.9.001 @ LAPTOP-K87G2CTR

File Edit Process View Connections Settings Extensions Help

Design Results Turbo Prep Auto Model Deployments

Result History PerformanceVector (Performance [0]) Tax (Decision Tree [2])

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

Table View Plot View

precision: 41.67% (positive class: true)

	true false	true true	class precision
pred false	454	11	97.63%
pred true	14	10	41.67%
class recall	97.01%	47.62%	

Local Repository/processes/evaluate cross validation - RapidMiner Studio Educational 9.9.001 @ LAPTOP-K87G2CTR

File Edit Process View Connections Settings Extensions Help

Design Results Turbo Prep Auto Model Deployments

Result History PerformanceVector (Performance [0]) Tree (Decision Tree [2])

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

Table View Plot View

precision: 41.67% (positive class: true)

	true false	true true	class precision
pred false	454	11	97.63%
pred true	14	10	41.67%
class recall	97.01%	47.62%	

