

**KLASIFIKASI AKSES SARANA AIR MINUM
MENGUNAKAN ALGORITMA NAÏVE
BAYES DAN DECISION TREE C4.5
(STUDI KASUS: PAMSIMAS III KEMENPUPR)**

TESIS

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

Oleh:

NUR YULIAS

NPM: 2019210104



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

**KLASIFIKASI AKSES SARANA AIR MINUM
MENGUNAKAN ALGORITMA NAÏVE
BAYES DAN DECISION TREE C4.5
(STUDI KASUS: PAMSIMAS III KEMENPUPR)**

Oleh:

NUR YULIAS

NPM: 2019210104

Bandung, 03 Juni 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**

*Thanks to Allah SWT
To my wife Nisa Uswatun Hasanah
And my daughter Ghaida Rizghita Aisyah*

ABSTRAK

KLASIFIKASI AKSES SARANA AIR MINUM MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN DECISION TREE C4.5 (STUDI KASUS: PAMSIMAS III KEMENPUPR)

Oleh:
NUR YULIAS
NPM: 2019210104

PAMSIMAS merupakan program unggulan nasional yang bertujuan untuk membantu masyarakat agar dapat memenuhi akses dasar terhadap sarana air minum dan sanitasi. Program ini sudah berjalan di 33 provinsi di Indonesia mulai dari tahun 2008 hingga saat ini. Status keberfungsian seluruh desa-desa PAMSIMAS; sarana dan jumlah pemanfaatannya selalu dilaporkan secara periodik melalui aplikasi SIM berbasis *website* untuk mendapatkan laporan secara *realtime*.

Desa-desa yang sudah selesai membangun sarana PAMSIMAS, selanjutnya akan disebut sebagai desa keberlanjutan atau desa paska PAMSIMAS. Seluruh desa-desa keberlanjutan tersebut masih bisa mendapatkan pendanaan lainnya seperti dana hibah dengan tujuan agar akses masyarakat terhadap sarana air minum dapat tercapai hingga 100%. Oleh karena itu, dalam menetapkan desa keberlanjutan yang tepat untuk mendapatkan dana hibah maka diperlukan sebuah pengetahuan yang dapat memudahkan dalam memutuskan desa-desa mana saja yang bisa mendapatkan dana tersebut.

Salah satu strategi yang dapat digunakan adalah memanfaatkan *data mining* sebagai alat pendukung keputusan untuk mencari informasi dan pengetahuan dengan cara menganalisa data-data desa keberlanjutan menggunakan teknik klasifikasi dengan memanfaatkan algoritma Naive Bayes dan *Decision Tree* C4.5 pada *dataset* desa keberlanjutan PAMSIMAS yang bersifat *supervised learning*.

Berdasarkan proses tahapan *data mining* pada algoritma *Decision Tree* C4.5 telah menghasilkan sebuah pohon keputusan yang terbentuk dari akar dan cabang dengan nilai *Gain* tertinggi. Atribut kondisi SAM menjadi akar pohon keputusan karena memiliki nilai *Gain* tertinggi daripada atribut-atribut lainnya. Melalui pohon keputusan inilah lahir suatu pengetahuan berupa *rule* atau pola yaitu atribut-atribut yang berpengaruh dominan dalam mencapai akses universal sarana air minum.

Penggunaan algoritma Naive Bayes dan *Decision Tree* C4.5 yang telah diimplementasikan menggunakan RapidMiner, selanjutnya dilakukan pengukuran kinerja klasifikasi menggunakan *confusion matrix* untuk melihat *performance metrics* pada masing-masing algoritma. Hasil *performance metrics* terhadap pengukuran tingkat akurasi berdasarkan perhitungan yang telah dilakukan dengan menggunakan 29.378 data sample pada masing-masing model, lalu dilakukan *balancing data* agar data kelas menjadi seimbang, maka diperoleh hasil akhir yaitu perhitungan dengan menggunakan algoritma *Decision Tree* C4.5 mendapatkan tingkat akurasi sebesar 56.59% dan perhitungan menggunakan algoritma Naive Bayes mendapatkan tingkat akurasi sebesar 54.40%. Sedangkan untuk pengukuran presisi, *Decision Tree* C4.5 mendapatkan nilai sebesar 56.26% dan Naive Bayes mendapatkan nilai 52.91%. Selanjutnya untuk hasil pengukuran *recall*, *Decision Tree* C4.5 mendapatkan nilai sebesar 59.22% sedangkan Naive Bayes mendapatkan nilai sebesar 80.22%.

Kata Kunci: pamsimas, *classification*, *data mining*, *naive bayes*, *decision tree* C4.5

ABSTRACT

CLASSIFICATION OF ACCESS DRINKING WATER FACILITIES USING NAÏVE BAYES ALGORITHM AND DECISION TREE C4.5 (CASE STUDY: PAMSIMAS III KEMENPUPR)

By:
NUR YULIAS
NPM: 2019210104

PAMSIMAS is a national flagship program that aims to help the community to fulfill basic access to drinking water and sanitation facilities. This program has been running in 33 provinces in Indonesia from 2008 to the present. Functional status of all PAMSIMAS villages; the facilities and the number of users are always reported periodically through a website-based SIM application to get real-time reports.

Villages that have finished building PAMSIMAS facilities will be referred to as sustainability villages or post PAMSIMAS villages. All of these sustainability villages can still get other funding, such as grants with the aim that community access to drinking water facilities can be achieved up to 100%. Therefore, in determining the right sustainability village to get grant funds, knowledge is needed that can make it easier to decide which villages can get these funds.

One strategy that can be used is to use data mining as a decision support tool to find information and knowledge by analyzing sustainability village data using classification techniques using the Naive Bayes algorithm and Decision Tree C4.5 on the PAMSIMAS sustainability village dataset which is supervised learning.

Based on the process of data mining stages in the Decision Tree C4.5 algorithm, it has produced a decision tree formed from root and branch with the highest Gain value. The SAM condition attribute becomes the root of the decision tree because it has the highest Gain value than other attributes. Through this decision tree, knowledge in the form of rules or patterns is born, namely the attributes that have a dominant influence in achieving universal access to drinking water facilities.

The use of the Naive Bayes algorithm and Decision Tree C4.5 which has been implemented using RapidMiner, then measures the classification performance using a confusion matrix to see the performance metrics for each algorithm. The results of performance metrics for measuring the level of accuracy are based on calculations that have been carried out using 29,378 sample data in each model, then data balancing is carried out so that the class data becomes balanced, then the final result calculation using the Decision Tree C4.5 algorithm for level of accuracy get 56.59% and using the Naive Bayes algorithm get 54.40%. As for precision measurement, Decision Tree C4.5 gets a value of 56.26% and Naive Bayes gets a value of 52.91%. Furthermore, for the recall measurement results, Decision Tree C4.5 gets a value of 59.22% while Naive Bayes gets a value of 80.22%.

Keywords: pamsimas, classification, data mining, naive bayes, decision tree c4.5

KATA PENGANTAR

Segala puji bagi Allah SWT atas limpahan rahmat dan rezeki serta petunjuk-Nya, sehingga penulis dapat menyelesaikan tesis dengan judul : “KLASIFIKASI AKSES SARANA AIR MINUM MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN DECISION TREE C4.5 (Studi Kasus PAMSIMAS III Kementerian Pekerjaan Umum dan Perumahan Rakyat)” 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. Rekan-rekan Tim ROMS 7 PAMSIMAS Provinsi Banten yang telah meluangkan waktunya untuk membantu penulis dalam berdiskusi dan bertukar pikiran.
3. Orang tua tercinta dan tersayang, ayahanda almarhum H. Dzakaria dan almarhumah Hj. Tati Sumiati.
4. Keluarga besar Bani Akoy yang selalu memberikan motivasi dan supportnya.
5. Istri tercinta Nisa Uswatun Hasanah, S.Pd.I dan anakku tersayang Ghaida Rizqhita Aisyah yang selalu mendampingi dan memberikan energi dan dukungan positif bagi penulis untuk menyelesaikan tesis ini.
6. Rekan-rekan *Data Management Anaylist* PAMSIMAS seluruh Indonesia yang selalu mendukung dan memberikan semangat.
7. Seluruh Bapak dan Ibu dosen program Magister Sistem Informasi STMIK LIKMI Bandung, atas ilmu yang tiada ternilai harganya.

8. Teman-teman seperjuangan seluruh mahasiswa S2 STMIK LIKMI yang selalu semangat dalam menyelesaikan studi.
9. Seluruh staf pengelola dan karyawan program Magister Sistem Informasi STMIK LIKMI Bandung.
10. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu dalam penulisan tesis ini.

Bandung, 03 Juni 2021

Penulis

DAFTAR ISI

ABSTRAK.....	iii
ABSTRACT	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR RUMUS	xii
DAFTAR LAMPIRAN.....	xii
DAFTAR SINGKATAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	5
1.4 Ruang Lingkup Penelitian.....	5
1.5 Metode Penelitian.....	5
1.6 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 <i>Data Mining</i>	8
2.1.1 Peran utama <i>data mining</i>	12
2.1.2 Tahapan <i>data mining</i>	13
2.1.3 Arsitektur <i>data mining</i>	14
2.2 Klasifikasi.....	15
2.2.1 Algoritma Naïve Bayes.....	17
2.2.2 Algoritma <i>Decision Tree</i> C4.5.....	19
2.2.3 Pengukuran kinerja klasifikasi.....	22
2.2.4 <i>Balancing dataset</i>	23
2.2.5 <i>Uji T-test</i>	24

2.3	Knowledge Discovery In Database (KDD)	24
2.4	Rapidminer	27
2.5	PAMSIMAS	28
2.6	Penelitian Terkait	29
BAB III OBJEK DAN METODOLOGI PENELITIAN		31
3.1	Profil PAMSIMAS	31
3.1.1	Tujuan dan sasaran PAMSIMAS	32
3.1.2	Struktur organisasi PAMSIMAS	34
3.1.3	Proses bisnis	35
3.2	Metodologi Penelitian	37
3.2.1	Persiapan penelitian	39
3.2.2	Seleksi data	39
3.2.3	Praproses data	40
3.2.4	Transformasi data	41
3.2.5	<i>Data mining</i>	42
3.2.6	Interpretasi dan evaluasi	44
3.2.7	Penggunaan pengetahuan	45
BAB IV HASIL DAN PEMBAHASAN		46
4.1	Seleksi Data	46
4.2	Praproses Data	48
4.3	Transformasi Data	48
4.4	<i>Data Mining</i>	49
4.4.1	Algoritma Naive Bayes	51
4.4.2	Algoritma <i>Decision Tree</i> C4.5	53
4.5	Interpretasi dan Evaluasi	56
4.6	Pembahasan Hasil Eksperimen	59
BAB V KESIMPULAN DAN SARAN		62
5.1	Kesimpulan	62
5.2	Saran	63

DAFTAR PUSTAKA.....	64
LAMPIRAN-LAMPIRAN	66

DAFTAR GAMBAR

Gambar 1.1 Trend perkembangan penelitian jurnal <i>data mining</i>	1
Gambar 2.1 Pengelompokan Metode <i>Data Mining</i>	9
Gambar 2.2 <i>Life Cycle of a Data Mining Project</i>	13
Gambar 2.3 Arsitektur <i>Data Mining</i>	15
Gambar 2.4 Proses Klasifikasi Data	16
Gambar 2.5 Contoh bentuk pohon keputusan untuk data klasifikasi	19
Gambar 2.6 Pembentukan pohon keputusan menggunakan <i>Decision Tree C4.5</i>	22
Gambar 2.7 Proses <i>Knowledge Discovery in Database (KDD)</i>	26
Gambar 3.1 Struktur Organisasi Pengelola dan Pelaksana PAMSIMAS	34
Gambar 3.2 Proses Bisnis Pelaporan Data SIM PAMSIMAS.....	36
Gambar 3.3 Proses <i>Data Mining</i> Pada <i>Dataset</i> PAMSIMAS	37
Gambar 3.4 Tahapan Penelitian Berdasarkan KDD	38
Gambar 3.5 Cara Kerja Algoritma Naive Bayes.....	43
Gambar 3.6 Pembentukan pohon keputusan dengan algoritma <i>Decision Tree C4.5</i>	44
Gambar 4.1 Proses seleksi data dengan RapidMiner	46
Gambar 4.2 Proses <i>cleaning</i> data menggunakan RapidMiner	48
Gambar 4.3 Grafik <i>imbalancing dataset</i>	50
Gambar 4.4 Grafik hasil <i>balancing dataset</i>	50
Gambar 4.5 Proses validasi model Naive Bayes dengan <i>Cross Validation</i>	51
Gambar 4.6 Proses <i>modeling</i> menggunakan algoritma Naive Bayes	52
Gambar 4.7 Hasil <i>ExampleSet</i> data menggunakan <i>modeling</i> Naive Bayes.....	53
Gambar 4.8 Proses validasi model <i>Decision Tree C4.5</i> dengan <i>Cross Validation</i>	54
Gambar 4.9 <i>Setting</i> sub proses <i>training</i> dan <i>testing</i> pada <i>Decision Tree C4.5</i>	55
Gambar 4.10 Hasil <i>ExampleSet</i> menggunakan <i>modeling</i> <i>Decision Tree C4.5</i>	55
Gambar 4.11 Pembentukan akar dan cabang pohon keputusan	56
Gambar 4.12 <i>Performance test</i> pada algoritma.....	57

DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix</i>	22
Tabel 2.2 Tabel Pengukuran Kinerja Klasifikasi	23
Tabel 2.3 Penelitian Terkait	29
Tabel 3.1 Pemilihan Atribut Pada <i>Dataset</i>	40
Tabel 3.2 Data Hasil Transformasi	41
Tabel 4.1 Hasil Tahapan Seleksi	47
Tabel 4.2 Transformasi Data	49
Tabel 4.3 <i>Confusion matrix</i> algoritma Naive Bayes	58
Tabel 4.4 <i>Confusion Matrix</i> algoritma <i>Decision Tree C4.5</i>	58
Tabel 4.5 Hasil <i>performance</i> algoritma Naive Bayes dan <i>Decision Tree C4.5</i>	58
Tabel 4.6 Hasil perbandingan uji T-test.....	58
Tabel 4.7 Pengetahuan yang dihasilkan menggunakan algoritma <i>Decision Tree C4.5</i> ..	60

DAFTAR RUMUS

Rumus 2.1 Formula umum Teorema Bayes	18
Rumus 2.2 Menghitung nilai entropy	21
Rumus 2.3 Menghitung nilai Gain	21

DAFTAR LAMPIRAN

Lampiran 1 <i>Cleaned data</i>	67
Lampiran 2 <i>Transformed Data</i>	71
Lampiran 3 ExampleSet Decision Tree C4.5	74
Lampiran 4 ExampleSet Naive Bayes	78

DAFTAR SINGKATAN

BABS	: Buang Air Besar Sembarangan
CTPS	: Cuci Tangan Pakai Sabun
JSP	: Jamban Sehat Permanen
JSSP	: Jamban Sehat Semi Permanen
KPSPAM	: Kelompok Pengelola Sarana SPAM
PAMSIMAS	: Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat
SAM	: Sarana Air Minum
SAN	: Sarana Sanitasi
SBS	: Stop Buang air besar Sembarangan
SPAM	: Sarana Pengelolaan Air Minum

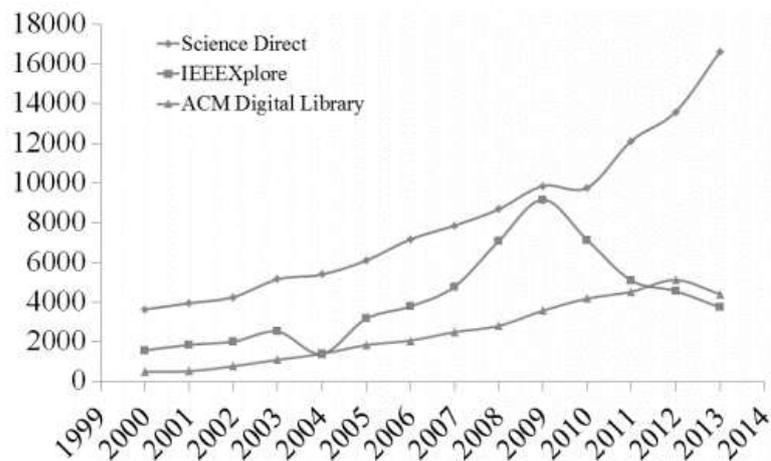
BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam penelitian ini penulis menggunakan proses *data mining* untuk menggali nilai tambah berupa informasi yang selama ini tidak dapat diketahui sebelumnya melalui proses manual dari penggalian suatu data. Proses penggalian *data mining* dilakukan dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari suatu data dan akhirnya dapat digunakan untuk menentukan keputusan dari pengetahuan yang sudah diperoleh.

Secara sadar atau tidak sadar manusia telah memproduksi beragam data yang jumlah dan ukurannya sangat besar di berbagai bidang kehidupan. Didukung oleh kemudahan teknologi yang saat ini ada telah dimungkinkan untuk dapat mengumpulkan data dalam jumlah yang besar dengan waktu yang sangat cepat. Namun disisi lain kecepatan dalam menemukan informasi dan pengetahuan yang berguna dari data, ternyata jauh tertinggal jika dibandingkan dengan kecepatan dalam mengumpulkannya.



Gambar 1.1
Trend perkembangan penelitian jurnal *data mining*
(Matsunaga *et al.*, 2014)

Gambar 1.1 menampilkan informasi berupa grafik tentang peningkatan penelitian terkait *data mining* dari tahun ke tahun. Dari informasi tersebut dapat disimpulkan bahwa penggalian data menjadi *trend* yang terus meningkat. Data harus diolah menjadi suatu

pengetahuan agar lahir nilai tambah dan dapat dimanfaatkan. Sebagai langkah pertama dalam melakukan konversi data secara besar-besaran menjadi sebuah informasi dan pengetahuan yang berguna yaitu dengan melakukan penambangan data dengan menggunakan algoritma yang tersedia pada *data mining*, hal ini untuk mengotomatiskan penambangan berbagai jenis pola data dan selanjutnya dapat dimanfaatkan pada domain-domain tertentu sesuai kebutuhannya masing-masing.

Penelitian ini dilakukan pada PAMSIMAS III Kementerian Pekerjaan Umum dan Perumahan Rakyat (KEMENPUPR). Program ini sudah berjalan dari tahun 2008 dan dilaksanakan untuk mendukung agenda nasional yaitu dalam rangka meningkatkan cakupan penduduk terhadap pelayanan air minum dan sanitasi yang layak dan berkelanjutan.

Dalam proses penetapan desa PAMSIMAS, terlebih dahulu dilakukan sejumlah kegiatan proses perencanaan yang dilaksanakan oleh masyarakat bersama-sama dengan fasilitator PAMSIMAS. Kegiatan perencanaan ini akan menghasilkan sebuah dokumen Rencana Kegiatan Masyarakat yang dijadikan syarat wajib untuk dapat masuk dalam penetapan desa PAMSIMAS yang akan ditetapkan oleh pejabat berwenang. Kemudian setelah desa selesai melaksanakan semua tahapan kegiatan PAMSIMAS dan sudah melakukan kegiatan serah terima sarana, maka desa-desa tersebut memiliki kewajiban untuk melaporkan status kondisi sarana dan perkembangan progres pemanfaatnya secara periodik melalui aplikasi SIM PAMSIMAS.

Jika capaian progres akses terhadap sarana air minum tinggi, maka untuk desa-desa keberlanjutan PAMSIMAS dapat dimungkinkan untuk mendapatkan bantuan dana lagi yaitu berupa dana hibah guna mempercepat akses universal terhadap sarana air minum. Desa-desa keberlanjutan PAMSIMAS tersebut adalah desa yang sudah melaksanakan seluruh kegiatan pembangunan sarana air minum dan sanitasi, desa-desa ini akan masuk sebagai desa paska atau desa keberlanjutan PAMSIMAS yang tetap di *monitoring* progres perkembangan jumlah pemanfaat dan kondisi sarananya. *Trend* dari data-data desa keberlanjutan ini selanjutnya bisa dilakukan penambangan data dengan

menggunakan teknik *data mining* untuk mendapatkan pengetahuan yang berguna dan dapat dimanfaatkan untuk melakukan klasifikasi akses sarana air minum.

Penggalian pengetahuan ini dilakukan dengan proses *data mining* menggunakan model klasifikasi dengan algoritma Naïve Bayes dan *Decision Tree* C4.5. Data akan diproses menggunakan Rapidminer sebagai *tool data mining* untuk mempermudah dalam pengolahan data, sehingga *dataset* yang besar akan berubah menjadi pengetahuan dan selanjutnya dapat dijadikan informasi untuk proses penetapan klasifikasi akses sarana air minum pada PAMSIMAS III.

Salah satu dasar penggunaan algoritma Naive Bayes dan Decision Tree C4.5 pada *dataset* PAMSIMAS yaitu dikarenakan jenis karakteristik data yang terdapat pada *dataset* tersebut. Dimana *dataset* PAMSIMAS memiliki beberapa atribut yang terdiri dari beberapa jenis data nominal dan terdapat satu atribut label atau atribut target yang jenis datanya berupa data nominal. Sedangkan penggunaan algoritma Naive Bayes pada penelitian ini dikarenakan kemampuannya yang sangat sederhana untuk diterapkan dan hasil yang baik telah diperoleh dalam banyak kasus serta dapat dengan mudah melakukan skala pada *dataset* yang besar dengan waktu yang *linier*, hal ini berbeda jika dibandingkan dengan beberapa algoritma klasifikasi lainnya yang terlalu lama karena proses pengolahan datanya dilakukan secara berulang. Sedangkan penggunaan Decision Tree C4.5 adalah kelebihanannya dalam membuat pola pohon keputusan yang dibentuk berdasarkan rasio *gain* tertinggi sehingga dengan pohon keputusan ini akan dengan mudah memahami data dari setiap atributnya. C4.5 merupakan versi terbaru dari Decision Tree yang versi sebelumnya adalah Decision Tree ID3. Oleh karena itu tentu Decision Tree C4.5 merupakan versi terbaik jika dibandingkan dengan versi Decision Tree sebelumnya. Kedua algoritma ini merupakan *predictive model* dan *supervised learning* yang dapat memetakan pengamatan tentang suatu *record* kepada suatu kesimpulan untuk menentukan nilai target.

Klasifikasi akses pada sarana air minum ini sangat penting dilakukan untuk dapat memproyeksikan kemungkinan dapat tercapainya 100% akses sarana air minum atau universal akses pada desa-desa PAMSIMAS berdasarkan pengetahuan yang telah dihasilkan melalui proses *data mining*. Proses *data mining* dapat membantu melakukan

penyeleksian secara otomatis dan terkomputerisasi dengan menggunakan salah satu aplikasi *data mining* seperti Rapidminer yang dapat mengurangi terjadinya *human error* pada proses klasifikasi akses sarana air minum desa-desa PAMSIMAS.

Data Mining adalah sebuah langkah opsional yang paling tepat dan dapat digunakan pada data PAMSIMAS guna melakukan penyeleksian desa-desa keberlanjutan yang akan mendapatkan bantuan dana hibah untuk mempercepat kemungkinan untuk dapat tercapainya akses universal pada sarana air minum. Sehingga penggunaan *data mining* ini diharapkan dapat membantu dalam proses klasifikasi desa-desa yang diprediksi memiliki probabilitas untuk dapat mencapai akses sarana air minum 100%.

Oleh karena itu, atas dasar pemikiran tersebut di atas, penelitian ini akan melakukan penerapan *data mining* pada data PAMSIMAS dengan menggunakan model klasifikasi dan memanfaatkan algoritma Naive Bayes dan algoritma *Decision Tree* C4.5 untuk menentukan klasifikasi akses sarana air minum desa-desa keberlanjutan PAMSIMAS. Penulis menetapkan judul tesis yaitu Klasifikasi Akses Sarana Air Minum Menggunakan Algoritma Naïve Bayes dan *Decision Tree* C4.5 (Studi Kasus pada PAMSIMAS III Kementerian Pekerjaan Umum dan Perumahan Rakyat).

1.2 Rumusan Masalah

Dari beberapa uraian yang dikemukakan pada latar belakang, penulis akan mengangkat beberapa rumusan masalah mengenai penerapan *data mining* pada data desa-desa keberlanjutan PAMSIMAS III. Adapun rumusan masalah tersebut adalah sebagai berikut:

1. Bagaimana algoritma Naive Bayes dan algoritma *Decision Tree* C4.5 diimplementasikan untuk menentukan klasifikasi akses sarana air minum pada data desa keberlanjutan PAMSIMAS?
2. Algoritma apa yang paling tinggi tingkat akurasi untuk klasifikasi akses sarana air minum PAMSIMAS?
3. Atribut-atribut apa saja yang dapat mempengaruhi klasifikasi dan menentukan pola akses sarana air minum?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, penelitian ini akan membahas tujuan penelitian. Diantaranya adalah sebagai berikut:

1. Menganalisa cara keputusan klasifikasi akses sarana air minum dengan algoritma Naive Bayes dan algoritma *Decision Tree* C4.5 berdasarkan data desa keberlanjutan PAMSIMAS.
2. Menentukan algoritma yang paling tinggi tingkat akurasi pada klasifikasi akses sarana air minum PAMSIMAS.
3. Melakukan identifikasi faktor atau atribut yang mempengaruhi klasifikasi akses sarana air minum PAMSIMAS.

1.4 Ruang Lingkup Penelitian

Berdasarkan tujuan penelitian yang telah dijelaskan, penelitian ini akan membahas ruang lingkup penelitian. Adapun ruang lingkup pembahasan dalam tesis adalah sebagai berikut:

1. Pengambilan data untuk penelitian ini diperoleh dari aplikasi SIM PAMSIMAS yang terdapat pada portal resmi PAMSIMAS dan merupakan satu-satunya sumber resmi data PAMSIMAS III. Data yang digunakan terhitung untuk desa-desa yang telah melaksanakan kegiatan PAMSIMAS pada tahun 2008 sampai dengan tahun 2020.
2. Metode yang digunakan adalah klasifikasi dengan memanfaatkan algoritma Naive Bayes dan algoritma *Decision Tree* C4.5 sebagai proses pemodelan *data mining* guna melakukan klasifikasi akses sarana air minum pada desa-desa keberlanjutan PAMSIMAS. Penggunaan algoritma Naive Bayes dan *Decision Tree* C4.5 diterapkan berdasarkan tipe karakter datanya serta akurasi yang baik. Sedangkan untuk pemilihan atribut didasarkan pada keputusan program berdasarkan keadaan *real* di lapangan.

1.5 Metode Penelitian

Penulisan tesis ini menggunakan metode yang memiliki urutan sebagai berikut:

1. Identifikasi Permasalahan

Berisi tentang permasalahan–permasalahan yang akan dirumuskan dan dikembangkan.

2. Studi Literatur

Berisi tentang uraian teori, temuan dan bahan penelitian lain yang diperoleh sebagai acuan untuk dijadikan landasan kegiatan penelitian baik bersumber dari jurnal, buku atau *website*.

3. Tujuan dan Ruang Lingkup Penelitian

Berisi tentang tujuan yang akan dicapai terhadap permasalahan dan batasan objek yang akan dibahas dan diteliti.

4. Pengumpulan Data

Berisi tentang teknik pengumpulan data yang akan dijadikan *dataset* pada proses *data mining* melalui *data preprocessing*, sehingga format data dapat disesuaikan dengan algoritma yang akan digunakan.

5. Metode *Data Mining*

Berisi tentang proses *data mining* menggunakan metode klasifikasi menggunakan algoritma Naive Bayes dan algoritma *Decision Tree C4.5*.

6. Evaluasi

Berisi tentang kesimpulan dari hasil proses *data mining* berupa lahirnya suatu pengetahuan.

1.6 Sistematika Penulisan

Berdasarkan latar belakang yang telah dijelaskan, penelitian ini akan membahas sistematika penulisan. Di antaranya sebagai berikut:

BAB I : Pendahuluan

Bab ini berisi latar belakang masalah terkait dengan penelitian PAMSIMAS III, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, metode penelitian, dan sistematika penulisan.

BAB II : Landasan Teori

Bab ini berisi tentang studi pustaka berupa teori-teori yang mendukung dan digunakan dalam penulisan tesis, baik bersumber dari buku, jurnal ataupun *website*. Teori-teori akan digunakan sebagai landasan pada proses penelitian di bab-bab selanjutnya.

BAB III : Objek dan Metodologi Penelitian

Bab ini menguraikan secara rinci mengenai objek penelitian, gambaran PAMSIMAS III dan metodologi penelitian yang digunakan yaitu *data mining classification rule* algoritma Naive Bayes dan algoritma *Decision Tree C4.5*.

BAB IV : Hasil dan Pembahasan

Pada bab ini akan diuraikan secara rinci mengenai hasil dan pembahasan pada penelitian. Melakukan proses analisa menggunakan *classification rule* algoritma Naive Bayes dan algoritma *Decision Tree C4.5*.

BAB V : Kesimpulan dan Saran

Bab ini berisi tentang kesimpulan dan saran dari hasil penelitian dan hal-hal lain yang dapat dikembangkan dari tesis ini.

BAB II LANDASAN TEORI

2.1 *Data Mining*

Didalam sub bab ini akan dibahas beberapa definisi tentang *data mining* yang berkaitan dalam penelitian ini. Beberapa definisi *data mining* telah dikemukakan oleh beberapa ahli seperti yang telah dikemukakan oleh (Bulolo, 2020:5) berikut ini:

Data mining atau kadang disebut juga Knowledge Discovery in Database (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar. Output dalam data mining dapat digunakan sebagai alternatif dalam pengambilan keputusan atau untuk memperbaiki pengambilan keputusan di masa yang akan datang.

Begitupula dengan Brammer yang telah mengemukakan bahwa *data mining* merupakan ilmu yang memiliki potensi untuk memecahkan masalah gelombang data. Sebagaimana berikut:

“Machine learning technology, some of it very long established, has the potential to solve the problem of the tidal wave of data that is flooding around organisations, governments and individuals.” (Bramer, 2016:2)

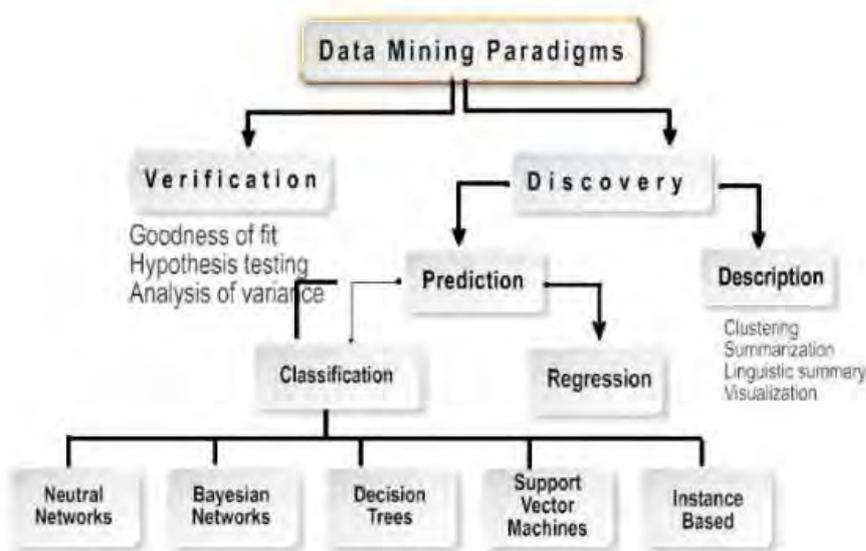
Didukung pula oleh Nong Ye yang menyatakan bahwa *data mining* bertujuan untuk menemukan pola data yang berguna dari sejumlah data yang besar.

“Data mining aims at discovering useful data patterns from massive amounts of data.” (Ye, 2014:3)

Data mining seperti disebutkan di atas sebagai gelombang pasang data yang dimungkinkan dapat membanjiri individu dan kelompok manapun termasuk organisasi pada pemerintahan dan swasta, merupakan kejadian yang sudah lama terjadi. Data saat ini bisa datang dengan cara apapun, dan bahkan sistem komputer *modern* saat ini dapat memungkinkan untuk mengumpulkan data dengan kecepatan yang hampir tak terbayangkan yang berasal dari berbagai sumber. Oleh karena itu data yang muncul tidak bisa dibiarkan sehingga menjadi tidak bermakna, padahal dengan hadirnya data bisa dilakukan proses *data mining* untuk menggali informasi yang tersembunyi berupa pengetahuan yang dapat bermanfaat.

Hal di atas dikuatkan oleh (Karim and Rahman, 2013) yang menurutnya bahwa penemuan pengetahuan dengan *data mining* adalah proses menemukan pola dan hubungan yang sebelumnya tidak diketahui dan berpotensi menarik dalam *database* besar. Prediksi dan keputusan di masa depan dapat dibuat berdasarkan penemuan pengetahuan melalui *data mining*.

“In data mining, the data is stored electronically and the search is automated—or at least augmented—by computer.” (Witten *et al.*, 2017:5)



Gambar 2.1
Pengelompokan Metode *Data Mining*
Sumber: (Wanto *et al.*, 2020:70)

Gambar 2.1 mengilustrasikan pengelompokan metode *data mining* yang secara otomatis mengidentifikasi pola dalam data dan melibatkan metode prediksi dan deskripsi. Metode *discovery* atau penemuan akan secara otomatis mengidentifikasi pola dalam data dan melibatkan metode prediksi dan deskripsi. Metode deskripsi fokus pada pemahaman data yang mendasarinya, dan sementara metode prediksi bertujuan untuk membangun model perilaku untuk mendapatkan sampel baru berupa nilai atau variabel yang terkait dengan sampel tersebut.

Data mining sudah selayaknya dimanfaatkan pada berbagai bidang yang memiliki kumpulan data agar dapat lahir suatu pengetahuan yang dapat digunakan untuk menentukan suatu keputusan ataupun kebijakan. Penggunaan *data mining* secara baik

serta disesuaikan dengan kebutuhan akan memberikan nilai lebih pada proses evaluasi. Jika suatu kebijakan organisasi memanfaatkan *data mining* sebagai proses perencanaan kegiatannya, tentu dapat memberikan landasan yang sangat baik bagi pengembangan organisasi tersebut dalam menentukan suatu keputusan. *Data mining* juga mampu meningkatkan efisiensi dan efektivitas organisasi berdasarkan pengetahuan yang diperoleh.

Data mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak data menjadi pengetahuan berupa pola, rumus, aturan atau model dari suatu data yang besar. *Data mining* bertujuan untuk menemukan pola data yang berguna dari sejumlah kumpulan data. *Data mining* merupakan suatu bidang ilmu yang keberadaannya tidak bisa berdiri sendiri, *data mining* membutuhkan bidang ilmu lainnya seperti statistik, *database*, pencarian informasi, dan *artificial intelligent*.

“The science of extracting useful information from large datasets or databases.”

(Gorunescu, 2011:4)

“Data mining techniques are used to operate on large amount of data to discover hidden patterns and relationships helpful in decision making.”

(Bhardwaj and Pal, 2011)

Dengan adanya data dan kemampuan untuk memanfaatkannya akan menjadi semakin penting bagi organisasi dan individu manapun untuk menjadikan data sebagai sumber utama dalam penggalian informasi dan pengetahuan, sebab saat ini atau di masa yang akan datang dalam mengatasi tantangan dan masalah yang kompleks dapat memanfaatkan data sebagai solusi dan menjadi lebih kompetitif.

Dengan melakukan proses penambangan data dapat memberikan wawasan yang belum pernah terjadi sebelumnya. Dimungkinkan akan lahir keputusan yang lebih informatif seperti memberikan prakiraan dan membantu memecahkan masalah yang semakin kompleks. Saat ini organisasi terus berproses dalam mengumpulkan data dalam jumlah yang terus bertambah dan berasal dari berbagai jenis sumber data, baik yang berasal dari internal ataupun eksternal. Oleh karena itu dibutuhkan alat yang ampuh untuk menguasai analitik data.

Data merupakan suatu fakta yang terekam dan tidak membawa arti jika tidak diubah menjadi informasi yang kemudian akan lahir suatu pengetahuan. *Data mining* adalah tentang memecahkan masalah dengan menganalisis data yang sudah ada dalam *database*. Sebagaimana yang dikatakan oleh Witten:

“Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage.”

(Witten *et al.*, 2017:6)

Proses penggalan data atau *data mining* merupakan data yang disimpan secara elektronik dan dapat dilakukan pencarian secara otomatis atau setidaknya ditambah oleh komputer secara otomatis. *Data mining* juga dikenal sebagai bagian integral dari *Knowledge Discovery in Databases* (KDD) seperti yang dikemukakan berikut ini:

“Data mining is the most instrumental tool in discovering knowledge from transactions. Also data mining is known as an integral part of knowledge discovery in databases (KDD).”

(Moslehi *et al.*, 2011)

“Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.”

(Raval, 2012)

Dalam memperoleh informasi berharga dari data yang sangat banyak, menggunakan teknik *data mining* merupakan cara yang efektif dan telah diterapkan secara luas di berbagai bidang. Pemanfaatan informasi yang telah dihasilkan dapat langsung digunakan sesuai kebutuhan pada domain masing-masing. Hal ini senada dengan apa yang ditulis oleh Aifeng Li berikut ini.

“Data mining (DM) technique is an effective mean to obtain valuable information from mass data and has been widely applied to various fields.”

(Li *et al.*, 2015)

Dari beberapa definisi tentang *data mining* yang telah dijelaskan, berikut hal-hal penting yang terkait dengan *data mining*:

1. Proses *data mining* dilakukan pada kumpulan data yang besar.

2. Proses pengolahan *data mining* menggunakan proses statistika dan matematika atau yang disebut sebagai algoritma.

Untuk memulai proses *data mining* terdapat beberapa langkah yang perlu dilakukan antara lain melakukan pengumpulan *dataset* atau himpunan data, memilih metode *data mining* sesuai tipe karakter datanya, memahami model dan pengetahuan yang sesuai, dan yang terakhir melakukan evaluasi analisis model dan kinerja metode. Adapun rincian dari proses *data mining* sebagaimana berikut dijelaskan pada sub bab berikut.

2.1.1 Peran utama *data mining*

Pada *data mining* sangat banyak teknik penggalian pengetahuan dan algoritma yang dapat dipilih dan disesuaikan dengan tujuan *mining* dan tipe karakter datanya. Semuanya memiliki fungsi dan tujuan yang berbeda-beda tergantung dari kebutuhannya masing-masing. Pada pemilihan metode *data mining* atau pemilihan algoritma wajib disesuaikan dengan *tipe* karakter datanya, baik berupa nominal ataupun numerik, dimana pada tahap ini ditentukan *modeling* atau penggunaan algoritma yang tepat untuk digunakan dan dipilih dari beberapa pilihan model algoritma yang tersedia. Berikut adalah pengelompokan peran *data mining* sebagaimana yang dikemukakan oleh (Bulolo, 2020):

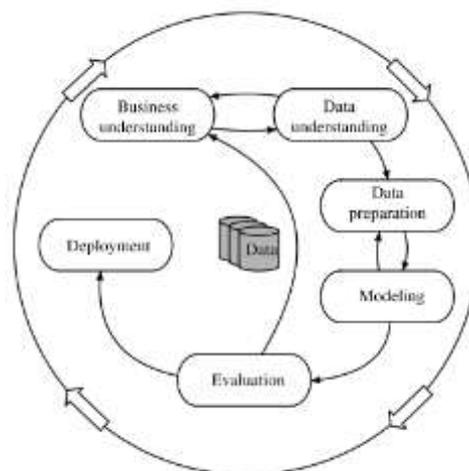
1. *Description* digunakan untuk mengidentifikasi pola yang sering muncul dan mengubah pola tersebut menjadi aturan yang dapat digunakan untuk memudahkan suatu aktivitas.
2. *Classification* digunakan jika atributnya bisa berupa nominal atau numerik tetapi labelnya harus berupa nominal. Pada klasifikasi dilakukan pengelompokan berdasarkan hubungan antara variabel kriteria dengan variabel target.
3. *Prediction* digunakan jika atributnya numerik dan labelnya numerik serta terdapat *time series* pada atributnya. Peran *data mining* prediksi ini secara umum hampir sama dengan klasifikasi.
4. *Estimation* digunakan jika *dataset* atributnya numerik dan labelnya numerik. Estimasi adalah perkiraan atau prediksi sehingga estimasi hampir sama dengan klasifikasi yang membedakan hanya pada jenis *dataset* yang digunakan.

5. *Clustering* digunakan jika *dataset* tidak memiliki label dan atributnya numerik. Pengklasteran merupakan pengelompokkan data yang memiliki kemiripan nilai.
6. *Association* digunakan jika *dataset* ingin diketahui seberapa dekat hubungan antar atributnya. Dalam *data mining* proses asosiasi adalah pencarian atribut yang muncul atau sering muncul dalam waktu yang bersamaan.

Jika peran *data mining* sudah dipahami maka proses *data mining* akan mudah dilakukan untuk dapat melahirkan suatu pengetahuan yang dapat digunakan sesuai dengan kebutuhan. Pengetahuan tersebut merupakan jenis yang dihasilkan dari proses *data mining* atau penambangan data, bisa berupa pola seperti formula, *tree*, *cluster*, *rule* ataupun *correlation*. Teknologi saat ini sangat memungkinkan untuk melakukan penyimpanan data dan menangkap dalam jumlah yang sangat besar. Dari data-data tersebut kemudian bisa dicari pola, *trend* dan anomali dari *dataset* nya, kemudian menyimpulkannya dengan model kuantitatif sederhana. Sehingga *data mining* akan menjadi tantangan tersendiri dalam mengubah data menjadi informasi dan pengetahuan.

2.1.2 Tahapan *data mining*

Dalam menerapkan teknik *data mining* perlu dipahami terlebih dahulu tujuan yang akan dicapai dengan mengimplementasikannya. Sehingga hasil yang didapatkan akan sesuai dengan *domain* yang dibutuhkan. Berikut adalah siklus hidup dari *data mining* sebagaimana yang terdapat pada buku yang di tulis oleh Witten, et al.



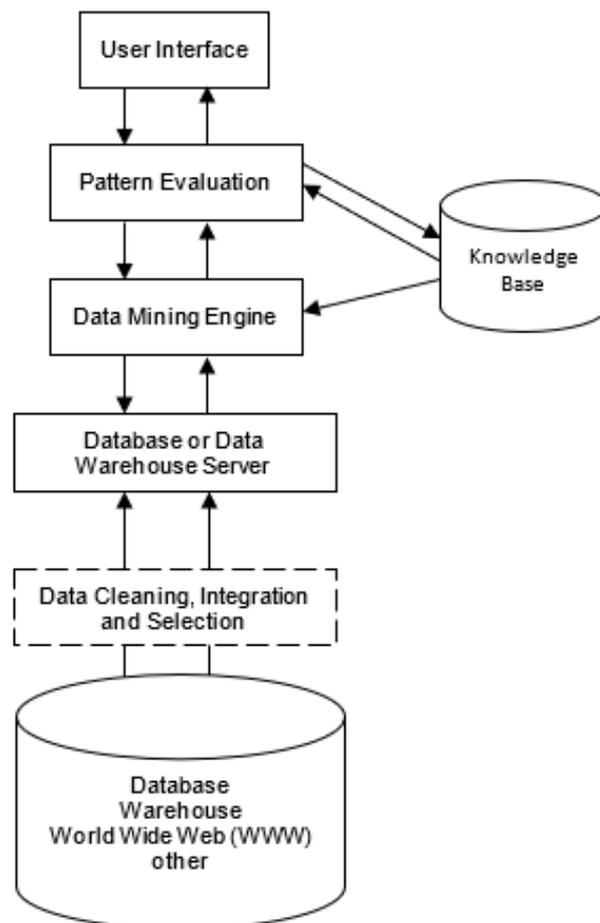
Gambar 2.2
Life Cycle of a Data Mining Project
 Sumber: (Witten et al., 2017:29)

Gambar 2.2 menerangkan tentang siklus hidup dari proyek *data mining*. Pada fase "*business understanding*" akan dilakukan analisa tujuan dan persyaratan bisnis untuk dapat memutuskan apakah *data mining* bisa diterapkan untuk memenuhi tujuan tersebut serta menentukan jenis data apa yang dapat dihimpun untuk membangun model yang dapat diterapkan, sehingga menjadi dasar pemikiran yang sangat fundamental untuk menentukan apakah *data mining* atau KDD diperlukan atau tidak. Selanjutnya adalah fase "*data understanding*" yaitu mempelajari himpunan data awal untuk menentukan pemrosesan selanjutnya. Pada fase ini harus menjamin data dapat memiliki kualitas yang baik sehingga dapat dilakukan penerapan proses *data mining*.

Untuk tiga fase berikutnya yaitu *data preparation*, *modeling*, dan evaluasi merupakan tahapan awal yang melakukan pemrosesan data mentah kemudian melakukan pemodelan dengan menggunakan algoritma, dan akhirnya dapat memberikan wawasan baru. Fase selanjutnya adalah tahapan penyebaran setelah penerapan *data mining* dengan melakukan pemodelan data dengan menggunakan algoritma.

2.1.3 Arsitektur *data mining*

Arsitektur *data mining* menggambarkan aliran proses *data mining* yang dimulai dari data yang digunakan bersumber dari *database*, *warehouse*, ataupun *website* serta sumber-sumber lainnya. Data-data ini membutuhkan volume data historis yang besar agar proses *data mining* dapat berhasil dilakukan. Data-data ini kemudian dilakukan proses penggabungan dan seleksi serta pembersihan untuk membuang data-data yang tidak berguna dan tidak ada kaitannya dengan tujuan yang ingin dicapai. Selanjutnya ditransformasi agar dapat sesuai dengan format yang dibutuhkan pada proses *data mining*. Kemudian data digali agar lahir suatu pengetahuan atau informasi yang diinginkan, selanjutnya ditampilkan agar dapat dilihat oleh *user*. Komponen penting dari sistem *data mining* sebagaimana yang ditunjukkan pada Gambar 2.3.



Gambar 2.3
Arsitektur *Data Mining*
Sumber: Buulolo, 2020:9

2.2 Klasifikasi

Klasifikasi adalah suatu bentuk analisis data yang mengekstrak model yang mendeskripsikan *class* atau label data. Model semacam ini disebut pengklasifikasi atau memprediksi kategori label dari suatu *record*. Klasifikasi juga bisa mengacu pada tugas untuk memprediksi atribut label terhadap atribut lain yang tak berlabel sehingga memiliki keterikatan sebagaimana yang dikemukakan berikut:

“Klasifikasi merupakan teknik dalam data mining untuk mengelompokkan data berdasarkan keterikatan data terhadap data sampel.”

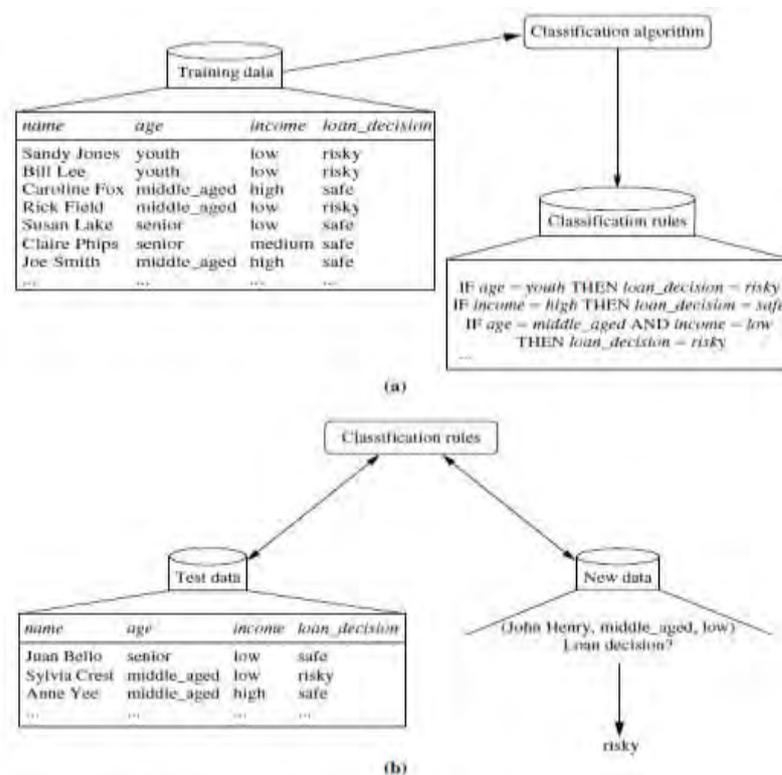
(Oktanisa and Supianto, 2018)

Salah satu teknik *data mining* yang digunakan dalam *supervised learning* adalah klasifikasi. Klasifikasi adalah teknik penambangan data untuk membuat prediksi tentang

nilai-nilai data menggunakan hasil yang diketahui dan ditemukan dari data yang berbeda. Sesuai dengan yang disampaikan oleh (Han *et al.*, 2012:328), bahwa klasifikasi merupakan sebuah model dalam *data mining* dimana, *classifier* dikonstruksi untuk memprediksi *catagorical* label. Contohnya seperti “aman” atau “berisiko” pada data aplikasi peminjaman uang, dan lain sebagainya.

The data analysis task is classification, where a model or classifier is constructed to predict class (categorical) labels, such as “safe” or “risky” for the loan application data; “yes” or “no” for the marketing data; or “treatment A,” “treatment B,” or “treatment C” for the medical data.

Data klasifikasi bekerja menggunakan dua langkah proses; proses data yang pertama adalah proses *learning*, dimana model klasifikasi akan dibangun berdasarkan *data training*, lalu proses pengklasifikasian akan menganalisa atau belajar dari *data training* yang akan menghasilkan suatu *rule* klasifikasi. Langkah proses yang kedua adalah proses pada *rule* klasifikasi, dimana model algoritma yang digunakan akan memprediksi data pada label untuk diterapkan pada *data testing*.



Gambar 2.4
Proses Klasifikasi Data
Sumber:(Han *et al.*, 2012:329)

Gambar 2.4 menunjukkan bahwa untuk gambar (a) merupakan proses *learning* menggunakan *data training* yang dianalisis dengan model klasifikasi. Dimana *loan decision* dijadikan sebagai atribut label *class* dan pengklasifikasi atau model yang dipelajari dan direpresentasikan dalam bentuk aturan klasifikasi atau *classification rule*. Sedangkan untuk gambar (b) yaitu proses klasifikasi yakni data uji digunakan untuk memperkirakan akurasi aturan klasifikasi. Dalam hal ini jika akurasi dapat diterima, maka *rule* atau aturan ini dapat digunakan pada klasifikasi data baru.

Klasifikasi disebut sebagai *supervised learning* karena pada *dataset* terdapat label pada salah satu atributnya. Data atribut pada label ini merupakan target keputusan yang diberikan untuk diprediksi dengan melakukan pembelajaran dari hasil proses pengklasifikasi pada *data learning*. Berbeda dengan *unsupervised learning*, di mana label dari setiap *data learning* tidak diketahui, dan jumlah atau kumpulan *class* yang akan dipelajari mungkin tidak diketahui sebelumnya.

Menurut (Marlina *et al.*, 2016) klasifikasi adalah suatu tindakan yang diberikan kepada suatu data yang berisi sekumpulan atribut, dan salah satunya adalah atribut kelas. Metode klasifikasi ini perlu mencari model yang dapat menjelaskan atribut kelas sebagai fungsi dari atribut masukan. Beberapa algoritma yang dapat digunakan untuk proses klasifikasi diantaranya Naive Bayes dan *Decision Tree C4.5*.

2.2.1 Algoritma Naïve Bayes

Pada klasifikasi Naïve Bayes yang didasarkan pada Teorema Bayes digunakan dalam probabilitas dan statistik untuk menghitung probabilitas bersyarat, sebagaimana yang dikemukakan oleh Vlandari berikut ini:

“Teori keputusan Bayes adalah pendekatan statistik yang fundamental dalam data mining.”
(Vlandari, 2017)

Salah satu tugas dari *data mining* adalah melakukan klasifikasi dengan memanfaatkan algoritma Naive Bayes yang digunakan untuk menentukan keputusan klasifikasi dengan menggunakan probabilitas. Kelebihan penggunaan Naive Bayes menurut (Vlandari, 2017) adalah kemampuannya untuk mengolah data yang beragam seperti data yang konsisten, data tidak konsisten, dan data yang bias. Data yang konsisten

adalah data yang setiap atributnya memiliki nilai dan target yang sama, sedangkan data yang tidak konsisten adalah jika setiap atributnya memiliki target yang sama tetapi nilai pada setiap atributnya berbeda, sedangkan untuk data bias adalah jika data memiliki target atau keputusan yang berbeda namun *instance* pada semua atributnya sama.

Klasifikasi Naïve Bayes dikemukakan oleh seorang ilmuwan berkebangsaan Inggris bernama Thomas Bayes yaitu seorang yang sudah melakukan penelitian awal tentang probabilitas dan teori keputusan selama abad ke-18. Ciri utama dari klasifikasi Naive Bayes adalah adanya asumsi yang kuat atau naif akan independensi dari masing-masing kejadian yang digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

“Algoritma Naïve Bayes classifier adalah suatu teknik dalam data mining di mana dalam pemecahan masalah menggunakan prinsip probabilitas (kemungkinan).”

(Buuolo, 2020:53)

Berikut formula algoritma Naive Bayes yang digunakan untuk mencari nilai peluang sebagaimana berikut (Buuolo, 2020):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Rumus 2.1 Formula umum Teorema Bayes

Penjelasan dari formula tersebut sebagai berikut:

X : Sampel data yang memiliki kelas atau label.

H : Hipotesis X data kelas atau label.

$P(H|X)$: Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti X terjadi.

$P(X|H)$: Probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis H.

$P(H)$: Probabilitas awal hipotesis H terjadi tanpa memandang bukti apapun.

$P(X)$: Probabilitas awal bukti X terjadi tanpa memandang hipotesis/bukti yang lain.

Naïve Bayes mengasumsikan bahwa efek dari nilai atribut pada kelas tertentu tidak tergantung pada nilai atribut lainnya. Asumsi ini disebut kebebasan bersyarat *class* untuk

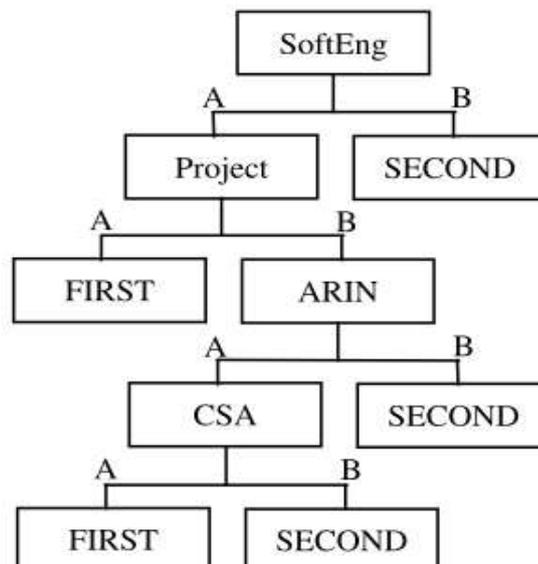
menyederhanakan perhitungan yang terlibat dalam pengertian ini dianggap "Naïve" atau asumsi yang sangat kuat.

2.2.2 Algoritma *Decision Tree* C4.5

Algoritma yang sering digunakan pada teknik klasifikasi lainnya adalah algoritma *Decision Tree* C4.5 yang menggunakan *gain ratio* dan bertujuan untuk mengeksplorasi data untuk diubah kedalam bentuk pohon keputusan agar memudahkan dalam membaca pengetahuan yang dihasilkan seperti yang dikemukakan Vulandari berikut ini:

"Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki." (Vulandari, 2017)

Algoritma ini merupakan *tool* yang sangat bermanfaat untuk melakukan klasifikasi karena strukturnya yang *simple* dan mudah untuk diinterpretasikan. Dari sisi waktu, algoritma *Decision Tree* C4.5 hanya membutuhkan waktu yang tidak terlalu lama (Mantas and Abellán, 2014). Dengan algoritma *Decision Tree* C4.5 akan dibentuk suatu pohon keputusan dengan akar terbalik, dimana akar atau *root* berada pada paling atas sedangkan bagian daun atau *leaf* berada pada bagian bawah (Rahayu *et al.*, 2015). Bentuk pohon keputusan diilustrasikan oleh Bramer pada Gambar 2.5 berikut ini:



Gambar 2.5
Contoh bentuk pohon keputusan untuk data klasifikasi
Sumber: (Bramer, 2016:6)

“A decision tree is a flowchart-like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes or class distributions.”

(Han et al., 2012:18)

Proses algoritma *Decision Tree* C4.5 akan mengubah data tabel menjadi model pohon atau *tree* dan selanjutnya mengubah model pohon tersebut menjadi *rule* atau aturan. Sehingga pada masalah pengambilan keputusan yang dilakukan melalui pemetaan mengenai beberapa alternatif pemecahan masalah maka pohon keputusan ini akan memperlihatkan faktor-faktor probabilitas atau kemungkinan yang akan mempengaruhi alternatif-alternatif keputusan. Pohon keputusan merupakan sebuah metode yang populer dan mudah diinterpretasikan. Mengenai hal ini (Buuololo, 2020) menjelaskan:

Algoritma C4.5 merupakan algoritma yang digunakan untuk mengeksplorasi data yaitu untuk menemukan hubungan variabel input atau attribute kriteria dengan variabel target atau decision attribute (attribute keputusan). Hasil dari eksplorasi data diubah dan bentuk pohon keputusan (decision tree).

Algoritma *Decision Tree* C4.5 ini akan bermanfaat untuk proses pengambilan keputusan karena kemampuannya untuk melakukan *breakdown* pengambilan keputusan yang kompleks menjadi lebih *simple*. Dengan algoritma *Decision Tree* C4.5 akan membentuk pohon keputusan yang merupakan salah satu dalam metode klasifikasi dan prediksi pada proses penambangan data. Algoritma *Decision tree* C4.5 berguna untuk mengeksplorasi data dan menemukan hubungan yang tersembunyi pada data. Sejatinnya pohon keputusan merupakan proses mengubah data pada *dataset* menjadi model pohon, kemudian dari model pohon diubah menjadi *rule*. Pohon keputusan yang dihasilkan merupakan hasil algoritma *Decision Tree* C4.5 dan dapat merepresentasikan serta memodelkan hasil eksplorasi data penting, sehingga pengetahuan atau informasi dari data tersebut lebih mudah diidentifikasi.

Untuk melakukan proses pembentukan pohon keputusan dengan menggunakan algoritma *Decision Tree* C4.5 dibutuhkan nilai *Entropy(S)* dan *Gain(S, A)*. Berikut adalah rumus untuk mencari nilai *Entropy* pada algoritma *Decision Tree* C4.5 sebagaimana berikut.

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

Rumus 2.2 Menghitung nilai entropy

Keterangan:

S = Jumlah sampel data (*sampling*)

n = Jumlah partisi S

p_i = Proporsi dari S_i terhadap S

$Entropy(s)$ merupakan parameter yang digunakan untuk mengukur keberagaman setiap nilai atribut terhadap atribut keputusan dalam suatu himpunan data. Tinggi dan rendahnya nilai $Entropy$ akan mempengaruhi tingkat keberagaman suatu himpunan data. Setelah melakukan proses penghitungan $Entropy$ selanjutnya adalah menghitung nilai $Gain(S, A)$ dengan rumus berikut ini.

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

Rumus 2.3 Menghitung nilai Gain

Keterangan:

S = Jumlah kasus (*sampling*)

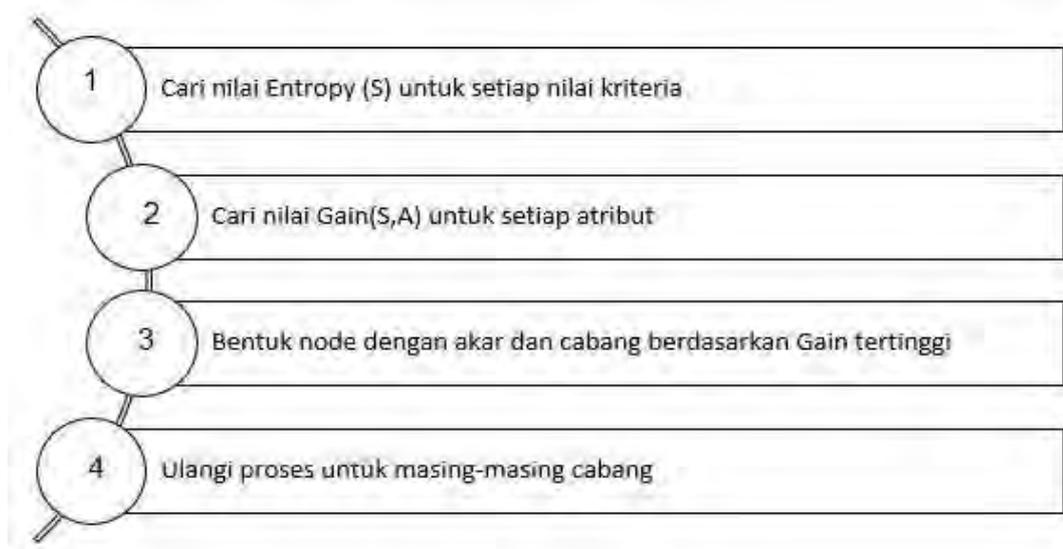
A = Atribut

n = Jumlah partisi S

$|S_i|$ = Jumlah kasus pada partisi ke- i

$|S|$ = Jumlah kasus dalam S

$Gain(S,A)$ berfungsi untuk mengukur efektivitas masing-masing atribut kriteria dalam mengklasifikasikan data. Nilai $Gain$ digunakan sebagai dasar pembentukan *node* atau akar dan cabang pada pohon keputusan. Kesimpulan proses algoritma *Decision Tree* C4.5 bisa dilihat pada gambar berikut:



Gambar 2.6
Pembentukan pohon keputusan menggunakan *Decision Tree C4.5*
Sumber: (Bulolo, 2020:37)

2.2.3 Pengukuran kinerja klasifikasi

Setelah dilakukan proses *data mining* maka selanjutnya dilakukan pengukuran kinerja klasifikasi. Umumnya pada pengukuran kinerja klasifikasi dapat menggunakan *confusion matrix* yang mengindikasikan properti aturan klasifikasi atau melakukan pengukuran performa untuk masalah klasifikasi. *Confusion matrix* dipresentasikan dengan sebuah tabel yang terdapat 4 kombinasi berbeda dari nilai prediksi dan nilai aktual yaitu *true positive*, *true negative*, *false positive* dan *false negative*. *Confusion matrix* bertujuan untuk menghitung akurasi klasifikasi, hal ini senada sebagaimana yang telah dikemukakan oleh Dangare dan Apte berikut ini:

“A *confusion matrix* is obtained to calculate the accuracy of classification. A *confusion matrix* shows how many instances have been assigned to each class.”

(Dangare and Apte, 2012)

Berikut tampilan *confusion matrix* pada Tabel 2.1 berikut ini:

Tabel 2.1 Tabel *Confusion Matrix*

Correct Classification	Classified as	
	+	-
+	<i>True Positives</i>	<i>False Negatives</i>
-	<i>False Positives</i>	<i>True Negatives</i>

Sumber: (Bramer, 2016:90)

Keterangan interpretasi:

True Positives (TP) : Diprediksi positif dan hasilnya benar.

True Negatives (TN) : Diprediksi negatif dan hasilnya benar.

False Positives (FP) : Diprediksi positif dan hasilnya salah.

False Negatives (FN) : Diprediksi negatif dan hasilnya salah.

Selanjutnya adalah mengukur *performance metrics* pada *confusion matrix* dengan menggunakan formula berikut ini:

Tabel 2.2 Tabel Pengukuran Kinerja Klasifikasi

Metrics	Formula	Evaluation Focus
<i>Accuracy</i>	$\frac{TP + TN}{TP + FP + TN + FN}$	Secara umum, metrik akurasi mengukur rasio prediksi yang benar atas jumlah total <i>instance</i> (model) yang dievaluasi.
<i>Precision</i>	$\frac{TP}{TP + FP}$	Presisi digunakan untuk mengukur pola positif yang diprediksi dengan benar dari total pola prediksi dalam kelas positif.
<i>Recall</i>	$\frac{TP}{TP + FN}$	Metrik ini digunakan untuk mengukur pecahan pola positif yang diklasifikasikan dengan benar.

Sumber: (Hossin and Sulaiman, 2015)

Accuracy digunakan untuk menghitung seberapa akurat model yang digunakan dapat mengklasifikasi dengan benar. Sedangkan *precision* melakukan pengukuran terhadap tingkat keakuratan pada pola positif yang diprediksi benar dari total pola prediksi dalam kelas positif yang merupakan hasil prediksi yang diberikan oleh *model*. Pengukuran selanjutnya adalah menggunakan *recall*. *Recall* digunakan untuk mengukur pecahan pola positif yang diklasifikasikan dengan benar.

2.2.4 *Balancing dataset*

Pada umumnya, *dataset* yang akan digunakan bisa dimungkinkan terjadinya kasus tidak seimbang antara kelas positif dan negatif atau kelas minoritas dan kelas mayoritas pada atribut label atau target. Ketidakeimbangan ini akan menyebabkan masalah untuk beberapa klasifikasi pembelajaran pada *data mining*, khususnya jika akan menggunakan *performance metrics* untuk melihat tingkat akurasi dari hasil pemodelan. Akurasi yang akan dihasilkan pada *dataset* yang tidak *balance* akan menjadi akurasi yang semu atau bias, sebagaimana yang dikatakan oleh Patel.

“The term imbalance refers to uneven distribution of data into classes that severely affects the performance of traditional classifiers, that is, classifiers become biased toward the class having larger amount of data.”(Patel et al., 2020)

Teknik *balancing* dapat menggunakan *resampling the training set* dengan cara *over-sampling*, yaitu dengan menggunakan SMOTE Upsampling yang akan bekerja menjadikan data di kelas yang kecil akan seimbang dengan kelas yang besar.

The Synthetic Minority Oversampling Technique (SMOTE) preprocessing algorithm is considered “de facto” standard in the framework of learning from imbalanced data. This is due to its simplicity in the design of the procedure, as well as its robustness when applied to different type of problems.(Fernandez et al., 2018)

2.2.5 Uji T-test

Selain melakukan pengukuran dengan menggunakan *confusion matrix*, pada penelitian ini akan dilakukan uji signifikansi masing-masing *performance* menggunakan operator T-test. Sehingga akan didapatkan model yang terbaik dari algoritma yang digunakan pada penelitian ini.

“T-Test adalah metode pengujian untuk menentukan dua sampel yang tidak berhubungan yang memiliki nilai rata-rata berbeda.” (Rahayuningsih, 2019)

Dengan uji T-test akan melakukan pengujian dengan batasan nilai alpha atau nilai signifikansi yang telah ditentukan. Jika hasil dari uji beda T-test ini menghasilkan nilai diatas batas nilai alpha maka bisa diinterpretasikan bahwa tidak ada perbedaan yang signifikan antara performance dari algoritma Naive Bayes dan Decision Tree C4.5, namun sebaliknya jika hasilnya kurang dari nilai alpha bisa disimpulkan terjadi perbedaan yang signifikan dan distribusi data yang tidak normal.

2.3 Knowledge Discovery In Database (KDD)

Banyak terdapat definisi yang menerangkan tentang *Knowledge Discovery in Database* (KDD), salah satunya sebagaimana yang dikemukakan oleh Dharminder dan Deepak berikut ini:

“Data Mining (as called as Knowledge discovery in databases (KDD), aims at the discovery of useful information from large collections of data.”

(Dharminder Kumar and Deepak Bhardwaj, 2011)

Setiap detik pada aktivitas kehidupan manusia dapat dipastikan selalu menghasilkan data dari mulai kegiatan yang kecil sampai dengan kegiatan yang besar, mulai bangun tidur sampai dengan tidur kembali, data bisa dihasilkan dan dapat digali pengetahuannya untuk dapat dimanfaatkan di berbagai bidang yang sesuai dengan kebutuhan.

Sebelum adanya *Knowledge Discovery in Database (KDD)*, data-data tersebut dianggap tidak bermanfaat dan dibiarkan bahkan dibuang. Namun seiring perkembangan pengetahuan dan teknologi akhirnya dapat merubah cara berpikir seperti ini, dari semula data dianggap sebagai sampah, pada akhirnya data dianggap seperti tambang emas yang dapat memberikan informasi dan pengetahuan yang bermanfaat bagi manusia.

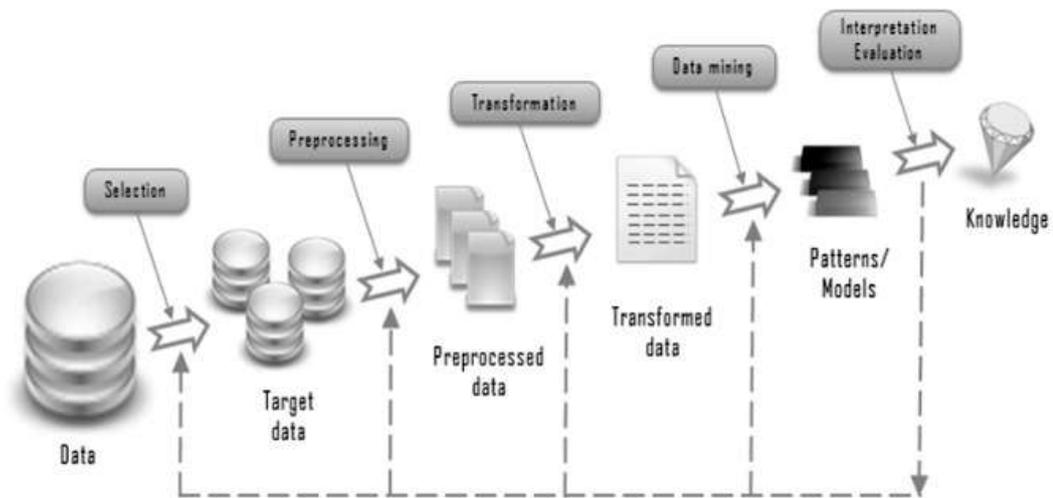
Kumar menjelaskan bahwa *data mining* yang biasanya disebut sebagai KDD bertujuan untuk menemukan informasi yang berguna dari kumpulan data yang besar. Kumpulan data yang besar setelah pengolahan *preprocessing* dilakukan, data disebut sebagai *dataset* atau himpunan data yang siap dilakukan *modeling*. *Data mining* sebenarnya merupakan bagian dari *Knowledge Discovery in Database (KDD)* seperti yang diutarakan oleh Holzinger.

“Data Mining is an important subset of the knowledge discovery process, the current most used notion is Knowledge Discovery and Data Mining (KDD).”

(Holzinger *et al.*, 2014)

Menurut Raval proses *Knowledge Discovery in Database (KDD)* dimulai dari pengumpulan data dan disimpulkan dengan hasil evaluasi. *Knowledge Discovery in Database (KDD)* merupakan tahapan penggalan data yang dapat diilustrasikan seperti pada Gambar 2.7. Untuk mendapatkan hasil yang diharapkan maka proses tahapan harus dilakukan secara berurut. Kesalahan yang dilakukan pada tahapan tertentu akan berakibat pada buruknya hasil yang diperoleh.

Setiap tahapan pada *Knowledge Discovery in Database (KDD)*, masing-masing akan menghasilkan *output* yang jelas. Seperti tahapan pertama yang menghasilkan *output* berupa himpunan data dan diakhiri dengan tahapan evaluasi yang akan menghasilkan suatu pengetahuan dari pemodelan yang sudah dilakukan dengan menggunakan algoritma.



Gambar 2.7
 Proses *Knowledge Discovery in Database* (KDD)
 Sumber: (Raval, 2012)

Pada Gambar 2.7 dapat dijelaskan sebagaimana berikut ini:

1. *Data*

Setelah tujuan *data mining* ditetapkan, pada proses KDD selanjutnya adalah memilih dan mengumpulkan *dataset* atau himpunan data yang akan digunakan. *Dataset* tersebut harus dipisahkan dari data operasional yang masih dapat berubah menggunakan proses *create*, *update* dan *delete*.

2. *Selection*

Diperlukan proses pemilihan data karena tidak semua data yang telah diambil dapat digunakan pada proses *data mining*. Perlu dilakukan pengumpulan *data target*, penentuan variabel dan menyimpannya pada suatu berkas.

3. *Preprocessing*

Proses ini diperlukan untuk melakukan pembersihan data atau *cleaning data* seperti membuang data yang tidak konsisten, data yang *noise*, atau memperbaiki data yang masih terdapat kesalahan.

4. *Transformation*

Pada tahap ini jika *dataset* sudah berhasil disiapkan maka selanjutnya data dilakukan transformasi untuk disesuaikan dengan format yang sesuai pada algoritma yang akan

digunakan pada proses *data mining*. Karena masing-masing algoritma membutuhkan *format* data yang sesuai agar dapat dilakukan penghitungan.

5. *Data Mining*

Proses *data mining* adalah tahapan utama dalam *Knowledge Discovery in Database* (KDD). Seperti yang sudah dijelaskan pada sub bab sebelumnya, *data mining* merupakan ilmu yang dapat digunakan untuk menggali data agar lahir suatu pengetahuan yang berguna dan dapat dimanfaatkan.

6. *Interpretation Evaluation*

Pada tahap ini setelah melakukan proses *data mining* pada data dan sudah melewati berbagai proses seperti penentuan model dan algoritma, maka dari sejumlah data tersebut akan menghasilkan suatu pengetahuan atau informasi yang bermanfaat dan ditampilkan kedalam bentuk yang mudah dimengerti seperti grafik, pohon keputusan ataupun *rule*. Pengetahuan yang dihasilkan dari proses *data mining* harus diperiksa kembali apakah sesuai dengan hipotesis yang sudah ditentukan.

7. *Knowledge*

Inilah tujuan utama dari KDD yaitu mendapatkan pengetahuan atau informasi yang berguna dan mudah dimengerti serta dapat dimanfaatkan sesuai dengan kebutuhan dan domainnya masing-masing.

2.4 Rapidminer

Kemunculan *data mining* sebagai ilmu yang berupaya menggali pengetahuan seperti pola, kaidah, aturan dan informasi berharga lainnya, diikuti pula dengan kemunculan berbagai aplikasi *data mining* sebagai *tool* pengolahan data secara mudah dan berbasis GUI. Aplikasi-aplikasi ini ada yang berbayar dan ada pula yang gratis, untuk yang berbayar, pengguna akan diminta untuk membeli lisensi terlebih dahulu sebelum dapat menggunakannya. Sedangkan yang versi gratis dapat digunakan oleh siapapun tanpa harus membeli lisensi terlebih dahulu. Untuk Rapidminer, walaupun aplikasi ini berbayar, namun telah disediakan aplikasi versi *student* sehingga dapat digunakan secara gratis tanpa mengurangi fungsi yang tersedia pada versi berbayar.

“Rapidminer is the most popular open source software in the world for data mining, and strongly supports text mining and other data mining techniques that are applied in combination with text mining.”

(Hofmann and Klinkenberg, 2014:243)

Pada penelitian ini, penulis menggunakan aplikasi Rapidminer untuk mengolah data menggunakan algoritma Naive Bayes dan algoritma *Decision Tree* C4.5 dengan teknik klasifikasi. Penggunaan aplikasi ini dapat mempermudah dalam mengolah data menggunakan algoritma yang tersedia. Rapidminer yang memiliki basis antarmuka pengguna grafis atau GUI, sangat mudah untuk digunakan pada proses penambangan data secara interaktif.

RapidMiner memungkinkan seseorang untuk merancang proses *data mining* dengan melakukan teknik *drag and drop* secara sederhana dari kotak yang mewakili modul fungsional yang disebut operator dipindahkan ke dalam kotak kerja, lalu operator-operator ini dihubungkan untuk menentukan aliran kontrol yang kompleks dan bersarang tanpa melakukan aktivitas pemrograman.

2.5 PAMSIMAS

PAMSIMAS merupakan singkatan dari Program Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat yang sudah dilaksanakan sejak tahun 2008. Program berbasis masyarakat ini berada di bawah Kementerian Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia yang menggunakan dana bersumber dari APBN dan sharing APBD serta pendanaan non pemerintah lainnya seperti dana CSR. PAMSIMAS bertujuan untuk meningkatkan akses masyarakat terhadap sarana air minum dan sanitasi dasar. (Profil Pamsimas, 2021)

PAMSIMAS dilaksanakan dengan pendekatan berbasis masyarakat melalui keterlibatan masyarakat baik perempuan dan laki-laki, kaya ataupun miskin serta pendekatan yang tanggap terhadap kebutuhan masyarakat. Pendekatan ini dilakukan melalui proses pemberdayaan masyarakat untuk menumbuhkan prakarsa, inisiatif, dan partisipasi aktif masyarakat dalam memutuskan, merencanakan, menyiapkan, melaksanakan, mengoperasikan dan memelihara sarana yang telah dibangun, serta

melanjutkan kegiatan peningkatan derajat kesehatan di masyarakat dan di lingkungan sekolah.

Pada penelitian ini data PAMSIMAS yang akan digunakan untuk proses *data mining* adalah data modul 7.3. Data modul 7.3 adalah data-data desa keberlanjutan yang telah melaporkan keberfungsian sarana dan jumlah masyarakat yang sudah akses sarana air minum serta data-data lainnya yang dilaporkan secara rutin setiap triwulan. Data modul 7.3 diambil dari data lapangan dan dilaporkan oleh fasilitator PAMSIMAS yang bekerja di desa kemudian di *input* melalui aplikasi Sistem Informasi Manajemen PAMSIMAS yang beralamat di <https://www.pamsimas.org>.

2.6 Penelitian Terkait

Setelah *data mining* dipublikasikan sebagai *tool* proses penambangan data, banyak dari para peneliti melakukan proses penelitian sesuai dengan domain dan kebutuhannya masing-masing. Berikut penelitian terkait yang berhubungan dengan penggunaan algoritma Naive Bayes dan algoritma *Decision Tree* C4.5 sebagaimana yang terdapat pada tabel 2.3 berikut ini:

Tabel 2.3 Penelitian Terkait

Penulis	Metode	Objek	Kesimpulan
Karim & Rahman, 2013	Naive Bayes & <i>Decision Tree</i> C4.5	<i>Decision Tree</i> dan Naive Bayes untuk Klasifikasi dan Pengetahuan <i>direct marketing</i>	Untuk menentukan algoritma yang terbaik didasarkan pada tingkat akurasi yang tertinggi. Hasil penelitian menunjukkan <i>Decision Tree</i> memiliki tingkat akurasi yang tertinggi yaitu sebesar 0.9396 dibandingkan Naive Bayes 0.8491.
Rosandy, 2016	Naive Bayes & <i>Decision Tree</i> C4.5	Perbandingan Metode Naive Bayes dan C4.5 dalam menganalisa kelancaran pembiayaan	Disimpulkan berdasarkan tingkat akurasi, bahwa pemodelan metode <i>Decision Tree</i> C4.5 sebesar 56% dan Naive Bayes 50%, sehingga untuk akurasi <i>Decision Tree</i> C4.5 lebih unggul. Namun untuk metode Naive Bayes lebih mudah digunakan karena tidak memiliki alur yang panjang seperti pada <i>Decision Tree</i> C4.5.
Septiani, 2017	<i>Decision Tree</i> C4.5 & Naive Bayes	Komparasi data mining algoritma C4.5 dan Naive Bayes	Disimpulkan bahwa kedua metode ini akurat dalam melakukan prediksi untuk penyakit hepatitis. Menurut pengujian berdasarkan Accuracy, algoritma terbaik adalah Naive Bayes, sedangkan menurut

Penulis	Metode	Objek	Kesimpulan
		untuk prediksi penyakit hepatitis	pengujian berdasarkan ROC Curve (AUC) algoritma yang terbaik adalah Algoritma <i>Decision Tree C4.5</i> .
Annur, 2018	Naive Bayes	Klasifikasi masyarakat miskin	Berdasarkan hasil pengujian <i>confussion matrix</i> dengan teknik <i>split</i> validasi, penggunaan metode klasifikasi Naive Bayes terhadap <i>dataset</i> yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori <i>Good</i> . Sementara nilai <i>Precision</i> sebesar 92% dan <i>Recall</i> sebesar 86%.
Noh, 2019	Naive Bayes	Penentuan golongan pelanggan air minum	Metode Naive Bayes yang diterapkan mampu memberikan hasil sebagaimana mestinya, dengan menampilkan nilai probabilitas peluang dari setiap golongan pada setiap pelanggan air.
Wibowo dkk, 2019	Naive Bayes & <i>Decision Tree C4.5</i>	Komparasi algoritma Naive Bayes dan <i>Decision Tree</i> untuk memprediksi lama studi mahasiswa	Hasil komparasi tingkat akurasi menunjukkan hasil dengan metode Naive Bayes sebesar 30% sedangkan dengan <i>Decision Tree C4.5</i> sebesar 55%. Sehingga dapat disimpulkan dengan <i>Decision Tree C4.5</i> lebih tinggi tingkat akurasinya dibandingkan dengan Naive Bayes.
Gustiana, 2020	<i>Decision Tree C4.5</i>	Sistem pendukung keputusan evkin fasilitator PAMSIMAS	<i>Data Mining</i> sebagai teknik klasifikasi menggunakan algoritma <i>Decision Tree C4.5</i> dapat diterapkan pada evaluasi kinerja fasilitator PAMSIMAS, dengan mengikuti beberapa tahapan dalam KDD.

Perbedaan penelitian ini dengan penelitian sebelumnya adalah sebagai berikut:

1. Objek pada penelitian ini menggunakan data laporan pada desa-desa keberlanjutan tahun 2008-2020 PAMSIMAS III periode Desember 2020 Kementerian Pekerjaan Umum dan Perumahan Rakyat.
2. Penelitian ini fokus pada prediksi klasifikasi akses sarana air minum.
3. Penggunaan algoritma Naive Bayes untuk melakukan prediksi klasifikasi probabilitas berdasarkan tingkat kemunculan peluang pada data akses sarana air minum yang bersifat tidak konsisten dan bias.
4. Penggunaan algoritma *Decision Tree C4.5* untuk menemukan pola pengetahuan akses sarana air minum untuk mensortir hasil prediksi klasifikasi pada Naive Bayes.
5. Melakukan komparasi akurasi metode Naive Bayes dan *Decision Tree C4.5*.

BAB III OBJEK DAN METODOLOGI PENELITIAN

3.1 Profil PAMSIMAS

Dalam penelitian ini, objek penelitian berupa data desa keberlanjutan PAMSIMAS III Kementerian Pekerjaan Umum dan Perumahan Rakyat yang bersumber dari situs resmi PAMSIMAS sebagai satu-satunya pusat pelaporan kegiatan PAMSIMAS. Adapun tujuan penelitian ini dilakukan untuk mengetahui klasifikasi keberfungsian sarana air minum menggunakan teknik *data mining* dan algoritma yang digunakan Naive Bayes dan *Decision Tree C4.5*.

Cara kerja PAMSIMAS menggunakan pola pemberdayaan masyarakat yang menjadikan masyarakat sebagai pelaku utama dalam menjalankan tahapan kegiatan PAMSIMAS, dari mulai kegiatan perencanaan sampai dengan pelaksanaan. Program ini sudah mulai dilaksanakan sejak tahun 2008 dan bersifat keberlanjutan dalam melakukan pemeliharaan sarana dan pemantauan progres akses pemanfaatnya. PAMSIMAS saat ini dibagi menjadi tiga fase, dimulai dari fase pertama yaitu fase PAMSIMAS I yang dilaksanakan pada tahun 2008-2012. Karena keberhasilan PAMSIMAS I dalam mengurangi angka penduduk yang tidak mendapatkan akses sarana air minum untuk keperluan sehari-hari maka program ini berlanjut ke fase PAMSIMAS II yang dilaksanakan pada tahun 2013-2015. Kemudian dilanjutkan kembali pada fase selanjutnya yaitu fase PAMSIMAS III yang dilaksanakan pada tahun 2016 sampai dengan saat ini.

Pada PAMSIMAS III ditargetkan akan mencapai 15.000 desa baru dan lebih dari 27.000 desa lama yang akan mengelola keberlanjutan program di seluruh Indonesia. Sejak awal pelaksanaan dari mulai tahun 2008, PAMSIMAS telah berhasil menambah angka penduduk yang telah mendapatkan akses air minum dan sanitasi yang layak dengan total jiwa kurang lebih 10,4 juta jiwa yang berasal lebih dari 12.000 desa/kelurahan yang tersebar di 32 provinsi di Indonesia.

Pemerintah Indonesia sudah berkomitmen pada kesepakatan *Millennium Development Goals – Water Supply and Sanitation (MDG-WSS)* untuk mencapai target

akses universal 100% penduduk dapat mengakses air minum dan sanitasi dasar. Program ini akhirnya menjadi program andalan nasional bagi pemerintah pusat dan daerah, dengan menggelontorkan dana yang bersumber dari APBN dan APBD serta membuka kesempatan bagi para pihak yang ingin berkontribusi melalui pendanaan non pemerintah sebagai wujud kolaborasi seperti pendanaan CSR dan lainnya.

Melalui PAMSIMAS, sejumlah warga miskin perdesaan dan pinggiran kota terdorong untuk mendapatkan akses layanan air minum dan sanitasi guna meningkatkan perilaku hidup bersih dan sehat. PAMSIMAS memiliki pola kegiatan yang bersifat pemberdayaan masyarakat dengan tujuan menyediakan layanan sarana air minum yang terjangkau dari segi jarak dan biaya. Program ini masih terus ditingkatkan dan dilanjutkan di desa-desa di Indonesia yang masih memiliki tingkat akses air minum dan sanitasi yang rendah namun dengan syarat ketersediaan sumber air.

PAMSIMAS dalam proses kegiatannya sangat mengoptimalkan pendekatan terhadap keterlibatan masyarakat baik laki-laki atau perempuan, kaya atau miskin serta pendekatan tanggap terhadap kebutuhan masyarakat. Kedua pendekatan ini dilakukan dengan metode pemberdayaan masyarakat dengan tujuan untuk dapat menumbuhkan partisipasi masyarakat dalam proses kegiatan PAMSIMAS. Hal ini dilandasi oleh faktor keberlanjutan yang harus diusung oleh masyarakat setelah sarana terbangun agar dapat berkembang dan cepat mencapai target akses universal.

3.1.1 Tujuan dan sasaran PAMSIMAS

Berikut adalah tujuan dan sasaran PAMSIMAS:

1. Tujuan PAMSIMAS

Program ini bertujuan untuk membantu masyarakat yang kurang terlayani dan berpendapatan rendah untuk dapat mengakses air minum dan sanitasi serta meningkatkan perilaku hidup bersih dan sehat guna mencapai target MDGs pada bidang air minum dan sanitasi dan akses universal pada tahun 2030.

2. Sasaran PAMSIMAS

Agar tujuan di atas dapat terwujud maka PAMSIMAS memiliki *Key Performance Indicator* (KPI) sebagai alat ukur kemajuan program, yaitu:

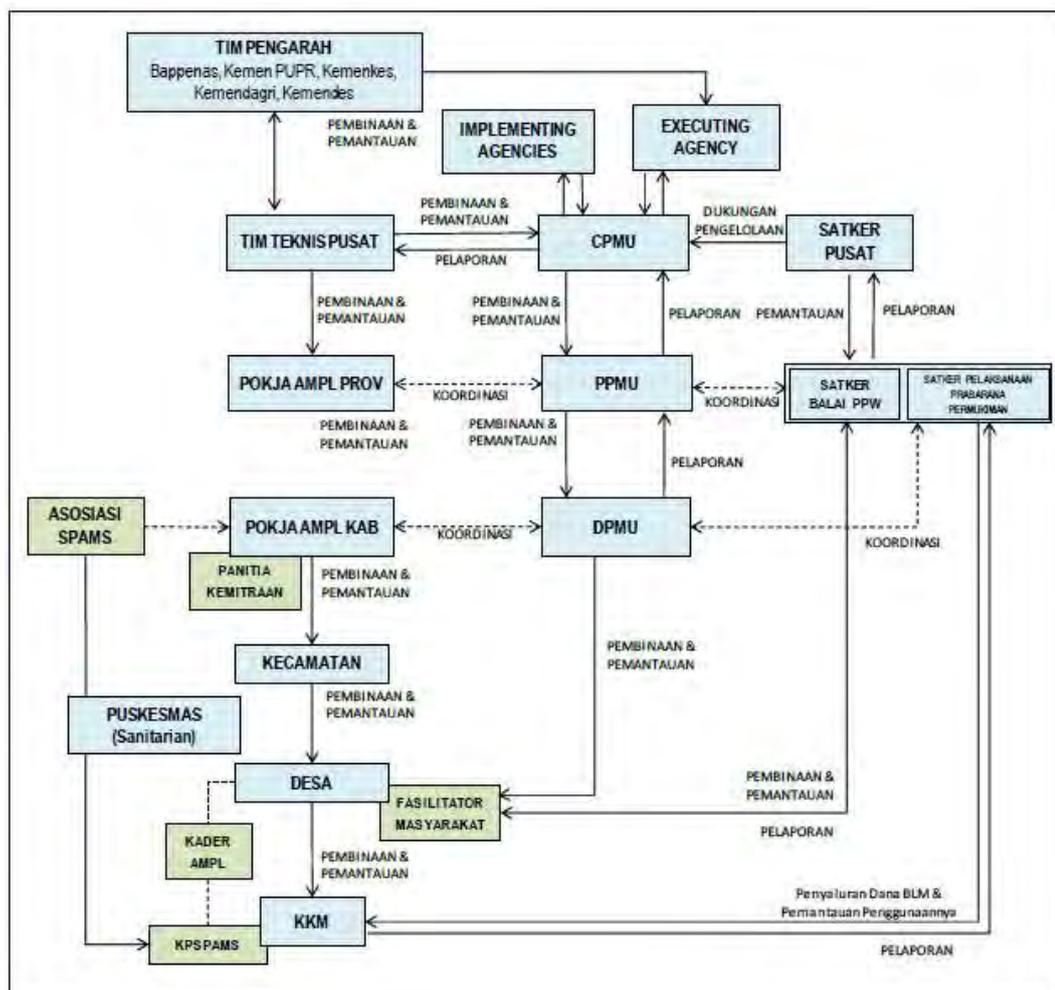
- a. Terdapat tambahan 22.1 juta penduduk yang dapat mengakses sarana air minum aman dan berkelanjutan;
- b. Terdapat tambahan 14.9 juta penduduk yang dapat mengakses sarana sanitasi yang layak dan berkelanjutan;
- c. Minimal 60% masyarakat menerapkan stop Buang Air Besar Sembarangan (BABS);
- d. Minimal 70% masyarakat mengadopsi perilaku program Cuci Tangan Pakai Sabun (CTPS);
- e. Minimal 70% Pemerintah Kabupaten memiliki dokumen perencanaan daerah bidang air minum dan sanitasi;
- f. Minimal 60% Pemerintah Kabupaten mempunyai peningkatan belanja di bidang air minum dan sanitasi.

3. Sasaran lokasi program

Pemilihan desa-desa sasaran PAMSIMAS dilakukan oleh Pemerintah Kabupaten melalui proses kegiatan sosialisasi PAMSIMAS di tingkat kabupaten yang mengundang seluruh desa yang berminat untuk mengikuti PAMSIMAS. Desa yang berminat pada kegiatan PAMSIMAS, akan dimulai kegiatan sosialisasi PAMSIMAS di tingkat desa dan dilanjutkan dengan serangkaian proses tahapan kegiatan perencanaan PAMSIMAS lainnya sesuai dengan tahapan yang sudah ditetapkan pada Pedoman Umum (PEDUM) PAMSIMAS. Sebelumnya untuk kabupaten yang akan ditetapkan sebagai lokasi PAMSIMAS adalah kabupaten yang belum mencapai 100% akses cakupan pelayanan air minum aman perdesaan. Penetapan kabupaten sasaran dilakukan oleh Pemerintah Pusat berdasarkan minat yang diusulkan dari Pemerintah Kabupaten. Untuk penetapan desa PAMSIMAS, desa-desa yang menggunakan sumber dana APBD maka penetapan desa PAMSIMAS cukup dilakukan oleh Kepala Daerah setempat, sedangkan untuk desa yang menggunakan dana yang bersumber dari APBN maka penetapan akan dilakukan oleh Direktorat Jenderal Cipta Karya (DJCK) Kementerian Pekerjaan Umum dan Perumahan Rakyat (KEMENPUPR).

3.1.2 Struktur organisasi PAMSIMAS

PAMSIMAS dikelola secara bersama-sama oleh Pemerintah Pusat, Pemerintah Daerah, Pemerintah Desa dan masyarakat dengan menjangkau seluruh kabupaten provinsi di Indonesia kecuali Provinsi DKI Jakarta. Untuk pembagian peran dan tanggung jawab kegiatan PAMSIMAS, serta jalur koordinasi dan pelaporan dibuat secara jelas dan terstruktur. Untuk bagan struktur organisasi PAMSIMAS dapat ditunjukkan pada Gambar 3.1.



Gambar 3.1
Struktur Organisasi Pengelola dan Pelaksana PAMSIMAS
(DJCK, 2021:38)

Gambar 3.1 mendeskripsikan susunan organisasi PAMSIMAS yang menerangkan pembagian peran dan tanggungjawab dimulai dari tingkat pusat, provinsi dan kabupaten

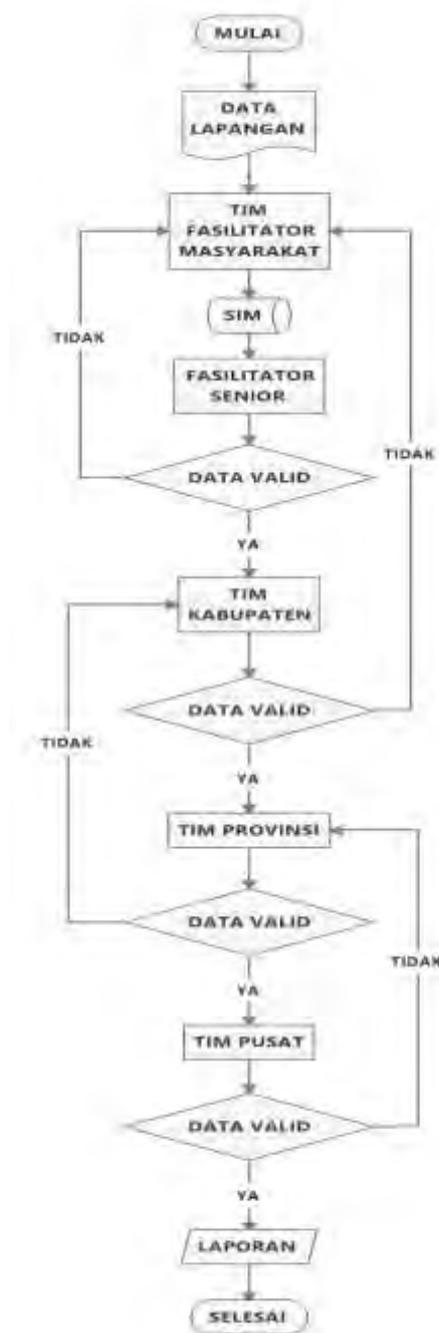
sampai dengan tingkat desa serta masyarakat sebagai pengelola sarana air minum yang sudah dibangun. Pada program ini juga terdapat peran konsultan dan fasilitator sebagai tenaga pendamping yang membantu masyarakat dalam perencanaan dan pelaksanaan PAMSIMAS.

3.1.3 Proses bisnis

PAMSIMAS, sebuah program yang dilaksanakan dalam jangka waktu yang cukup panjang dengan cakupan wilayah yang cukup luas, maka program menerapkan Sistem Informasi Manajemen (SIM) terpadu berbasis *web* yang dapat menyajikan informasi dan laporan secara *real time* serta dilengkapi dengan fasilitas pengisian data secara *online* dengan tujuan untuk meningkatkan efisiensi waktu dan biaya dalam pelaporan data pelaksanaan kegiatan PAMSIMAS III di lapangan.

Alur proses bisnis ini untuk merancang alur data dan tingkat validasinya yang dilakukan secara bertahap dan cukup ketat untuk menghasilkan data yang bersumber dari lapangan dan dapat disajikan sebagai laporan yang dapat digunakan oleh *stakeholder*. Penggunaan sistem informasi berbasis *online* pada PAMSIMAS menjadi sangat penting agar program dapat di *monitoring* dengan baik.

Pihak-pihak yang terlibat langsung pada program ini berasal dari berbagai kalangan, mulai dari pusat sebagai pengelola kebijakan sampai dengan masyarakat di desa sebagai pengelola dan penerima sarana PAMSIMAS. Sehingga penggunaan Sistem Informasi Manajemen PAMSIMAS akan menjadi sumber data PAMSIMAS yang sangat penting untuk pengendalian program.



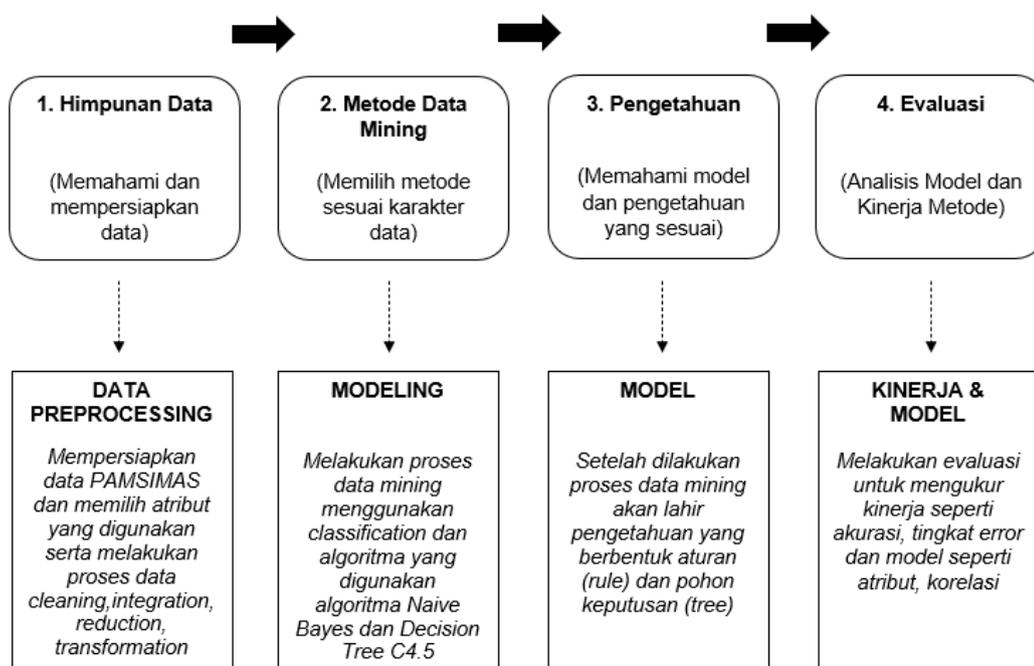
Gambar 3.2
 Proses Bisnis Pelaporan Data SIM PAMSIMAS
 Sumber: (NMC, 2017)

Gambar 3.2 memperlihatkan alur proses pelaporan data SIM PAMSIMAS yang bersumber dari data lapangan berupa data dari Pemerintahan Desa dan instansi kesehatan serta dari lembaga resmi PAMSIMAS di desa seperti Kelompok Keswadayaan Masyarakat (KKM) dan Kelompok Pengelola Sarana Prasarana Air Minum dan Sanitasi

(KPSPAMS). Selanjutnya data oleh Tim Fasilitator Masyarakat (TFM) akan di *input* ke dalam aplikasi SIM PAMSIMAS. Data yang sudah di *input* oleh fasilitator selanjutnya akan di verifikasi oleh Fasilitator Senior (FS), jika sudah *valid* akan di verifikasi kembali oleh *District Coordinator* atau Koordinator Kabupaten. Data yang sudah lolos verifikasi ditingkat kabupaten selanjutnya akan di verifikasi kembali di tingkat provinsi oleh seluruh tenaga ahli tingkat provinsi sesuai dengan bidang keahliannya masing-masing. Data yang sudah lolos verifikasi ditingkat provinsi selanjutnya akan diverifikasi kembali oleh tim pusat.

3.2 Metodologi Penelitian

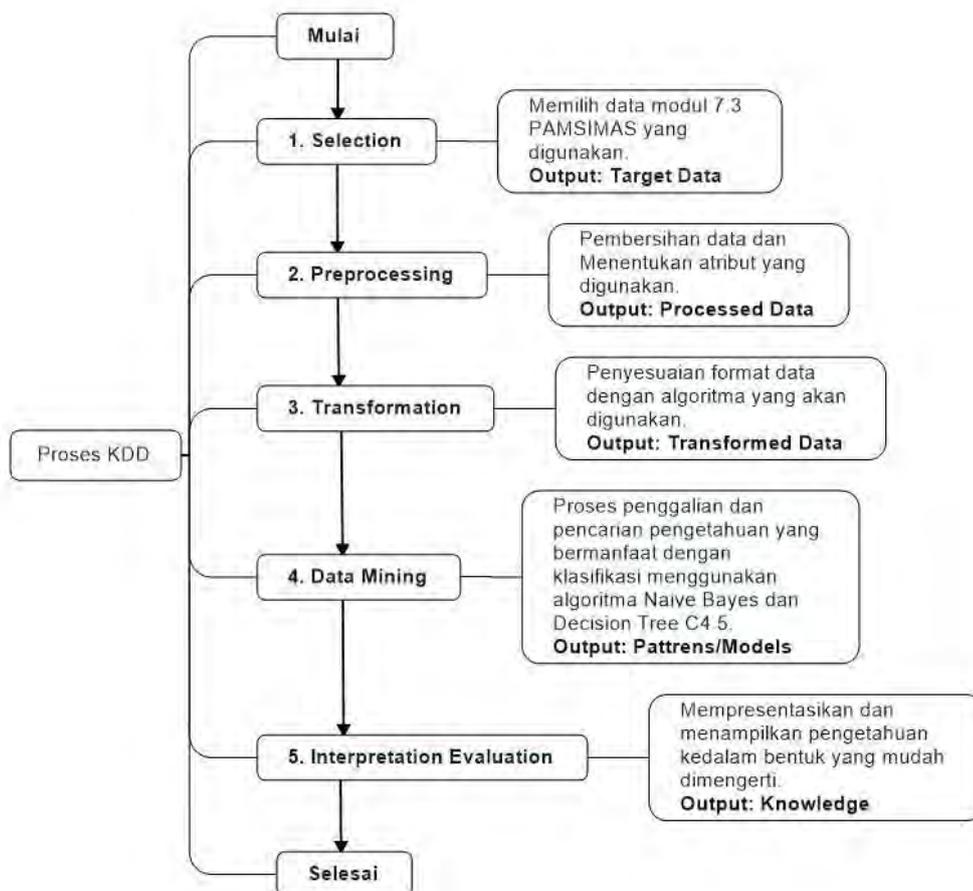
Metodologi penelitian *data mining* merupakan kegiatan untuk melakukan penggalian pengetahuan atau informasi yang dikenal sebagai *knowledge discovery*. Penelitian ini menggunakan alur proses *Knowledge Discovery In Databases* (KDD), sedangkan algoritma yang digunakan Naive Bayes dan *Decision Tree* C4.5 yang akan mengeksekusi *dataset* desa-desa keberlanjutan PAMSIMAS III. Berikut proses *data mining* ditunjukkan dalam gambar 3.3.



Gambar 3.3
Proses *Data Mining* Pada *Dataset* PAMSIMAS

Proses *data mining* pada Gambar 3.3 merupakan tahapan proses yang ada dalam *Knowledge Discovery in Database* (KDD). Tahap 1 adalah memahami dan mempersiapkan data sehingga akan menjadi suatu *dataset* yang melewati proses *data preprocessing*. Tahap 2 memilih *modeling* yang sesuai dengan tipe karakter data, pada penelitian ini akan menggunakan *classification rule* dengan algoritma Naive Bayes dan *Decision Tree* C4.5. Tahap 3 merupakan pengetahuan yang dihasilkan oleh algoritma Naive Bayes berupa probabilitas dan algoritma *Decision Tree* C4.5 berupa *rule* dan pohon keputusan. Tahap 4 melakukan evaluasi kinerja untuk mengukur tingkat akurasi dan lainnya serta evaluasi model untuk menganalisa atribut dan faktor lainnya.

Proses *data mining* di atas yang merupakan bagian dari proses KDD, berikut dijelaskan proses penelitian pada Gambar 3.4.



Gambar 3.4
Tahapan Penelitian Berdasarkan KDD

3.2.1 Persiapan penelitian

Pada tahap ini penelitian melakukan studi literatur untuk mengetahui teori-teori yang berkaitan dengan topik yang akan diteliti khususnya mengenai hal-hal yang terkait dengan *data mining* dan PAMSIMAS. Selanjutnya mempersiapkan data yang akan digunakan dengan mengunduh melalui portal resmi PAMSIMAS III. Data yang diunduh adalah data desa-desa keberlanjutan PAMSIMAS periode 2008-2020 yang akan digunakan untuk proses *data training* menggunakan algoritma Naive Bayes dan *Decision Tree C4.5*.

Hasil studi pustaka yang dilakukan telah menghasilkan informasi bagaimana algoritma Naive Bayes dan *Decision Tree C4.5* dapat diimplementasikan pada penelitian ini untuk menentukan klasifikasi akses sarana air minum desa PAMSIMAS III dengan menggunakan atribut dan label pada *dataset* yang sesuai dengan tujuan pengetahuan yang ingin dihasilkan pada penelitian dengan menggunakan tahapan KDD. Dengan dilakukannya proses *data mining* KDD adalah untuk menganalisis cara menentukan klasifikasi akses sarana air minum menggunakan algoritma Naive Bayes dan algoritma *Decision Tree C4.5* pada data desa-desa keberlanjutan PAMSIMAS.

3.2.2 Seleksi data

Pada tahap ini dilakukan proses pemilihan data sebagai target objek penelitian dan membuang data-data yang tidak dapat dipergunakan, karena tidak semua data yang diambil dapat digunakan dalam penelitian ini dan oleh karena itu dilakukan proses pemilihan data-data yang terkait saja. Data telah diunduh melalui *portal* resmi program PAMSIMAS yaitu data-data desa keberlanjutan PAMSIMAS tahun 2008 sampai dengan tahun 2020 periode Desember 2020.

Data modul 7.3 merupakan sekumpulan data laporan triwulan yang dilaporkan oleh desa-desa keberlanjutan yang sudah membangun sarana PAMSIMAS. Data modul 7.3 desa-desa keberlanjutan akan dijadikan sebagai *data training* dan *data testing* pada penelitian ini. Data-data ini akan dilakukan proses *modeling* pada algoritma Naive Bayes dan algoritma *Decision Tree C4.5* sehingga akan dihasilkan suatu pengetahuan yang bermanfaat.

3.2.3 Praproses data

Pada tahap ini dilakukan proses *data cleaning* dengan memeriksa data yang tidak konsisten dan memperbaiki kesalahan pada data seperti kesalahan antara *record* dan atributnya yang tidak sesuai. Setelah dilakukan *cleaning data* yang sebelumnya berjumlah 29.819 *record* menjadi 29.378 *record*. Beberapa *record* dihapus karena tidak memenuhi syarat untuk proses selanjutnya. Pada *dataset* ini terdapat 35 atribut yang terdiri dari rencana kerja, pembukuan, daftar aset, iuran, jumlah KK, jumlah jiwa, jumlah perempuan, dan data-data lainnya.

Selanjutnya pada praproses data, dilakukan seleksi atribut yang berkaitan dengan akses sarana air minum, hal ini dikarenakan tidak semua atribut yang ada terdapat hubungan langsung dengan sarana air minum dan tidak dapat digunakan untuk proses selanjutnya. Berikut tabel 3.1 untuk melihat beberapa atribut yang akan digunakan dan tidak digunakan pada proses *data mining*.

Tabel 3.1
Pemilihan Atribut Pada *Dataset*

Atribut	Keterangan	Tipe Data
Tahun	Tidak	Nominal
Provinsi	Tidak	Nominal
Kabupaten	Tidak	Nominal
Kecamatan	Tidak	Nominal
Desa	Tidak	Nominal
Pengelolaan SPAM (Sarana Penyediaan Air Minum)	Ya	Nominal
Rencana Kerja KPSPAM	Ya	Nominal
Administrasi Pembukuan	Ya	Nominal
Daftar Aset	Ya	Nominal
Iuran	Ya	Nominal
Jumlah KK	Tidak	Numerik
Jumlah Jiwa	Tidak	Numerik
Jumlah Jiwa Perempuan	Tidak	Numerik
Jumlah Jiwa Disabilitas	Tidak	Numerik
Akses SAM (Sarana Air Minum) KK	Tidak	Numerik
Akses SAM (Sarana Air Minum) Jiwa	Tidak	Numerik
Akses SAM (Sarana Air Minum) Disabilitas	Tidak	Numerik
Sambungan Rumah dengan Meter	Tidak	Numerik
Sambungan Rumah tanpa Meter	Tidak	Numerik
Belum Akses SAM (Sarana Air Minum) KK	Tidak	Numerik
Belum Akses SAM (Sarana Air Minum) Jiwa	Tidak	Numerik
JSP (Jamban Sehat Permanen)	Tidak	Numerik
JSSP (Jamban Sehat Semi Permanen)	Tidak	Numerik

Atribut	Keterangan	Tipe Data
Jamban Sharing	Tidak	Numerik
BABS KK	Tidak	Numerik
Dusun Intervensi	Tidak	Numerik
Dusun SBS	Tidak	Numerik
Dusun CTPS	Tidak	Numerik
Kondisi SAM (Sarana Air Minum)	Ya	Nominal
Kondisi SAN (Sarana Sanitasi)	Tidak	Nominal
Kondisi CTPS (Sarana Cuci Tangan Pakai Sabun)	Tidak	Nominal
Kualitas Air	Ya	Nominal
Rencana Kemitraan	Ya	Nominal
Angka Diare	Tidak	Numerik
Universal Akses	Ya	Nominal

3.2.4 Transformasi data

Pada fase transformasi ini yang dilakukan adalah mentransformasikan *dataset* ke dalam bentuk format yang sesuai dengan algoritma yang akan digunakan, sehingga *dataset* siap untuk dilakukan proses *data mining*. Berikut adalah hasil transformasi data yang dapat dilihat pada tabel 3.2.

Tabel 3.2
Data Hasil Transformasi

Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	...	Akses Universal
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Tidak Ada	Ada	...	Belum
...
Mandiri	Ada	Ada	Ada	...	Belum
Mandiri	Ada	Tidak Ada	Ada	...	Belum

3.2.5 Data mining

Setelah selesai melakukan transformasi pada *dataset* modul 7.3 desa-desa keberlanjutan atau desa paska PAMSIMAS yang diperoleh dari aplikasi SIM PAMSIMAS melalui portal resminya. Maka untuk tahap selanjutnya adalah melakukan *modeling* pada *dataset* dengan menggunakan algoritma yang sesuai untuk mendapatkan pengetahuan yang diinginkan yaitu berupa informasi akses sarana air minum dengan menggunakan teknik *classification*. Dasar penggunaan *classification* disesuaikan dengan tipe karakter data yang telah dihasilkan dari data desa-desa paska PAMSIMAS, hal ini bertujuan agar mudah dalam menentukan algoritma yang akan digunakan untuk mengolah *dataset*. Pada *dataset* klasifikasi ini, akan memanfaatkan algoritma Naive Bayes dan *Decision Tree* C4.5. Penggunaan dua algoritma ini, selain baik dalam menghasilkan nilai akurasi, juga bertujuan untuk mendapatkan hasil yang lebih variatif lagi diantaranya pengetahuan yang berbentuk aturan dari hasil proses oleh Naive Bayes dan pohon keputusan dari hasil proses *Decision Tree* C4.5. *Dataset* ini diolah dan dianalisa menggunakan aplikasi Rapidminer untuk mempermudah dan mempercepat dalam menghasilkan pengetahuan.

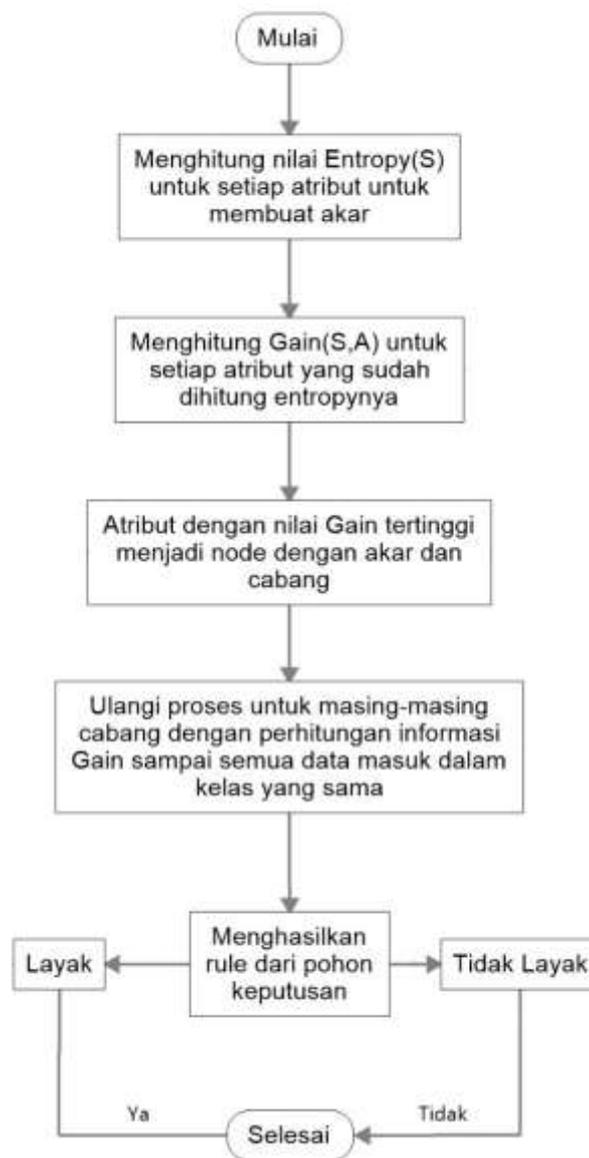
Pada tahap ini pula secara bersamaan akan dilakukan proses *balancing dataset*, sehingga kelas yang minoritas akan di *over sampling* jumlahnya menjadi sama dengan jumlah kelas mayoritas dan dilakukan secara *random*. Teknik *balancing* ini merupakan langkah agar nilai akurasi dapat digunakan sebagai analisa keputusan klasifikasi pada tahap interpretasi dan evaluasi pada *dataset* yang tidak *balance*. *Balancing* ini akan menggunakan operator *sample* dan SMOTE sebelum *performance metrics* dihasilkan dari proses *modeling* dengan menggunakan algoritma.

Pada algoritma Naive Bayes yang merupakan salah satu teknik dalam *data mining* selalu digunakan untuk pemecahan masalah menggunakan prinsip probabilitas dengan membandingkan nilai antar label dan nilai yang paling tinggi yang akan menjadi hasilnya. Cara kerja proses penghitungan *dataset* menggunakan algoritma Naive Bayes dapat dilihat pada Gambar 3.5.



Gambar 3.5
Cara Kerja Algoritma Naive Bayes
(Bulolo, 2020)

Sedangkan untuk mengetahui cara kerja penghitungan dengan menggunakan algoritma *Decision Tree* C4.5 dapat dilihat pada Gambar 3.6. *Decision Tree* C4.5 ini merupakan algoritma yang tidak hanya dapat melakukan prediksi hasil klasifikasi namun juga dapat menghasilkan sebuah informasi pohon keputusan yang dibangun berdasarkan *gain* tertinggi dari setiap datanya. Pohon keputusan ini akan mempermudah dalam proses interpretasi dan evaluasi untuk menentukan jalur dominan yang dapat menentukan akses sarana air minum dapat tercapai.



Gambar 3.6
Pembentukan pohon keputusan dengan algoritma *Decision Tree C4.5*

3.2.6 Interpretasi dan evaluasi

Hasil dari proses *data mining* akan dipresentasikan ke dalam bentuk pohon keputusan dan bentuk aturan atau *rule*. Kemudian dari pengetahuan atau informasi yang sudah dihasil berdasarkan proses *data mining* sebelumnya kemudian akan dilakukan proses evaluasi untuk menjawab hipotesis yang masih bersifat sementara terkait akses sarana air minum pada data desa-desa paska PAMSIMAS.

Evaluasi yang dilakukan akan menguji pengetahuan dan informasi yang sudah didapatkan pada proses *data mining* menggunakan *confusion matrix* yang telah tersedia pada aplikasi Rapidminer dan selanjutnya akan dilihat *performance metrics* dari *confusion matrix* tersebut menggunakan *accuracy*, *precision* dan *recall*. Pengujian ini dilakukan untuk mengetahui tingkat akurasi dari pola yang sudah dihasilkan dari *model* yang sudah digunakan serta presisi dan *recall* yang dapat memberikan informasi ketepatan hasil prediksi klasifikasi.

3.2.7 Penggunaan pengetahuan

Penelitian ini akan menghasilkan suatu pengetahuan yang diperoleh melalui proses *data mining*. Informasi *confusion matrix* beserta *performance metrics* nya dapat digunakan untuk menganalisa tingkat akurasi dari *dataset* berdasarkan algoritma yang digunakan pada penelitian ini yaitu algoritma Naive Bayes dan *Decision Tree C4.5*. Sehingga dapat disimpulkan bahwa metode *data mining* dapat digunakan untuk memprediksi berdasarkan probabilitas yang dimungkinkan dapat digunakan untuk mengetahui akses universal sarana air minum berdasarkan data desa-desa keberlanjutan atau desa-desa paska PAMSIMAS.

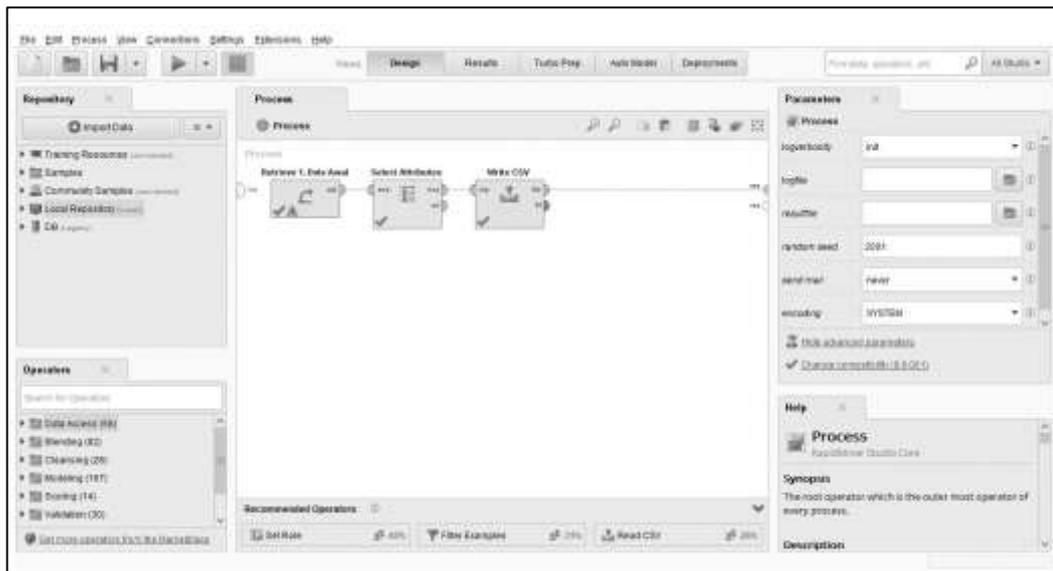
Informasi prediksi akses sarana air minum ini akan menjadi *urgent* untuk menentukan desa-desa paska yang akan mendapatkan dana hibah berdasarkan desa yang dimungkinkan bisa mencapai akses 100% air minum. Dana hibah yang merupakan dana untuk memberikan apresiasi kepada desa dengan tingkat akses air minum yang tinggi yang diharapkan dengan dana hibah ini akan dapat mendorong semua masyarakat dapat mengakses sarana air minum.

BAB IV HASIL DAN PEMBAHASAN

4.1 Seleksi Data

Berdasarkan data yang telah diperoleh tidak semua atribut yang terdapat pada data laporan progres desa keberlanjutan PAMSIMAS tahun 2008-2020 periode Desember 2020 dapat digunakan seluruhnya karena tidak ada kaitan dengan akses sarana air minum. Proses seleksi atribut berdasarkan hasil wawancara dengan tenaga ahli bidang *Water Supply And Sanitation (WSS)* yang bertanggung jawab langsung terhadap keberfungsian sarana air minum PAMSIMAS.

Seleksi data meliputi penentuan *data target* atau label dan pengumpulan sampel data serta menyimpannya dalam sebuah berkas sehingga menjadi *dataset*. Pada Gambar 4.1 dapat dilihat proses seleksi data yang dilakukan menggunakan aplikasi RapidMiner dengan memanfaatkan operator *Select Attributes* dan *Write CSV*.



Gambar 4.1
Proses seleksi data dengan RapidMiner

Tabel 4.1 merupakan hasil proses seleksi data yang berjumlah berjumlah 9 atribut dari 35 atribut yang terdapat pada data laporan desa keberlanjutan PAMSIMAS periode Desember 2020 dengan sampel data seluruhnya sebanyak 29.819 *record*.

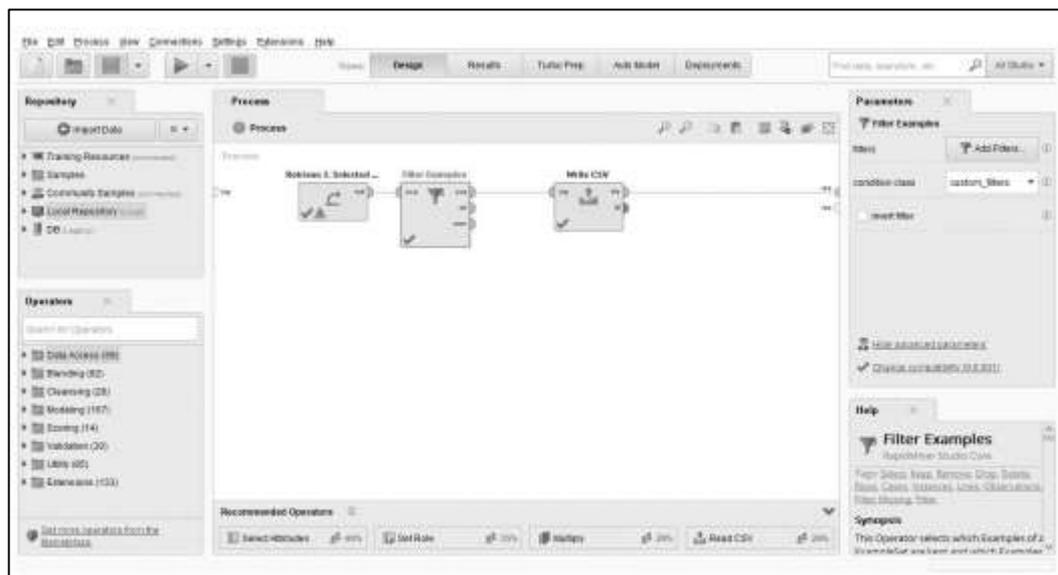
Tabel 4.1 Hasil Tahapan Seleksi

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	...	Akses Universal
1	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
2	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
3	Mandiri	Ada	Pembukuan lengkap	...	Belum
4	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
5	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
6	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
7	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
8	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
9	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
10	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
11	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
12	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
13	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
14	Mandiri	Ada	Tidak ada pembukuan	...	Belum
15	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
16	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
17	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
18	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
...
29801	Mandiri	Ada	Pembukuan lengkap	...	Belum
29802	Mandiri	Ada	Pembukuan lengkap	...	Belum
29803	Mandiri	Ada	Pembukuan lengkap	...	Belum
29804	Mandiri	Ada	Pembukuan lengkap	...	Belum
29805	Mandiri	Ada	Tidak ada pembukuan	...	Belum
29806	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
29807	Mandiri	Ada	Pembukuan lengkap	...	Belum
29809	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
29810	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
29811	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
29812	Mandiri	Ada	Pembukuan lengkap	...	Belum
29813	Mandiri	Ada	Pembukuan lengkap	...	Belum
29814	Mandiri	Ada	Pembukuan tidak lengkap	...	Belum
29815	Mandiri	Ada	Pembukuan lengkap	...	Belum
29816	Mandiri	Ada	Pembukuan lengkap	...	Belum
29817	Mandiri	Ada	Tidak ada pembukuan	...	Belum
29818	Mandiri	Ada	Pembukuan lengkap	...	Belum
29819	Mandiri	Ada	Tidak ada pembukuan	...	Belum

Sumber: SIM PAMSIMAS

4.2 Praproses Data

Pada tahap ini, *dataset* yang sudah diseleksi atribut dan ditentukan label atau kelasnya, maka selanjutnya dilakukan proses *cleaning data* menggunakan aplikasi Rapidminer dengan memanfaatkan operator *Filter Examples* untuk membersihkan *data-data* yang kosong atau *missing data* dan menerapkan parameter *custom filter* dengan memilih atribut-atribut yang terdapat *missing datanya* sebagaimana yang ditampilkan pada Gambar 4.2.



Gambar 4.2
Proses *cleaning* data menggunakan RapidMiner

Data yang *missing* merupakan data dari desa-desa yang tidak melaporkan progres desa pasca periode Desember 2020 secara lengkap. Setelah melewati praproses data maka dihasilkan *dataset* yang sudah bersih dengan sampel data sebanyak 29.378 *record* yang sudah siap untuk dilakukan proses berikutnya.

4.3 Transformasi Data

Pada proses ini *dataset* yang sudah terseleksi dan bersih selanjutnya ditransformasikan ke dalam bentuk format yang sesuai dengan kebutuhan pada proses algoritma Naive Bayes dan algoritma *Decision Tree* C4.5. Proses transformasi data

menggunakan aplikasi Microsoft Excel 2019. Berikut pola transformasi data yang digunakan sebagaimana yang terdapat pada Tabel 4.2.

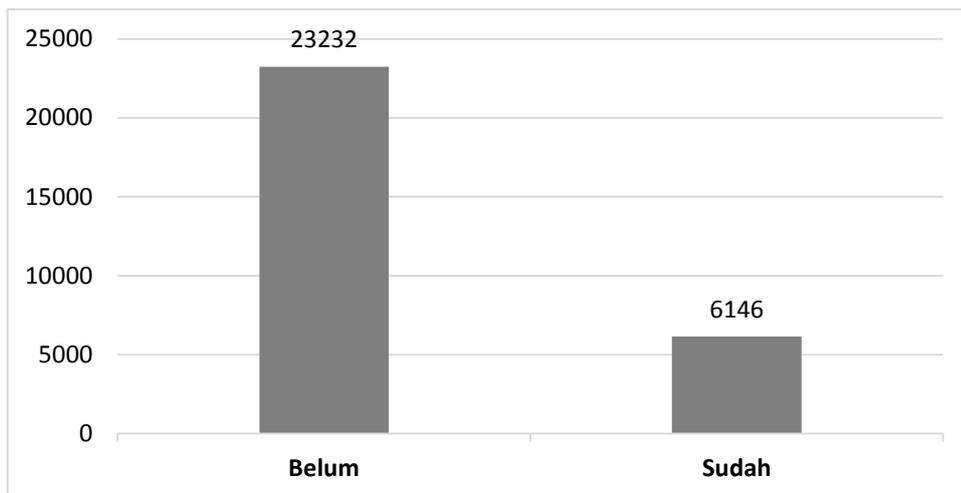
Tabel 4.2 Transformasi Data

Atribut	Nilai	Transformasi
Pengelolaan SPAM	BUMDes	BUMDes
	Mandiri	Mandiri
Rencana Kerja KPSPAMS	Ada	Ada
	Tidak ada	Tidak ada
Administrasi Pembukuan	Pembukuan lengkap	Ada
	Pembukuan tidak lengkap	
	Tidak ada Pembukuan	Tidak ada
Daftar Aset	Ada	Ada
	Tidak ada	Tidak ada
Iuran	Iuran < BOP	Ada
	Iuran >= BOP	
	Iuran >= BOP & CR	
	Tidak ada Iuran	Tidak ada
Kondisi SAM	Berfungsi Baik	Berfungsi Baik
	Berfungsi Sebagian	Berfungsi Sebagian
	Tidak Berfungsi	Tidak Berfungsi
Rencana Kemitraan	Realisasi 1%-59%	Ada
	Realisasi 60%-80%	
	Realisasi > 80%	
	Belum ada realisasi	Tidak Ada
	Tidak ada rencana kemitraan	
Akses Universal	Sudah	Sudah
	Belum	Belum

4.4 Data Mining

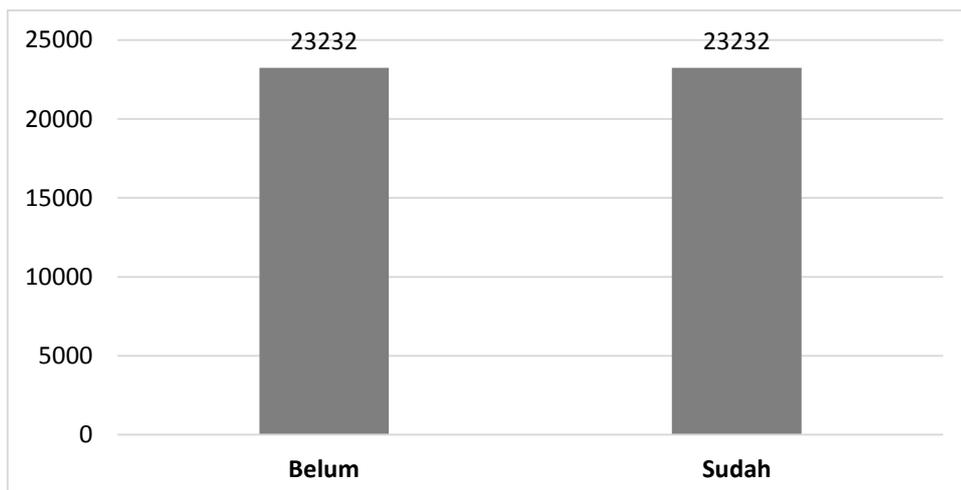
Dataset selanjutnya dilakukan proses *modeling* dengan menggunakan algoritma Naive Bayes dan algoritma *Decision Tree* C4.5. Pada algoritma Naive Bayes digunakan untuk mencari nilai peluang, sedangkan algoritma *Decision Tree* C4.5 untuk membuat pohon keputusan agar mudah dimengerti dan dapat menghasilkan sebuah *rule*. Penggunaan dua algoritma ini untuk melihat tingkat akurasi data melalui *confusion matrix* dan hasil prediksi klasifikasi.

Pada tahap ini pula, dikarenakan *dataset* terjadi *imbalancing* data antara data yang sudah dan belum pada atribut akses sarana air minum yang dijadikan sebagai label atau kelas, maka agar nilai akurasi pada *confusion matrix* dapat dimanfaatkan untuk proses interpretasi dan evaluasi, selanjutnya akan digunakan operator *sample* dan SMOTE untuk melakukan proses *balancing data*.



Gambar 4.3 Grafik *imbalancing dataset*

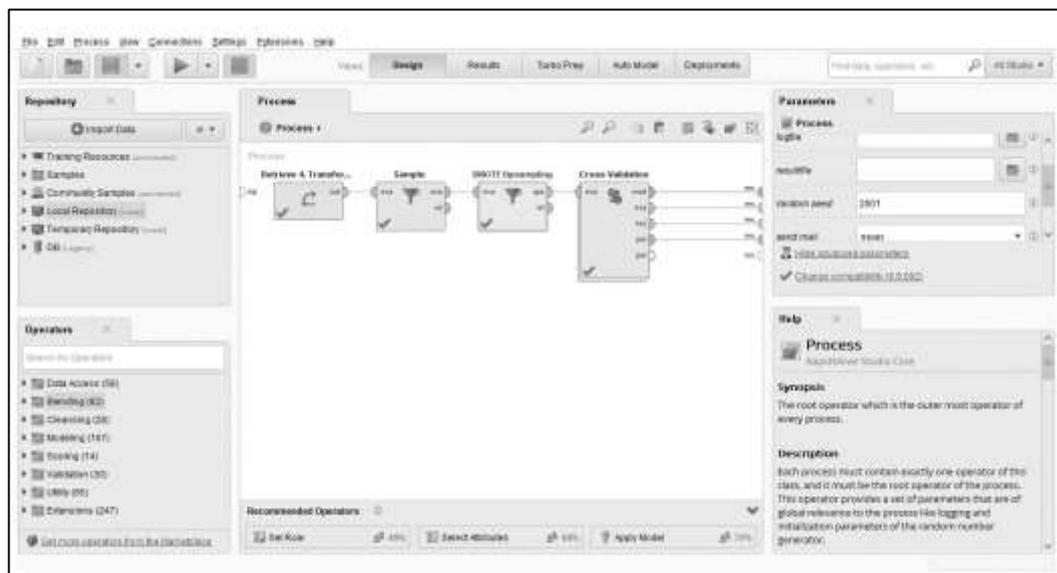
Gambar 4.3 menunjukkan bahwa pada *dataset* terdapat sejumlah *record* yang tidak *balance* antara kelas belum dan kelas sudah. Pada kelas belum terdapat 23.232 record sedangkan pada kelas sudah berjumlah 6.146. Jika *Imbalancing dataset* ini digunakan untuk proses *data mining* maka akan menghasilkan nilai akurasi yang semu dan berbahaya pada kasus-kasus tertentu. Oleh karena itu perlu dilakukan proses *balancing data* dengan menggunakan operator *sample* dan SMOTE sehingga kelas sudah yang merupakan kelas minoritas, datanya akan menjadi *balance* dengan kelas belum. Pada Gambar 4.4 menunjukkan *dataset* yang sudah dilakukan proses *balancing*.



Gambar 4.4 Grafik hasil *balancing dataset*

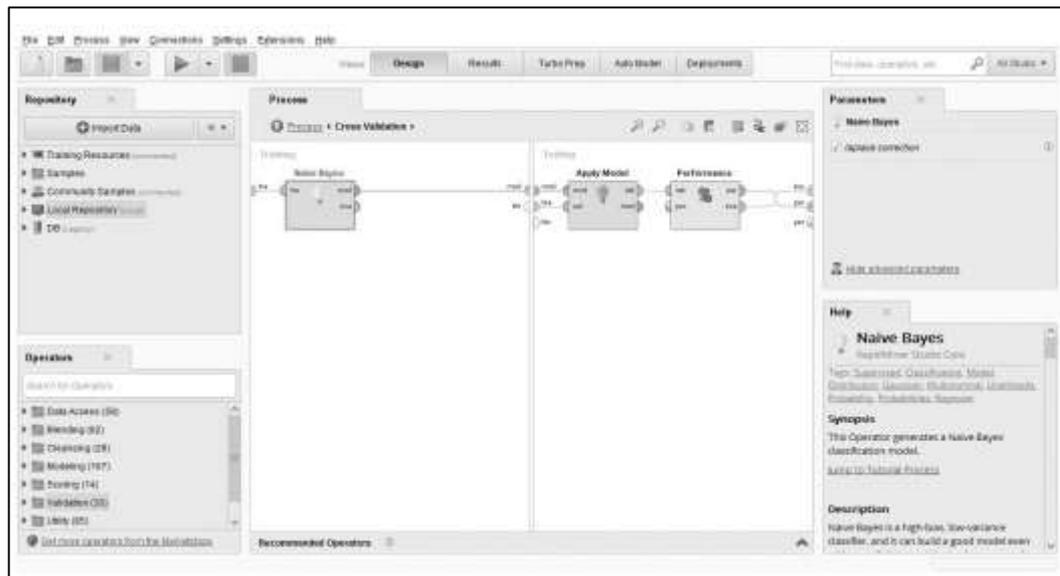
4.4.1 Algoritma Naive Bayes

Proses klasifikasi model dengan algoritma Naive Bayes menggunakan RapidMiner seperti yang terlihat pada Gambar 4.5. Pada proses algoritma Naive Bayes ini terdapat proses *balancing dataset* menggunakan operator *sample* dan SMOTE, yang bertujuan agar data kelas yang tidak *balance* akan menjadi seimbang, lalu dilanjutkan dengan proses validasi silang menggunakan operator *Cross Validation*.



Gambar 4.5
Proses validasi model Naive Bayes dengan *Cross Validation*

Selanjutnya Gambar 4.6 pada operator *cross validation* yang akan melakukan operasi validasi silang dengan sub proses pelatihan dan sub proses pengujian. Sub proses pelatihan digunakan untuk melatih model algoritma Naive Bayes kemudian diterapkan dalam sub proses pengujian. Pada *cross validation* ini akan melakukan validasi analisis pada model yang digunakan dan memilih hasil klasifikasi dengan kinerja prediksi yang baik dengan menggunakan *folds* atau lipatan yang digunakan sebanyak 10 *folds*. *Folds* ini berguna untuk membagi *dataset* menjadi 10 bagian yang akan dianalisa oleh model secara berurutan dan bergantian.



Gambar 4.6
Proses *modeling* menggunakan algoritma Naive Bayes

Data yang telah dilakukan uji pada kotak *training* pada *cross validation* dengan algoritma Naive Bayes merupakan proses untuk mengambil nilai tertinggi dari probabilitas data berdasarkan peluang yang muncul, baik untuk data-data yang bersifat konsisten ataupun tidak konsisten atau bahkan data-data yang bersifat bias. Sehingga pada setiap *record* yang terdapat pada *dataset*, akan terjadi penentuan hasil keputusan klasifikasi berdasarkan probabilitas. Kemudian pada kotak *testing* pada Gambar 4.6 merupakan proses pengujian berdasarkan model yang sudah terbentuk oleh algoritma Naive Bayes, lalu akan mengeluarkan hasil nilai prediksi klasifikasi untuk setiap *record*nya.

Proses *modeling* menggunakan operator Naive Bayes ini telah menghasilkan *ExampleSet* yang merupakan hasil prediksi akses SAM sebagaimana pada Gambar 4.7. Pada *ExampleSet*, *dataset* yang diuji dengan total awal sebanyak 29.378 *record* dimana kelas akses universal dengan klasifikasi belum berjumlah 23.232 *record* dan kelas sudah berjumlah 6.146 *record*. Setelah dilakukan proses *balancing*, kelas minoritas yaitu kelas dengan klasifikasi sudah akan diseimbangkan dengan kelas dengan klasifikasi belum sehingga masing-masing kelas berjumlah 23.232 *record*. Sehingga total *record* yang dilakukan uji pada *modeling* berjumlah 46.464 *record*.

Row No.	Akses Univ.	prediksbk.	confidence	confidence	Pengujian	Record No.	Administrasi	Daftar Ases	Ases
1	Belum	Sudah	0.528	0.491	Mandiri	Ada	Ada	Ada	Ada
2	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada
3	Belum	Sudah	0.528	0.491	Mandiri	Ada	Ada	Ada	Ada
4	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada
5	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada
6	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada
7	Belum	Sudah	0.528	0.491	Mandiri	Ada	Ada	Ada	Ada
8	Belum	Sudah	0.528	0.491	Mandiri	Ada	Ada	Ada	Ada
9	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada
10	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada
11	Belum	Sudah	0.528	0.491	Mandiri	Ada	Ada	Ada	Ada
12	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada
13	Belum	Sudah	0.528	0.491	Mandiri	Ada	Ada	Ada	Ada
14	Belum	Sudah	0.518	0.482	Mandiri	Ada	Ada	Ada	Ada

Gambar 4.7
Hasil *ExampleSet* data menggunakan *modeling* Naive Bayes

Berdasarkan hasil statistik pada *ExampleSet* diperoleh hasil prediksi klasifikasi akses universal sarana air minum dengan klasifikasi sudah sebanyak 34.756 *record*, sedangkan hasil prediksi klasifikasi belum sebanyak 11.708 *record*. Hasil pengujian dengan Naive Bayes ini merupakan penghitungan dengan teknik probabilitas yakni menghitung banyak peluang yang muncul dari suatu data.

4.4.2 Algoritma *Decision Tree* C4.5

Proses pengujian pada algoritma *Decision Tree* C4.5 menggunakan operator *Cross Validation* untuk membagi menjadi dua proses yaitu sub proses *training* dan sub proses *testing* seperti yang terlihat pada Gambar 4.8. Pada *modeling* dengan algoritma *Decision Tree* C4.5 ini untuk operator yang digunakannya sama seperti pada proses *modeling* yang dilakukan dengan algoritma Naive Bayes. Hanya saja pada kotak *testing* di *cross validation* menggunakan algoritma *Decision Tree* C4.5.

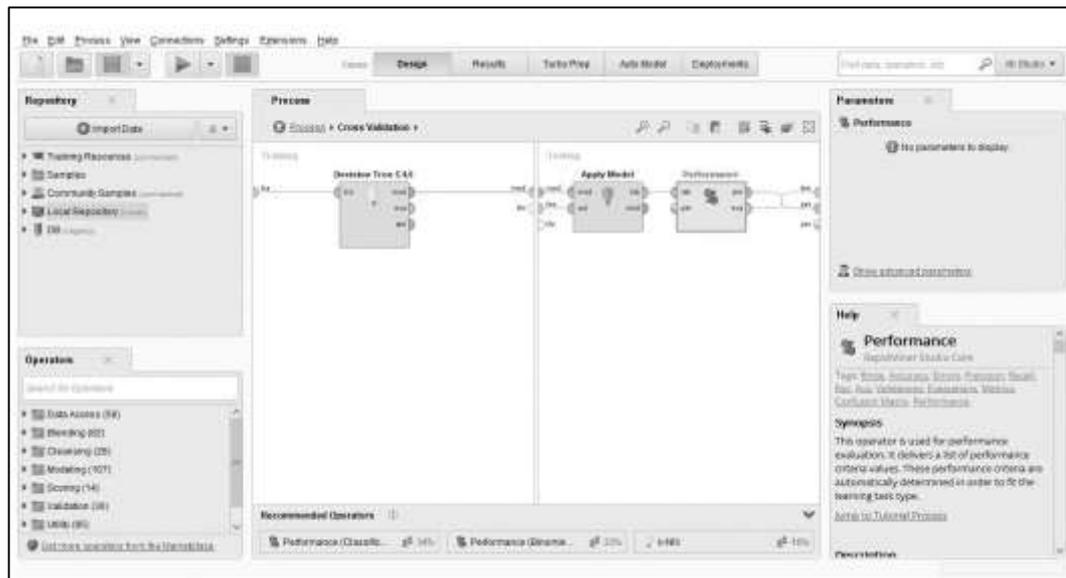
Pada pengujian dengan algoritma *Decision Tree* C4.5 ini, *dataset* yang digunakan sama seperti yang digunakan pada algoritma Naive Bayes, sehingga perlu dilakukan proses *balancing data* sebelum dilakukan *cross validation* untuk proses *modeling* dengan algoritma *Decision Tree* C4.5.



Gambar 4.8
Proses validasi model *Decision Tree* C4.5 dengan *Cross Validation*

Dengan menggunakan operator *Cross Validation* pada Gambar 4.9 dilakukan *setting* pada bagian sub proses *training* menggunakan algoritma *Decision Tree* C4.5 untuk melakukan proses *modeling* terhadap *record* pada atribut-atribut yang terdapat pada *dataset*. Atribut pada *dataset* yang dijadikan label atau kelas adalah atribut akses universal sebagai atribut prediksi klasifikasi dengan jenis data *nominal* dan di *setting* sebagai *binominal* pada saat proses *export file* karena atribut ini hanya memiliki dua jenis *record* yaitu *record* dengan kelas sudah dan *record* dengan kelas belum.

Atribut akses universal yang dijadikan label tidak akan dihitung pada proses *modeling Decision Tree* C4.5 dan hanya atribut-atribut lain saja yang dihitung karena merupakan atribut pendukung untuk melakukan mengujian terhadap prediksi klasifikasi pada atribut label yang sudah ditentukan. Pada bagian sub proses *testing* menggunakan operator *Apply Model* untuk menguji hasil model *Decision Tree* C4.5 yang sudah terbentuk sebagai pembelajaran model yang akan diterapkan pada data *testing* lalu hasil dari proses modeling ini akan dilihat *performancenya*.



Gambar 4.9
Setting sub proses *training* dan *testing* pada *Decision Tree C4.5*

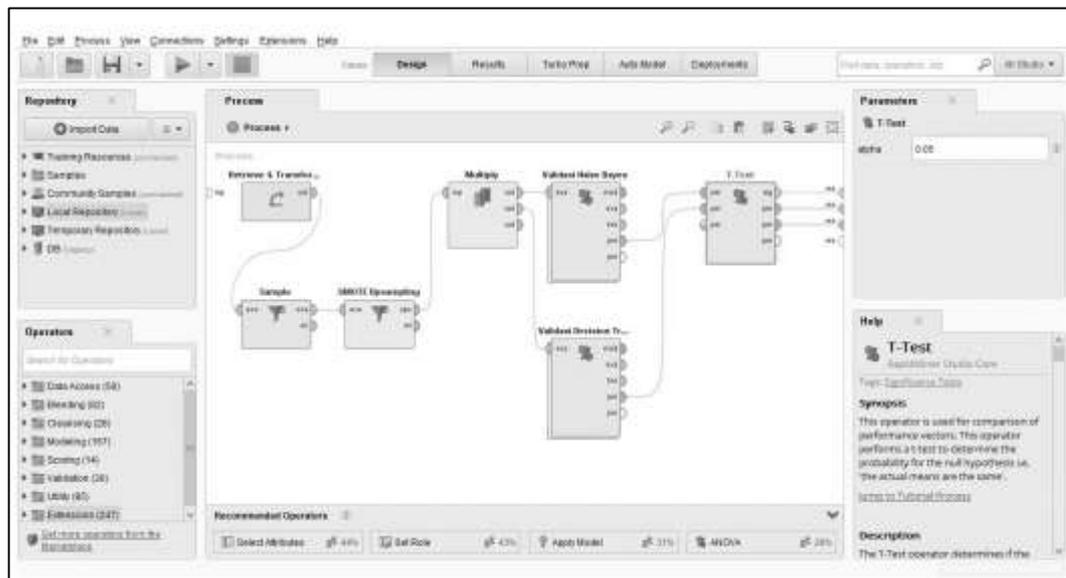
Dari proses *modeling Decision Tree C4.5* di atas dapat menghasilkan *ExampleSet* sebagaimana pada Gambar 4.10. Dari total data awal yang diuji dengan *Decision Tree C4.5* sebanyak 29.378 *record*, lalu dilakukan proses *balancing data* sehingga total menjadi 46.464 *record*. Hasil prediksi klasifikasi sudah berjumlah 24.454 *record* dan klasifikasi belum berjumlah 22.010 *record*.

Row No.	Actual Value	predicted	confidence	confidence	Population	Balance Ratio	Administration	Debt Asset	Age
1	Belum	Belum	0.488	0.504	Mandiri	Ada	Ada	Ada	Ada
2	Belum	Sudah	0.508	0.488	Mandiri	Ada	Ada	Ada	Ada
3	Belum	Belum	0.488	0.504	Mandiri	Ada	Ada	Ada	Ada
4	Belum	Sudah	0.508	0.488	Mandiri	Ada	Ada	Ada	Ada
5	Belum	Sudah	0.508	0.488	Mandiri	Ada	Ada	Ada	Ada
6	Belum	Sudah	0.508	0.488	Mandiri	Ada	Ada	Ada	Ada
7	Belum	Belum	0.488	0.504	Mandiri	Ada	Ada	Ada	Ada
8	Belum	Belum	0.488	0.504	Mandiri	Ada	Ada	Ada	Ada
9	Belum	Belum	0.488	0.504	Mandiri	Ada	Ada	Ada	Ada
10	Belum	Sudah	0.508	0.488	Mandiri	Ada	Ada	Ada	Ada
11	Belum	Belum	0.488	0.504	Mandiri	Ada	Ada	Ada	Ada
12	Belum	Sudah	0.508	0.488	Mandiri	Ada	Ada	Ada	Ada
13	Belum	Belum	0.488	0.504	Mandiri	Ada	Ada	Ada	Ada
14	Belum	Sudah	0.508	0.488	Mandiri	Ada	Ada	Ada	Ada

Gambar 4.10
Hasil *ExampleSet* menggunakan *modeling Decision Tree C4.5*

keputusan hanya dapat digunakan pada data yang seimbang saja. Oleh karena itu, disebabkan *dataset* PAMSIMAS tidak *balance* pada atribut akses universal antara kelas sudah yang minoritas dan kelas belum yang mayoritas, maka perlu dilakukan proses *balancing data* menggunakan operator *sample* dan SMOTE.

Untuk melakukan proses uji *performance* secara bersamaan dapat memanfaatkan T-test sebagai operator pada RapidMiner yang digunakan untuk uji beda atau uji perbandingan seperti terlihat pada Gambar 4.10. Operator statistik t-test ini akan menguji signifikansi *performance* dari masing-masing algoritma, sehingga akan diketahui apakah ada perbedaan yang signifikan antara algoritma Naive Bayes dan Decision Tree C4.5 dengan nilai alpha 0.05.



Gambar 4.12
Performance test pada algoritma

Dari proses di atas, setelah dilakukan *balancing data*, selanjutnya menggunakan operator *Multiply* untuk mengambil objek RapidMiner dari *port input* dan mengirimkan salinannya ke *port output*. Setiap *port* yang terhubung akan membuat salinan *dataset* secara independen dan menjadi satu salinan yang tidak berpengaruh pada salinan lainnya. Pada masing-masing algoritma ditambahkan operator *performance* untuk menghasilkan nilai akurasi, *recall* dan presisi. Selanjutnya pada proses ini pun menggunakan operator T-test untuk melakukan uji algoritma secara bersamaan.

Tabel 4.3 menunjukkan hasil *performance* pada proses *modeling* menggunakan algoritma algoritma Naive Bayes dan pada Tabel 4.4 merupakan hasil *performance* pada proses *modeling* menggunakan algoritma *Decision Tree* C4.5. Untuk melihat hasil *performance* dari masing-masing algoritma dapat memanfaatkan tabel *confusion matrix*.

Tabel 4.3 *Confusion matrix* algoritma Naive Bayes

	<i>True</i> Belum	<i>True</i> Sudah	<i>Class Precision</i>
<i>Pred.</i> Belum	6.640	4.595	59.10%
<i>Pred.</i> Sudah	16.592	18.637	52.90%
<i>Class Recall</i>	28.58%	80.22%	

Tabel 4.4 *Confusion Matrix* algoritma *Decision Tree* C4.5

	<i>True</i> Belum	<i>True</i> Sudah	<i>Class Precision</i>
<i>Pred.</i> Belum	12.535	9.475	56.95%
<i>Pred.</i> Sudah	10.697	13.757	56.26%
<i>Class Recall</i>	53.96%	59.22%	

Dari hasil *confusion matrix* di atas, pada Tabel 4.5 dapat dilihat *performance* dari masing-masing proses *test* yang telah dilakukan dan menghasilkan nilai akurasi, nilai presisi dan nilai *recall* dengan tujuan untuk mengetahui perhitungan dari masing-masing algoritma.

Tabel 4.5 Hasil *performance* algoritma Naive Bayes dan *Decision Tree* C4.5

	Naive Bayes	<i>Decision Tree</i> C4.5
<i>Accuracy</i>	54.40%	56.59%
<i>Precision</i>	52.91%	56.26%
<i>Recall</i>	80.22%	59.22%

Pada Tabel 4.6 merupakan hasil dari uji t-test yang telah dilakukan dan menunjukkan hasil 0.000 atau kurang dari nilai signifikansi yang telah ditentukan sebesar 0.05. Nilai yang lebih kecil dari alpha 0.05 menunjukkan perbedaan yang mungkin signifikan antara nilai rata-rata, sehingga dapat diinterpretasikan bahwa terdapat perbedaan signifikan antara algoritma Naive Bayes dan *Decision Tree* C4.5 pada masing-masing *performancenya*.

Tabel 4.6 Hasil perbandingan uji T-test

A	B	C
	Naive Bayes	<i>Decision Tree</i> C4.5
Naive Bayes		0.000
<i>Decision Tree</i> C4.5		

4.6 Pembahasan Hasil Eksperimen

Algoritma Naive Bayes dan *Decision Tree* C4.5 memiliki teknik masing-masing dalam melakukan penghitungan pada *dataset* desa-desa keberlanjutan atau desa paska PAMSIMAS. Dari hasil penelitian diatas, kedua algoritma tersebut masing-masing memiliki cara penghitungan yang berbeda dalam menghasilkan sebuah informasi. Namun keduanya dapat melakukan proses prediksi klasifikasi dengan baik. Bedanya, pada *Decision Tree* C4.5 dapat menghasilkan pohon keputusan sedangkan Naive Bayes hanya dapat menghitung nilai probabilitas.

Pada algoritma Naive Bayes, karena prinsip yang digunakan untuk menghasilkan informasi menggunakan nilai probabilitas, maka keputusan prediksi pada klasifikasi akan diperoleh berdasarkan pengalaman yang ada pada data sebelumnya atau pada proses *training*. Pada data *training*, Naive Bayes bekerja dengan menghitung kejadian yang terjadi untuk menentukan keputusan atas masalah yang dihadapi pada data *testing*.

Sedangkan algoritma *Decision Tree* C4.5 bekerja dengan cara mengelompokkan beberapa data *training* yang akan menghasilkan suatu pohon keputusan yang berdasarkan pada fakta-fakta yang ada pada *data training* tersebut. Berdasarkan proses tahapan *data mining* pada algoritma *Decision Tree* C4.5, pengetahuan yang telah dihasilkan berupa *rule* atau aturan yang berbentuk pola atribut guna melakukan identifikasi pada klasifikasi akses universal sarana air minum. Pola yang dilahirkan tersebut merupakan pola atribut dengan akar dan cabang pada pohon keputusan berdasarkan *Gain* tertinggi. *Gain* tertinggi yang telah dihasilkan dan menjadi akar pohon keputusan adalah atribut kondisi SAM yang selanjutnya diikuti oleh *Gain-Gain* tinggi lain yang menjadi cabang pada pohon keputusan tersebut.

Urutan atribut yang dominan dan telah menjadi pengetahuan berupa pola dari akar dan cabang pada pohon keputusan berdasarkan nilai gain tertinggi dapat dideskripsikan pada Tabel 4.7. Pola pada pohon keputusan ini akan bermanfaat pada proses penentuan desa paska PAMSIMAS yang akan mendapatkan bantuan dana berupa dana hibah dengan tujuan agar masyarakat dapat melakukan akses sarana air minum 100%. Dengan memanfaatkan akar dan cabang dari atribut-atribut yang dominan pada pohon keputusan,

merupakan langkah yang tepat dalam memberikan rekomendasi cepat untuk digunakan sebagai aspek prasyarat pada penentuan desa yang akan mendapatkan dana hibah.

Tabel 4.7 Pengetahuan yang dihasilkan menggunakan algoritma *Decision Tree C4.5*

Rule
Jika kondisi SAM berfungsi baik dan iuran ada dan pengelolaan SPAM mandiri dan rencana kerja KPSPAM ada dan daftar aset ada dan kualitas air memenuhi syarat dan administrasi pembukuan ada dan rencana kemitraan ada maka akses universal sudah.

Dari segi hasil *performance metrics* yang mengukur kinerja model menggunakan *confusion matrix* telah memberikan informasi perbandingan hasil klasifikasi dengan menggunakan algoritma Naive Bayes dan *Decision Tree C4.5*. Hasil *performance metrics* terhadap pengukuran tingkat akurasi berdasarkan perhitungan yang telah dilakukan pada 29.378 data *sample* yang sama pada masing-masing model; didapatkan nilai hasil akhir akurasi.

Dengan menggunakan metode *Decision Tree C4.5* mendapatkan hasil dengan tingkat akurasi 56.59% dan perhitungan menggunakan metode Naive Bayes mendapatkan tingkat akurasi sebesar 54.40%. Dari hasil perhitungan akurasi ini dapat disimpulkan bahwa model *Decision Tree C4.5* sedikit lebih tinggi nilai akurasinya dalam mengklasifikasikan data dengan benar dibandingkan tingkat akurasi yang dihasilkan dengan metode Naive Bayes.

Tingkat akurasi yang telah didapatkan dengan model Naive Bayes dan *Decision Tree C4.5* tersebut merupakan fakta untuk kasus *dataset* yang nyata untuk mengetahui berapa persen desa yang benar diprediksi akses universalnya, baik dengan kelas sudah maupun dengan kelas belum dari total keseluruhan desa. Kemudian nilai akurasi yang diperoleh dengan model Naive Bayes dan *Decision Tree C4.5* menunjukkan kisaran yang terpaut tidak terlalu jauh, hal ini menunjukkan bahwa tahapan *preprocessing* dan *cleaning data* sudah dilakukan dengan benar.

Sedangkan untuk pengukuran nilai presisi berdasarkan hasil *performance metrics* yang bertujuan untuk mengukur tingkat keakuratan antara data prediksi benar positif yang diminta dengan hasil prediksi yang diberikan oleh model menunjukkan dengan algoritma *Decision Tree C4.5* mendapatkan nilai presisi sebesar 56.26%, sedangkan dengan

algoritma Naive Bayes diperoleh 52.91%. Kemudian hasil pengukuran lain pada *performance metrics* yaitu *recall* yang dapat menggambarkan keberhasilan model untuk mengetahui berapa persen desa yang diprediksi sudah akses universal dibandingkan dengan keseluruhan desa yang sebenarnya sudah akses universal. Dengan menggunakan algoritma *Decision Tree* C4.5 diperoleh nilai *recall* sebesar 59.22%, sedangkan dengan Naive Bayes menghasilkan *recall* sebesar 80.22%.

Berdasarkan hasil *performance* pada pengujian *dataset* menggunakan dua algoritma *Decision Tree* C4.5 dan Naive Bayes, dapat direkomendasikan untuk menggunakan *Decision Tree* C4.5 untuk proses klasifikasi; selain nilai akurasi yang lebih tinggi, dengan algoritma ini dapat menghasilkan pohon keputusan yang dapat memudahkan untuk melakukan identifikasi atribut yang dapat berpengaruh cukup dominan pada pencapaian akses universal sarana air minum.

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan proses yang sudah dilakukan, hasil perhitungan pada analisa dan pengujian data desa keberlanjutan PAMSIMAS untuk menentukan algoritma yang akan dimanfaatkan untuk proses analisa penentuan desa-desa yang akan mendapatkan dana hibah PAMSIMAS guna mencapai 100% akses universal pada sarana air minum dapat disimpulkan sebagaimana berikut:

1. Pengujian yang sudah dilakukan secara bertahap pada *dataset* desa keberlanjutan PAMSIMAS setelah dilakukan proses *balancing data*; dengan menggunakan algoritma *Decision Tree C4.5* telah menghasilkan klasifikasi dengan *correct prediction* sebanyak 26.292 *record* dan *wrong prediction* sebanyak 20.172 *record*, sedangkan pada algoritma Naive Bayes untuk *correct prediction* diperoleh sebanyak 25.378 *record* dan *wrong prediction* 21.086 *record*. Dari hasil ini jelas bahwa algoritma *Decision Tree C4.5* paling tinggi dalam menghasilkan prediksi klasifikasi.
2. Dengan melakukan proses perhitungan menggunakan algoritma *Decision Tree C4.5* telah diperoleh nilai akurasi sebesar 56.59% lebih tinggi dari hasil yang sudah dihasilkan dengan menggunakan algoritma Naive Bayes yaitu sebesar 54.40%. Dengan demikian, pada penelitian ini algoritma *Decision Tree C4.5* memiliki tingkat akurasi yang lebih tinggi pada klasifikasi akses sarana air minum. Namun nilai akurasi yang diperoleh dari kedua algoritma ini masing-masing menunjukkan nilai yang belum cukup untuk proses klasifikasi.
3. Berdasarkan pengujian dengan algoritma *Decision Tree C4.5* telah diperoleh pengetahuan berupa pola atribut yang dapat mempengaruhi terhadap klasifikasi universal akses sarana air minum yang terbentuk dari akar dan cabang pada pohon keputusan melalui penghitungan *Gain* tertinggi. Atribut-atribut yang memiliki *Gain* tertinggi dan menunjukkan yang berpengaruh adalah sebagai berikut:
 - a. Kondisi SAM : Berfungsi baik

- b. Iuran : Ada
- c. Pengelolaan SPAM : Mandiri
- d. Rencana kerja KPSPAM : Ada
- e. Daftar Aset : Ada
- f. Kualitas Air : Memenuhi syarat
- g. Administrasi pembukuan : Ada
- h. Rencana kemitraan : Ada

5.2 Saran

Berdasarkan hasil kesimpulan, maka penulis memberikan beberapa saran yang bermanfaat sebagaimana berikut:

1. Diharapkan untuk penelitian selanjutnya dapat menggunakan pendekatan *deep learning* dan menggunakan metode algoritma klasifikasi lainnya untuk memperoleh algoritma dengan tingkat akurasi yang lebih tinggi dalam membuat klasifikasi akses sarana air minum PAMSIMAS.
2. *Dataset* pada penelitian selanjutnya dapat ditingkatkan dengan menambahkan data desa-desa PAMSIMAS pada periode berikutnya dan menerapkan teknik penyeleksian atribut yang paling berpengaruh.

DAFTAR PUSTAKA

- Bhardwaj, B.K. & Pal, S. (2011) 'Data Mining: A Prediction For Performance Improvement Using Classification', *IJCSIS*, 9.
- Bramer, M. (2016) *Principles of Data Mining*. Third Edit, Edited by Ian Mackie, London: Springer.
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Yogyakarta, Yogyakarta, Indonesia: Deepublish.
- Dangare, C.S. & Apte, S.S. (2012) 'Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques', *International Journal of Computer Applications*, 47, 44–48.
- Dharminder Kumar & Deepak Bhardwaj (2011) 'Rise of Data Mining: Current and Future Application Areas', *International Journal of Computer Science Issues*, 8, 256–260.
- DJCK, K. (2021) *Pedoman Umum Program PAMSIMAS*, Jakarta.
- Fernandez, A., Garcia, S. & Herrera, F. (2018) 'SMOTE for Learning from Imbalanced Data: Progress and Challenges , Marking the 15-year Anniversary', *Journal of Artificial Intelligence Research*, 61, 863–905.
- Gorunescu, F. (2011) *Data Mining Concepts, Models and Techniques*. 12th ed., Edited by J. Kacprzyk and L. C Jain, Berlin: Springer.
- Han, J., Kamber, M. & Pei, J. (2012) *Data Mining Concepts and Techniques*. Third Edit, Waltham USA: Morgan Kaufmann.
- Hofmann, M. & Klinkenberg, R. (2014) *Rapidminer Data Mining Use Cases and Business Analytics Applications*, Edited by V. Kumar, Suite: CRC Press.
- Holzinger, A., Dehmer, M. & Jurisica, I. (2014) 'Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art , future challenges and research directions', *BMC Bioinformatics*, 15, 1–9.
- Hossin, M. & Sulaiman, M. (2015) 'A Review on Evaluation Metrics for Data Classification Evaluations', *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 5.
- Karim, M. & Rahman, R.M. (2013) 'Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing', *Journal of Software Engineering and Applications*, 06, 196–206.
- Li, A., Liu, K. & Ge, Z. (2015) 'Application of Data Mining in the Colleges ' in-Class Teaching Quality Evaluation System', *Journal of Computer*, 10, 166–175.
- Mantas, C.J. & Abellán, J. (2014) 'Expert Systems with Applications Credal-C4 . 5: Decision tree based on imprecise probabilities to classify noisy data', *Expert Systems with Applications*, Elsevier Ltd, 41, 4625–4637.
- Marlina, L., Iim, M. & Utama Siahaan, A.P. (2016) 'Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)', *International Journal of Engineering Trends and Technology*, 38, 380–383.
- Matsunaga, F.T., Brancher, J.D. & Busto, R.M. (2014) 'Data mining applications and techniques: a systematic review', *ResearchGate*, 1.
- Moslehi, P., Bidgoli, B.M., Nasiri, M. & Salajegheh, A. (2011) 'Multi-objective Numeric Association Rules Mining via Ant Colony Optimization for Continuous Domains without Specifying Minimum Support and Minimum Confidence', *International Journal of Computer Science*, 8, 34–41.

- NMC, T.P. (2017) *Prosedur Operasional Baku (POB) Mekanisme Pelaporan Data Sistem Informasi Manajemen PAMSIMAS*, Jakarta.
- Oktanisa, I. & Supianto, A.A. (2018) 'Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing', *JTIK*, 5, 567–576.
- Patel, H., Rajput, D.S., Reddy, G.T., Iwendi, C., Bashir, A.K. & Jo, O. (2020) 'A review on classification of imbalanced data for wireless sensor networks', *SAGE Journals*, 16.
- Profil Pamsimas. (2021, January 9). Diambil kembali dari Situs resmi program PAMSIMAS: <http://www.pamsimas.org>
- Rahayu, E.S., Wahono, R.S. & Supriyanto, C. (2015) 'Penerapan Metode Average Gain , Threshold Pruning dan Cost Complexity Pruning untuk Split Atribut pada Algoritma C4 . 5', *Journal of Intelligent Systems*, 1, 91–97.
- Rahayuningsih, P.A. (2019) 'Komparasi Algoritma Klasifikasi Data Mining untuk Memprediksi Tingkat Kematian Dini Kanker dengan Dataset Early Death Cancer', *JOINTECS (Journal of Information Technology and Computer Science)*, 4, 63.
- Raval, K.M. (2012) 'Data Mining Techniques', *International Journal of Advanced Research in Computer Science and Software Engineering*, 2, 439–442.
- Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media.
- Wanto, A., Siregar, M.N.H., Windarto, A.P., Hartama, D., Ginantra, N.L.W.S.R., Napitupulu, D., Negara, E.S., Lubis, M.R., Dewi, S.V. & Prianto, C. (2020) *Data Mining : Algoritma dan Implementasi*. Cetakan 1, Edited by T. Limbong, Medan: Yayasan Kita Menulis.
- Witten, I.H., Frank, E., Hall, M.A. & Pal, C.J. (2017) *Data Mining Practical Machine Learning Tools and Techniques*. Fourth, *Data Mining*. Fourth, Cambridge: Morgan Kaufmann.
- Ye, N. (2014) *Data mining: Theories, Algorithms, and Examples*, *IEEE Potentials*, New York: CRC Press.

LAMPIRAN-LAMPIRAN

Lampiran 1 *Cleaned data*

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	Kondisi SAM	Kualitas Air	Rencana Kemitraan	Akses Universal
1	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Belum ada realisasi	Belum
2	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
3	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran >= BOP & CR	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
4	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Belum ada realisasi	Belum
5	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
6	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
7	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
8	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
9	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi > 80%	Belum
10	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
11	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Belum ada realisasi	Belum
12	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
13	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Sebagian	Memenuhi syarat	Realisasi 1%-59%	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	Kondisi SAM	Kualitas Air	Rencana Kemitraan	Akses Universal
14	Mandiri	Ada	Tidak ada pembukuan	Ada	Tidak ada iuran	Tidak Berfungsi	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
15	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
16	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
17	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Belum ada realisasi	Belum
18	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Belum ada realisasi	Belum
19	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran >= BOP & CR	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
20	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Belum ada realisasi	Belum
21	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
22	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
23	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
24	BUMDes	Ada	Pembukuan lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 60%-80%	Belum
25	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Memenuhi syarat	Realisasi 1%-59%	Belum
...
29351	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	tidak ada rencana kemitraan	Belum
29352	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Tidak ada iuran	Berfungsi Sebagian	Belum melakukan uji kualitas air	Belum ada realisasi	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	Kondisi SAM	Kualitas Air	Rencana Kemitraan	Akses Universal
29353	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Tidak ada iuran	Berfungsi Sebagian	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29354	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran < BOP	Berfungsi Sebagian	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29355	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Tidak ada iuran	Berfungsi Sebagian	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29356	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Tidak ada iuran	Berfungsi Sebagian	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29357	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29358	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Tidak ada iuran	Berfungsi Sebagian	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29359	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29360	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29361	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Sudah
29362	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran < BOP	Berfungsi Baik	Memenuhi syarat	tidak ada rencana kemitraan	Belum
29363	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29364	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran >= BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29365	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Sebagian	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29366	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	Kondisi SAM	Kualitas Air	Rencana Kemitraan	Akses Universal
29367	Mandiri	Ada	Tidak ada pembukuan	Tidak	Tidak ada Iuran	Tidak Berfungsi	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29368	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran < BOP	Berfungsi Sebagian	Belum melakukan uji kualitas air	tidak ada rencana kemitraan	Belum
29369	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29370	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29371	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29372	Mandiri	Ada	Pembukuan tidak lengkap	Tidak	Iuran >= BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29373	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29374	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29375	Mandiri	Ada	Pembukuan tidak lengkap	Ada	Iuran < BOP	Berfungsi Baik	Memenuhi syarat	tidak ada rencana kemitraan	Belum
29376	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Belum ada realisasi	Belum
29377	Mandiri	Ada	Pembukuan lengkap	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Realisasi 1%-59%	Belum
29378	Mandiri	Ada	Tidak ada pembukuan	Ada	Iuran < BOP	Berfungsi Baik	Belum melakukan uji kualitas air	Belum ada realisasi	Belum

Lampiran 2 *Transformed Data*

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	Kondisi SAM	Kualitas Air	Rencana Kemitraan	Akses Universal
1	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
2	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
3	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
4	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
5	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
6	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
7	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
8	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
9	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
10	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
11	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
12	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
13	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Sebagian	Memenuhi syarat	Ada	Belum
14	Mandiri	Ada	Tidak Ada	Ada	Tidak Ada	Tidak Berfungsi	Belum melakukan uji kualitas air	Ada	Belum
15	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
16	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
17	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
18	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
19	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
20	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
21	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
22	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
23	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
24	BUMDes	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	Kondisi SAM	Kualitas Air	Rencana Kemitraan	Akses Universal
25	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Ada	Belum
...
29351	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Tidak Ada	Belum
29352	Mandiri	Ada	Ada	Ada	Tidak Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29353	Mandiri	Ada	Ada	Ada	Tidak Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29354	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29355	Mandiri	Ada	Ada	Ada	Tidak Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29356	Mandiri	Ada	Ada	Ada	Tidak Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29357	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Tidak Ada	Belum
29358	Mandiri	Ada	Ada	Ada	Tidak Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29359	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29360	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29361	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Sudah
29362	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
29363	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29364	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	Kondisi SAM	Kualitas Air	Rencana Kemitraan	Akses Universal
29365	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29366	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29367	Mandiri	Ada	Tidak Ada	Tidak	Tidak Ada	Tidak Berfungsi	Belum melakukan uji kualitas air	Tidak Ada	Belum
29368	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Sebagian	Belum melakukan uji kualitas air	Tidak Ada	Belum
29369	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29370	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Tidak Ada	Belum
29371	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Tidak Ada	Belum
29372	Mandiri	Ada	Ada	Tidak	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Tidak Ada	Belum
29373	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29374	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29375	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Memenuhi syarat	Tidak Ada	Belum
29376	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Tidak Ada	Belum
29377	Mandiri	Ada	Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Ada	Belum
29378	Mandiri	Ada	Tidak Ada	Ada	Ada	Berfungsi Baik	Belum melakukan uji kualitas air	Tidak Ada	Belum

Lampiran 3 ExampleSet Decision Tree C4.5

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	Confidence (Belum)	Confidence (Sudah)	Prediction (Akses Universal)
1	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
2	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
3	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
4	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
5	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
6	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
7	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
8	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
9	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
10	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
11	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
12	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
13	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
14	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
15	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
16	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
17	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
18	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
19	Mandiri	Tidak	Ada	Ada	Ada	...	Belum	0.37593985	0.62406015	Sudah
20	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
21	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
22	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
23	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
24	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
25	Mandiri	Tidak	Ada	Ada	Ada	...	Belum	0.37593985	0.62406015	Sudah
26	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	Confidence (Belum)	Confidence (Sudah)	Prediction (Akses Universal)
27	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
28	Mandiri	Ada	Tidak Ada	Ada	Ada	...	Belum	0.451669596	0.548330404	Sudah
29	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
30	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
31	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
32	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
33	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
34	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.820512821	0.179487179	Belum
35	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
36	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
37	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.582312925	0.417687075	Belum
38	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.582312925	0.417687075	Belum
39	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
40	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
41	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
42	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
43	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
44	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
45	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
46	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
47	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.582312925	0.417687075	Belum
48	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
49	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
50	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.582312925	0.417687075	Belum
51	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.582312925	0.417687075	Belum
52	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
53	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	Confidence (Belum)	Confidence (Sudah)	Prediction (Akses Universal)
54	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
55	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
56	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
57	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.449750535	0.550249465	Sudah
58	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
59	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
60	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.582312925	0.417687075	Belum
61	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
62	Mandiri	Ada	Tidak Ada	Ada	Tidak Ada	...	Sudah	0.657071339	0.342928661	Belum
63	Mandiri	Ada	Tidak Ada	Ada	Tidak Ada	...	Sudah	0.17414248	0.82585752	Sudah
64	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.492790772	0.507209228	Sudah
65	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
66	BUMDes	Ada	Ada	Ada	Ada	...	Sudah	0.329524955	0.670475045	Sudah
67	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
68	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
69	Mandiri	Ada	Ada	Ada	Tidak Ada	...	Belum	0.838709677	0.161290323	Belum
70	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
71	Mandiri	Ada	Ada	Ada	Tidak Ada	...	Belum	0.138743455	0.861256545	Sudah
72	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
73	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
74	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.504473348	0.495526652	Belum
75	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.504473348	0.495526652	Belum
76	Mandiri	Tidak	Tidak Ada	Ada	Ada	...	Sudah	0.37593985	0.62406015	Sudah
77	Mandiri	Tidak	Ada	Ada	Ada	...	Sudah	0.37593985	0.62406015	Sudah
78	Mandiri	Ada	Tidak Ada	Ada	Ada	...	Belum	0.451669596	0.548330404	Sudah
79	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.504473348	0.495526652	Belum
80	BUMDes	Ada	Ada	Ada	Ada	...	Sudah	0.390998594	0.609001406	Sudah

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	Confidence (Belum)	Confidence (Sudah)	Prediction (Akses Universal)
81	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.504473348	0.495526652	Belum
82	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.666666667	0.333333333	Belum
83	BUMDes	Ada	Ada	Ada	Ada	...	Sudah	0.390998594	0.609001406	Sudah
84	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.666666667	0.333333333	Belum
85	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
86	BUMDes	Ada	Ada	Ada	Ada	...	Belum	0.390998594	0.609001406	Sudah
87	Mandiri	Ada	Tidak Ada	Ada	Ada	...	Belum	0.451669596	0.548330404	Sudah
88	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
89	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.492790772	0.507209228	Sudah
90	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.337770383	0.662229617	Sudah
91	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.492790772	0.507209228	Sudah
92	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.492790772	0.507209228	Sudah
93	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.666666667	0.333333333	Belum
94	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.504473348	0.495526652	Belum
95	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.666666667	0.333333333	Belum
96	Mandiri	Ada	Ada	Tidak	Tidak Ada	...	Belum	0.723076923	0.276923077	Belum
97	BUMDes	Ada	Ada	Ada	Ada	...	Belum	0.329524955	0.670475045	Sudah
98	BUMDes	Ada	Ada	Ada	Ada	...	Belum	0.390998594	0.609001406	Sudah
99	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.492790772	0.507209228	Sudah
100	Mandiri	Ada	Tidak Ada	Ada	Ada	...	Sudah	0.451669596	0.548330404	Sudah

Lampiran 4 ExampleSet Naive Bayes

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	confidence (Belum)	confidence (Sudah)	prediction (Akses Universal)
1	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
2	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
3	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
4	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
5	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
6	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
7	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
8	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
9	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
10	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
11	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
12	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
13	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
14	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
15	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
16	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
17	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
18	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
19	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
20	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.779175503	0.220824497	Belum
21	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
22	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
23	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.785670126	0.214329874	Belum
24	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
25	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
26	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	confidence (Belum)	confidence (Sudah)	prediction (Akses Universal)
27	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
28	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
29	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
30	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
31	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
32	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
33	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
34	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
35	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
36	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
37	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
38	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
39	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
40	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
41	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
42	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
43	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
44	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
45	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
46	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
47	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
48	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.511391272	0.488608728	Belum
49	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
50	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.511391272	0.488608728	Belum
51	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
52	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.511391272	0.488608728	Belum
53	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	confidence (Belum)	confidence (Sudah)	prediction (Akses Universal)
54	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.511391272	0.488608728	Belum
55	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
56	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
57	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
58	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
59	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
60	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.551104462	0.448895538	Belum
61	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
62	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
63	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
64	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
65	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.481557564	0.518442436	Sudah
66	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
67	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.511391272	0.488608728	Belum
68	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.501855068	0.498144932	Belum
69	Mandiri	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.58527764	0.41472236	Belum
70	Mandiri	Ada	Tidak Ada	Ada	Tidak Ada	...	Sudah	0.462641893	0.537358107	Sudah
71	Mandiri	Tidak	Tidak Ada	Ada	Tidak Ada	...	Belum	0.371003917	0.628996083	Sudah
72	Mandiri	Ada	Ada	Ada	Tidak Ada	...	Belum	0.748984011	0.251015989	Belum
73	Mandiri	Tidak	Ada	Ada	Ada	...	Sudah	0.397988956	0.602011044	Sudah
74	Mandiri	Tidak	Tidak Ada	Ada	Tidak Ada	...	Sudah	0.371003917	0.628996083	Sudah
75	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.491088304	0.508911696	Sudah
76	Mandiri	Tidak	Ada	Ada	Ada	...	Sudah	0.397988956	0.602011044	Sudah

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	confidence (Belum)	confidence (Sudah)	prediction (Akses Universal)
77	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.322801419	0.677198581	Sudah
78	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.322801419	0.677198581	Sudah
79	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.491088304	0.508911696	Sudah
80	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.322801419	0.677198581	Sudah
81	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.322801419	0.677198581	Sudah
82	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
83	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
84	Mandiri	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.650994712	0.349005288	Belum
85	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.322801419	0.677198581	Sudah
86	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
87	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.491088304	0.508911696	Sudah
88	Mandiri	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.45317162	0.54682838	Sudah
89	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
90	BUMDes	Ada	Tidak Ada	Ada	Ada	...	Sudah	0.369326435	0.630673565	Sudah
91	Mandiri	Ada	Ada	Ada	Ada	...	Sudah	0.491088304	0.508911696	Sudah
92	BUMDes	Ada	Tidak Ada	Ada	Tidak Ada	...	Belum	0.322801419	0.677198581	Sudah
93	Mandiri	Ada	Ada	Ada	Tidak Ada	...	Belum	0.439924055	0.560075945	Sudah
94	Mandiri	Tidak	Tidak Ada	Tidak	Tidak Ada	...	Belum	0.443298512	0.556701488	Sudah
95	BUMDes	Ada	Tidak Ada	Ada	Ada	...	Belum	0.369326435	0.630673565	Sudah
96	BUMDes	Ada	Ada	Ada	Ada	...	Belum	0.339613062	0.660386938	Sudah

No	Pengelolaan SPAM	Rencana Kerja KPSPAM	Administrasi Pembukuan	Daftar Aset	Iuran	...	Akses Universal	confidence (Belum)	confidence (Sudah)	prediction (Akses Universal)
97	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
98	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah
99	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.481557564	0.518442436	Sudah
100	Mandiri	Ada	Ada	Ada	Ada	...	Belum	0.491088304	0.508911696	Sudah