KLASIFIKASI KARAKTERISTIK CALON MAHASISWA PADA INSTITUSI PENDIDIKAN MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus: STMIK LIKMI)

TESIS

Disusun sebagai Salah Satu Syarat untuk
memperoleh gelar Magister Komputer dari
Sekolah Tinggi Manajemen Infomatika dan Komputer LIKMI

Oleh:

Wilianti Aliman

NPM: 2019210051

PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER - LIKMI
BANDUNG

2020

KLASIFIKASI KARAKTERISTIK CALON MAHASISWA PADA INSTITUSI PENDIDIKAN MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus: STMIK LIKMI)

Oleh:

Wilianti Aliman

NPM: 2019210051

Bandung, 21 Desember 2020
Menyetujui,

Dr. Djajasukma Tjahjadi, S.E., M.T.
Pembimbing

PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER - LIKMI
BANDUNG

2020

Dipersembahkan untuk keluarga & orang terkasih
Chanfru, Chandra, Debby, Vanesa, Kevin Christian

ABSTRAK

KLASIFIKASI KARAKTERISTIK CALON MAHASISWA PADA INSTITUSI PENDIDIKAN MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus: STMIK LIKMI)

Oleh:
Wilianti Aliman
NPM: 2019210051
wiliantialiman@gmail.com
Sekolah Tinggi Manajemen Informatika & Komputer - LIKMI

Strategi Pemasaran haruslah disesuaikan dengan adanya penyesuaian dengan kebutuhan dan keinginan calon mahasiswa, sehingga banyak calon mahasiswa yang tertarik untuk mendaftar menjadi mahasiswa di kampus tersebut.

Data yang dibutuhkan dalam melakukan segmentasi karakteristik calon mahasiswa pada STMIK LIKMI adalah Data profil mahasiswa yang sudah menjadi mahasiswa dan data pelanggan yang mendaftar (PMB Online) dan beberapa data pendukung lainnya untuk memperdalam hasil analisis. Dari data-data tersebut dilakukan algoritma Classification untuk menentukan kelompok-kelompok pelanggan.

Penelitian ini menggunakan metode *Knowledge Discovery in Databases* (KDD) dan memanfaatkan algoritma *Naïve Bayes* untuk menemukan pola dari data calon mahasiswa STMIK LIKMI. Penelitian dilakukan dari data calon mahasiswa tahun 2016-2020. Pola yang ditemukan merupakan karakteristik calon mahasiswa yang didpatkan.

Pola yang didapatkan memberikan pengetahuan untuk mengetahui karakteristik pelanggan mahasiswa STMIK LIKMI, karakteristik calon mahasiswa yang paling memberikan keuntungan untuk STMIK LIKMI, dan strategi untuk pemaaran kedepannya.

ABSTRACT

CLASSIFICATION OF STUDENT ENROLLMENT CHARACTERISTICS FOR EDUCATION INSTITUTIONS USING NAVE BAYES METHOD (Case Study: STMIK LIKMI)

By:
Wilianti Aliman
NPM: 2019210051
wiliantialiman@gmail.com
Sekolah Tinggi Manajemen Informatika & Komputer - LIKMI

The marketing strategy must be adjusted according to the needs and desires of prospective students, so that many prospective students are interested in applying to become students on the campus.

The data needed in segmenting the characteristics of prospective students at STMIK LIKMI are profile data of students who have become students and data of registered customers (PMB Online) and some other supporting data to deepen the results of the analysis. From these data, Classification algorithm is used to determine customer groups.

This study uses the Knowledge Discovery in Databases (KDD) method and utilizes the Naïve Bayes algorithm to find patterns from the data of STMIK LIKMI student candidates. The research was conducted from the data of prospective students in 2016-2020. The patterns found are the characteristics of prospective students who are obtained.

The pattern obtained provides knowledge to determine the characteristics of STMIK LIKMI student customers, characteristics of prospective students that provide the most benefits for STMIK LIKMI, and strategies for future exposure.

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kepada Tuhan YME sehingga penulis dapat menyelesaikan Tesis dengan judul "***Klasifikasi Karakteristik Calon Mahasiswa Pada Institusi Pendidikan Menggunakan Metode Naïve Bayes (Studi Kasus: Stmik Likmi)***", untuk memenuhi syarat kelulusan program Pendidikan Pascasarjana (S2) jurusan Sistem Informasi di STMIK LIKMI Bandung.

Penulisan Tesis ini tidak lepas dari dukungan bantuan dan bimbingan dari berbagai pihak, khususnya para dosen, keluarga, kerabat, dan rekan mahasiswa. Dalam kesempatan kali ini, penulis mengucapkan terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Bapak Djajasukma Tjahyadi, S.E., M.T. selaku dosen pembimbing yang telah memberi bimbingan, dukugan, arahan, serta referensi hingga terselesaikannya Tesis ini.
2. Seluruh dosen dan karyawan STMIK LIKMI yang telah membantu dan membimbing penulis dalam menyelesaikan penyusunan Tesis ini dengan lancer.
3. Seluruh rekan seperjuangan penulis, khususnya semua rekan di STMIK LIKMI yang telah membantu penulis dalam menyusun Tesis dengan tulus dan sukarela, baik berupa dorongan semangat, ilmu, maupun ide.
4. Seluruh pihak yang tidak dapat disebutkan pada halaman ini yang telah menjadi penyemangat dan pendukung bagi penulis untuk menyelesaikan tugas akhir ini.

Penulis menyadari bahwa Tesis ini masih jauh dari sempurna, baik dari segi materi yang dijelaskan, maupun segi penyajian secara visual. Oleh karena itu, kritik membangun serta saran dari pembaca akan sangat diharapkan demi menyempurnakan Tesis ini.

Akhir kata, penulis berharap bahwa Tesis ini dapat memberikan ilmu dan pengetahuan yang bermanfaat dalam menambah wawasan bagi para pembaca yang

tertarik dalam mendalami Segmentasi Calon Mahasiswa menggunakan *Data Mining*, khususnya bagi penulis yang menyusun Tesis dengan topik ini.

Bandung, 21 Desember 2020

Penulis

DAFTAR ISI

ABSTRAK i

ABSTRACT ii

KATA PENGANTAR iii

DAFTAR ISI vii

DAFTAR GAMBAR viii

DAFTAR TABEL ix

DAFTAR LAMPIRAN viii

BAB I PENDAHULUAN 1

1.1. Latar Belakang 1

1.2. Rumusan Masalah 4

1.3. Tujuan Penelitian 4

1.4. Batasan Masalah 4

1.5. Pendekatan dan Metode Penelitian 5

1.6. Sistematika Penulisan 5

BAB II LANDASAN TEORI 7

2.1. Knowledge Discovery in Databases (KDD) 7

2.2. *Data Mining* 9

2.3. Segmentasi Target Pemasaran 14

2.4. Algoritma *Naïve Bayes* 17

2.5. Strategi Pemasaran Segmen Target Pasar 19

2.6. *Data Mining* pada Institusi Pendidikan 22

2.7. Penelitian Terkait 24

BAB III OBJEK DAN METODE PENELITIAN 27

3.1. Profil Calon Mahasiswa/ Pendaftar di STMIK LIKMI 27
3.2. Metode Pengumpulan data 28
3.3. Metode Penelitian 29
BAB IV HASIL DAN PEMBAHASAN 32
4.1. Identifikasi Kebutuhan Bisnis 32
4.2. Data Cleanning 33
4.3. Data Selection 37
4.3.1. Analisis *Information Gain* Atribut 39
4.3.2. Analisis Deskriptif Atribut 40
4.4. Data Transformation 53
4.5. *Data Mining* 53
4.6. Data Evaluation/ Interpretasi 32
4.7. Hasil Rekomendasi 33
BAB V Kesimpulan 67
5.1. Kesimpulan 67
5.2. Saran 68
DAFTAR PUSTAKA 69
LAMPIRAN 72

DAFTAR GAMBAR

Gambar 3. 1 Model Penelitian 30
Gambar 3. 2 Potongan Data dari Tabel Data Penerimaan Mahasiswa Baru 33
Gambar 4. 1 Analisis deskriptif atribut Gelombang 40
Gambar 4. 2 Analisis deskriptif data pilihan utama 41
Gambar 4. 3 Analisis deskriptif atribut agama 42
Gambar 4. 4 Analisis deskriptif jenis kelamin 43
Gambar 4. 5 Analisis deskriptif asal SMU 43
Gambar 4. 6 Analisis deskriptif atribut USM 44
Gambar 4. 7 Analisis deskriptif atribut Asal kota 45
Gambar 4. 8 Analisis deskriptif Beasiswa yang diajukan 46
Gambar 4. 9 Analisis deskriptif atribut Hasil Beasiswa 47
Gambar 4. 10 Analisis deskriptif atribut daftar ulang 48
Gambar 4. 11 Analisis deskriptif atribut Angkatan 49
Gambar 4. 12 Analisis deskriptif atribut pendapatan orang tua 49
Gambar 4. 13 Analisis Deskriptif Uang Saku Mahasiswa 50
Gambar 4. 14 Analisis deskriptif atribut Informasi kampus 51
Gambar 4. 15 Analisis deskriptif atribut alasan 52
Gambar 4. 16 Classifier Summary 54

DAFTAR TABEL

Tabel 2. 1 Gambaran Confusion Matrix 12
Tabel 2. 2 Penelitian terdahulu 24
Tabel 3. 1 Atribut Data Penelitian 29
Tabel 4. 1 Tabel Penjelasan Atribut 33
Tabel 4. 2 Informasi Data atribut 35
Tabel 4. 3 Verifikasi Kualitas Data 37
Tabel 4. 4 Atribut yang terpilih dari data PMB 38
Tabel 4. 5 Atribut yang dipilih dari data pendukung 38
Tabel 4. 6 Information Gain Data PMB 40
Tabel 4. 7 Classifier Summary Per Angkatan 55
Tabel 4. 8 Hasil Model Data Tahun 2016-2020 55
Tabel 4. 9 Karakteristik Calon Mahasiswa 2016-2020 58
Tabel 4. 10 Karakteristik Calon Mahasiswa Tahun 2016 60
Tabel 4. 11 Karakteristik Calon Mahasiswa Tahun 2017 60
Tabel 4. 12 Karakteristik Calon Mahasiswa Tahun 2018 61
Tabel 4. 13 Karakteristik Calon Mahasiswa Tahun 2019 61
Tabel 4. 14 Karakteristik Calon Mahasiswa Tahun 2020 62
Tabel 4. 15 Confusion Matrix 32
Tabel 4. 16 Evaluasi Perhitungan Sistem 32

DAFTAR LAMPIRAN

Lampiran 1 **Hasil Data Mining Tahun 2016-2020** 73

Lampiran 2 **Hasil Data Mining Tahun 2016** 76

Lampiran 3 **Hasil Data Mining Tahun 2017** 77

Lampiran 4 **Hasil Data Mining Tahun 2018** 78

Lampiran 5 **Hasil Data Mining Tahun 2019** 80

Lampiran 6 **Hasil Data Mining Tahun 2020** 81

BAB I

PENDAHULUAN

1.1. Latar Belakang

Persaingan akan selalu terjadi di dalam dunia bisnis dan perusahaan akan terus mengembangkan strategi bisnisnya agar dapat tetap bersaing dan menjadi unggul di bidangnya. Perusahaan dapat bertahan di dalam persaingan dengan memiliki pemikiran kreatif dan inovatif untuk memenuhi kebutuhan konsumen dan membangun hubungan yang baik dengan pelanggan, agar terciptanya sebuah bisnis yang terus berkesinambungan. Berdasarkan Survei yang dilakukan oleh satu organisasi bernama *American Express Survei* pada tahun 2011, sebanyak 80% dari pelanggan tidak kembali melakukan transaksi setelah mendapatkan penanganan yang buruk. Hal tersebut tentunya tidak membuat umur bisnis suatu perusahaan bertahan lama, maka perlu penanggulangan untuk sistem yang berfokus pada pelanggan dari segi penawaran awalnya maupun setelah menjadi pelanggan dapat tercipta loyalitas dari pelanggan tersebut.

Begitu pula dengan institusi-institusi yang bergerak di bidang pendidikan, selain aktivitas utama bisnis sebuah institusi pendidikan yaitu menyelenggarakan pendidikan yang baik, tentunya institusi tersebut harus mempertahankan pelanggan dan mitra agar tetap banyak calon mahasiswa yang masuk ke dalam kampus tersebut. Institusi pendidikan juga bersaing dalam menawarkan produk pendidikan yang dimiliki agar banyak calon-calon mahasiswa yang mendaftar. Persaingan tersebut memerlukan strategi yang tepat sehingga menaikan kualitas pemasaran institusi tersebut. Strategi tersebut haruslah disesuaikan dengan adanya penyesuaian dengan kebutuhan dan keinginan calon mahasiswa, sehingga banyak calon mahasiswa yang tertarik untuk mendaftar menjadi mahasiswa di kampus tersebut. Data yang terintergrasi dan baik prosesnya akan menjadi pondasi utama untuk menentukan strategi dan juga sistem operasional dalam menjalankan manajemen hubungan dengan pelanggan (Buttle, 2015:289). Strategi yang diterapkan untuk mempertahankan pelanggan tersebut juga memiliki banyak alternatif langkah, seperti mengeluarkan produk baru, atau melakukan promosi yang lebih luas lagi, dan salah

satu langkah awal untuk mengetahui strategi tersebut adalah dengan memahami betul data. Sehingga dapat ditemukan sebuah pola dari pengalaman yang ada dan tercipta peluang-peluang atau kemungkinan lain nantinya.

Data pelanggan yang terintergrasi tersebut akan diproses untuk menentukan langkah nyata yang digambarkan dalam strategi atau perkiraan arus bisnis kedepannya (Buttle,2015:287). Pengolahan tersebut akan memberikan informasi baru yang membantu seorang marketing mengetahui kebutuhan pelanggan dan dari data tersebut seorang marketing juga dapat membuat sebuah segmentasi kebutuhan agar dapat mengidentifikasi pelanggan yang menjadi utama untuk diberikan pelayanan lebih.

STMIK LIKMI merupakan salah satu institusi yang sudah lama bersaing didunia pendidikan, bahkan STMIK LIKMI sudah meraih banyak sekali prestasi. Salah satu prestasinya adalah menjadi sekolah tinggi di bidang Sistem Informasi pertama di Jawa Barat dan ketiga di Indonesia. Setelah melihat langsung dilapangan, STMIK LIKMI telah melakukan banyak penjualan produk ke sekolah-sekolah yang sudah banyak berelasi baik. Sekolah sebagai salah satu sumber utama untuk kampus melakukan promosi dan mendapatkan pelanggan untuk menjadi mahasiswa ini. Selain memanfaatkan mitra (sekolah yang berelasi baik), promosi dilakukan melalui media sosial yaitu Website, Instagram, dll. Cara yang lebih ringan lainnya adalah melalui acara-acara yang di adakan. Sejauh ini kampus hanya berusaha mengingat sekolah mana yang paling menguntungkan dan memasang iklan promosi pada program-program yang diperkirakan menguntungkan. Sehingga belum dapat dijelaskan secara jelas kategori atau ciri-ciri target pasar yang bisa dibilang akan menguntungkan kampus atau target pasar yang apabila hilang akan sangat disayangkan.

Setelah beberapa orang yang terjaring melalui seluruh cara promosi yang telah dilakukan, maka seorang marketing perlu melakukan *Followup* dengan tepat agar Ketika mengejar calon pelanggan seorang marketing dapat fokus pada kategori pelanggan yang menguntungkan. Pengetahuan akan peluang keberhasilan seorang pelanggan dapat diketahui melalui *Data Mining*. Pengetahuan akan profil pelanggan yang berhasil menjadi mahasiswa dengan yang kemungkinan besar tidak akan menjadi mahasiswa ini dapat

digunakan seorang marketing untuk menentukan langkah promosi maupun *Followup*. Sehingga target pasar yang akan dikejar menjadi lebih jelas dan fokus. Setelah mengenal profil pelanggan yang menguntungkan ini, maka untuk menentukan perilaku atau kebutuhan menjadi lebih mudah sebagai bahan promosi untuk menjaring profil pelanggan tersebut.

Penelitian pada kali ini digunakan untuk menentukan karakteristik profil pelanggan yang memiliki peluang besar akan menjadi mahasiswa dan untuk menentukan strategi bagi seorang marketing meningkatkan kuantitas pelanggan. Segmentasi bertujuan untuk menentukan kelompok-kelompok *margin* pelanggan berdasarkan kebutuhan dan perilakunya. Segmentasi dilakukan dengan cara mengelompokan data yang sama atau mirip.

Data yang dibutuhkan dalam melakukan segmentasi karakteristik calon mahasiswa pada STMIK LIKMI adalah Data profil mahasiswa yang sudah menjadi mahasiswa dan data pelanggan yang mendaftar dan beberapa data pendukung lainnya untuk memperdalam hasil analisis. Dari data-data tersebut dilakukan algoritma *Classification* untuk menentukan kelompok-kelompok pelanggan.

Berdasarkan pembahasan berikut, maka dilakukan penelitian untuk menentukan strategi pemasaran baru untuk meningkatkan hasil penjualan produk dari STMIK LIKMI dengan judul **"*KLASIFIKASI KARAKTERISTIK CALON MAHASISWA PADA INSTITUSI PENDIDIKAN MENGGUNAKAN METODE NAÏVE BAYES (STUDI KASUS: STMIK LIKMI)*"**.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Profil calon mahasiswa seperti apa yang berhasil menjadi mahasiswa STMIK LIKMI?
2. Profil calon mahasiswa seperti apa yang paling berpeluang akan menjadi mahasiswa STMIK LIKMI?
3. Strategi marketing apa saja yang dapat dijadikan saran untuk meningkatkan pemasaran bagi kelompok-kelompok calon mahasiswa tersebut?

1.3. Tujuan Penelitian

Tujuan penelitian pada penelitian ini adalah sebagai berikut:

1. Mengetahui informasi profil setiap calon mahasiswa yang berhasil menjadi mahasiswa STMIK LIKMI
2. Mengklasifikasikan profil calon mahasiswa yang berpeluang untuk menjadi mahasiswa dan peluang pelanggan gagal menjadi mahasiswa.
3. Menganalisa dan mengidentifikasi kebutuhan setiap profil calon mahasiswa sebagai strategi bagi seorang marketing dapat melakukan promosi yang tepat sasaran dan berguna untuk meningkatkan kuantitas pelanggan yang menjadi mahasiswa di STMIK LIKMI.

1.4. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini dilakukan dengan memanfaatkan data Penerimaan Mahasiswa Baru (PMB)
2. Penelitian ini dilakukan pada data Angkatan 2016-2020
3. Penelitian ini dilakukan menggunakan jenis penelitian klasifikasi yang ada pada *Data Mining*

1.5. Pendekatan dan Metode Penelitian

Pada penelitian ini akan menggunakan *Data Mining* untuk mendapatkan pengetahuan baru dari data pelanggan yang ada. Tahapan untuk melakukan penelitian ini mengikuti tahapan *Knowledge Discover in Database* (KDD) dan menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan kategori profil pelanggan. Adapun cara untuk mendapatkan data pada penelitian ini adalah sebagai berikut:

1. Studi Literatur

   Studi literatur yang digunakan dalam penulisan penelitian ini adalah mengumpulkan bahan referensi mengenai metode klasifikasi dari berbagai buku, jurnal, dan beberapa referensi lainnya.

2. Dokumentasi

   Dokumentasi merupakan metode dengan melakukan pengumpulan data atau dokumen, sampai data terkumpul menjadi satu dan siap untuk kepentingan penelitian.

3. Survei

   Metode pengumpulan data yang dilakukan melalui pengisian survei yang diajukan secara *online* kepada responden untuk mendapatkan data atau informasi yang berkaitan dengan objek penelitian.

1.6. Sistematika Penulisan

BAB I PENDAHULUAN

Dalam pendahuluan berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, kegunaan hasil, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Landasan teori berisi teori-teori yang terkait dengan *Data Mining* di pendidikan dan Algoritma *Naïve Bayes*

BAB III OBJEK DAN METODE PENELITIAN

Dalam bab ini berisi mengenai gambaran perusahaan yang menjadi objek penelitian dan tahap-tahap melakukan penelitian

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi mengenai pola dan hasil penelitian yang didapatkan setelah melakukan pengolahan data

BAB V KESIMPULAN DAN SARAN

Bab ini berisikan mengenai kesimpulan dari hasil perancangan dan saran untuk pengembangan lebih lanjut.

BAB II

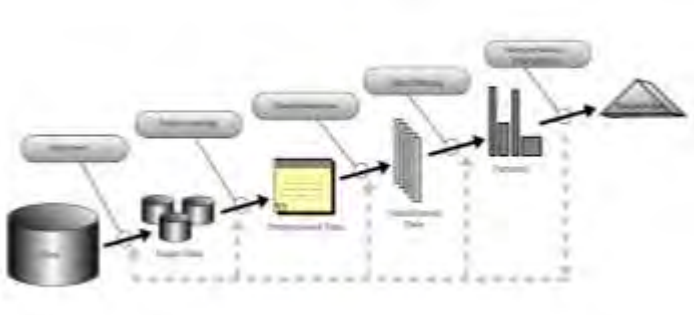
LANDASAN TEORI

2.1. Knowledge Discovery in Databases (KDD)

KDD merupakan salah satu alat bantu proses *Data Mining* yang sudah banyak digunakan sejak dulu dan bahkan banyak digunakan pada penelitian-penelitian baru-baru ini (Fayyad, 1996: 39). KDD adalah proses kompleks terkait penemuan pola dalam sekumpulan data yang ada (Oliveira, 2012: 26). KDD mampu melakukan pemetaan data yang tidak terstruktur menjadi data yang terstruktur dan rapih. Usama Fayyad dkk. memberikan sebuah pendapat mengenai KDD sebagai berikut:

"*The basic problem addressed by the KDD process is one of mapping low-level data into other forms that might be more compact, more abstract, or more useful*." (Fayyad, 1996: 40)

Alasan penggunaan KDD sebagai alat bantu adalah dikarenakan konsep dan tujuan hasil yang sama. Tujuan dari KDD adalah melakukan pengolahan data sehingga didapati satu informasi baru untuk menentukan strategi atau keputusan untuk dunia marketing. KDD juga memiliki konsep yang sama, fungsi dari KDD adalah mengolah data dari sebuah data yang jumlahnya banyak sehingga ditemukan pola-pola yang selanjutnya dapat dilakukan segmentasi sehingga menghasilkan informasi baru. Informasi yang dihasilkan tersebut dapat dikaitkan atau diimplematasikan pada banyak bidang, salah satunya dalam pemasaran. Pada Marketing, KDD dapat membantu pengolahan data pelanggan untuk mengidentifikasi kelompok pelanggan yang berbeda dan meramalkan perilaku mereka (Fayyad, 1996:37).



Gambar 2. 1
An Overview of the Steps That Compose the KDD Process.
(Sumber: Fayyad, 1996: 41)

KDD berfokus pada keseluruhan proses penemuan pengetahuan dari data, termasuk bagaimana data disimpan dan diakses, bagaimana algoritma dapat ditingkatkan ke aplikasi kumpulan data besar-besaran. Pada gambar 2.3 menjelaskan setiap tahapan pada KDD, proses-proses tersebut dijelaskan oleh oliveira sebagai berikut (Oliveira, 2012: 28):

1. *Data Selection*

   Tahap ini merupakan proses pemilihan data dan pemahaman mengenai penngaplikasian domain. Pengaplikasian domain bertujuan untuk memberikan konteks proyek dalam operasi perusahaan, dengan memahami bahasa bisnis dan menentukan tujuan proyek. Sedangkan pemilihan data bertujuan untuk memfokuskan analisis pada *subset* variabel atau sampel data, di mana penemuan harus dilakukan. Pada tahap ini, perlu untuk mengevaluasi *subset* minimum data yang akan dipilih, atribut yang relevan dan periode waktu yang tepat untuk dipertimbangkan.

2. *Data Processing*

   Tahap ini mencakup operasi dasar, seperti: menghilangkan *noise* atau *outlier*, mengumpulkan informasi yang diperlukan untuk memodelkan atau memperhitungkan kebisingan, memutuskan strategi untuk menangani atribut data yang hilang, dan menghitung informasi urutan waktu dan perubahan yang diketahui. Tahap ini juga mencakup masalah yang berkaitan dengan sistem manajemen basis data, seperti tipe data, skema, dan pemetaan nilai yang hilang dan tidak diketahui.

3. *Data Transformation*

   Tahap ini merupakan tahap pemrosesan menjadi data yang memiliki format yang sesuai untuk dilakukan *Data Mining.* Transformasi yang paling umum adalah: normalisasi data, agregasi data dan diskritisasi data. Normalisasi dilakukan pada setiap nilai dikurangi rata-rata dan dibagi dengan standar deviasi. Normalisasi dilakukan Ketika Data berasal dari beberapa tabel, karena beberapa algoritma hanya dapat menangani data kuantitatif dan beberapa lagi menangani data

kualitatif. Oleh karena itu, mungkin perlu untuk mendiskritisasi data, yaitu memetakan data kualitatif ke data kuantitatif, atau memetakan data kuantitatif ke data kualitatif.

4. *Data Mining*

Tahap ini terdiri dari menemukan pola dalam dataset yang sebelumnya disiapkan. Pada tahap ini juga dilakukan pemilihan algoritma yang tepat sesuai dengan kebutuhan, kemudian diterapkan pada data yang relevan, untuk menemukan hubungan implisit atau pola menarik lainnya.

5. Interpreter/ evaluation

Tahap ini terdiri dari menafsirkan pola yang ditemukan dan mengevaluasi kegunaan dan kepentingannya sehubungan dengan pengaplikasian *domain*.

KDD merupakan suatu lingkung yang lebih besar untuk menemukan sebuah pengetahuan baru dari data yang ada yang memanfaatkan *Data Mining* untuk mengekstrak polanya. Singkatnya KDD merupakan tahapan atau langkah-langkah yang digunakan untuk mengolah dan menjelajahi penemuan dari data, sedangkan *Data mining* merupakan pengaplikasian sebuah algoritma untuk menghasilkan pola-pola untuk KDD menemukan penemuan-penemuan dari data tersebut.

2.2. *Data Mining*

Data Mining merupakan tipe analisis yang lebih tinggi lagi, *Data Mining* menganalisis data dan memberikan informasi dengan pendekatan data seperti statistik deskriptif (frekuensi, rata-rata, median, mode, varian, standar deviasi), reduksi data, analisis statistik bivariat (tabulasi silang, korelasi), analisis statistik multivariat (regresi berganda, analisis faktor, analisis diskriminan, analisis klaster, penskalaan multidimensi, analisis konjoin), pohon keputusan dan jaringan saraf, dan visualisasi data (Buttle,2015 :258). *Data Mining* dapat diimplementasikan untuk keperluan pemasaran terutama untuk mengenal karakter dari calon pelanggan. *Data Mining* merupakan proses yang menggunakan teknik stastistik, matematika, kecerdasan buatan, dan *Machine learning*

untuk melakukan pengolahan data dan menentukan sebuah informasi baru yang berguna dari sebuah kumpulan data yang besar (Turban, 2006: 3).

Menurut Turban *Data Mining* bekerja dengan mencari pola dalam jumlah data yang besar:

"*the semi-automated process of extracting patterns from large amounts of unstructured data*" (Bygstad, 2012:59).

Data Mining digunakan untuk pemasaran memproses data yang menghasilkan sejumlah informasi, diantaranya adalah menghitung jumlah pembelian, menentukan target pasar untuk produk tertentu, memprediksi jumlah *customer* yang akan diberikan pelayanan lebih pada tahun mendatang, memprediksi jumlah *customer* yang akan kehilangan loyalitasnya, mendeskripsikan karakteristik pelanggan yang memberikan kita keuntungan lebih, mendeskripsikan perilaku konsumen yang digambarkan secara multidimensi pada segmentasi kebutuhan pelanggan, memberikan gambaran umum tentang kebutuhan-kebutuhan pelanggan untuk setiap segmen kebutuhan pelanggan, dan juga memberikan informasi panjang umur dari nilai yang diberikan oleh setiap pelanggan (Buttle, 2015:259)

Data Mining memiliki beberapa cakupan-cakupan teknik mengolah data sesuai dengan fungsinya masing-masing, diantaranya yaitu (Oliveira, 2012: 30):

1. *Association*

   Asosiasi bertujuan untuk menentukan hubungan antara beberapa atribut dalam database. Asosiasi berfokus pada memperoleh korelasi *multi-atribut*, dukungan yang memuaskan, dan batas kepercayaan.

2. *Classification*

   Klasifikasi bertujuan untuk menentukan peta setiap item data ke dalam sebuah kategori yang telah ditentukan. Sebagai contoh, klasifikasi dapat digunakan untuk mengidentifikasi resiko pemohon pinjaman.

3. *Clustering*

   Clustering memiliki kemiripan dengan klasifikasi, perbedaanya hanya pada kategori yang ditentukan. Pada klasifikasi, kategori tersebut telah ditentukan

namun pada *Clustering* tidak. *Clustering* menemukan kelompok alami dari item data, berdasarkan kesamaan metrik atau model kepadatan probabilitas

4. *Forecasting*

   Forecasting bertujuan memperkirakan nilai atribut tertentu di masa mendatang, berdasarkan pola rekaman. Pada *forecasting*, atribut yang diukur untuk setiap item untuk memprediksi perilaku di masa mendatang.

5. *Regression*

   Regresi bertujuan untuk memetakan setiap item data ke dalam kategori/ variable nilai prediksi.

6. *Sequence discovery*

   Sequence discovery bertujuan mengidentifikasi hubungan antar item dari waktu ke waktu Ini pada dasarnya dapat dianggap sebagai penemuan asosiasi atas basis data sementara. Misalnya, analisis urutan dapat dikembangkan untuk menentukan, jika pelanggan telah mendaftar untuk rencana A, maka rencana selanjutnya yang akan diambil oleh pelanggan dan dalam satu kerangka waktu diidentifikasi dalam *Sequence Discovery*.

7. *Visualization*

   Visualisasi digunakan untuk menyajikan data sehingga pengguna dapat melihat pola yang lebih kompleks, Visualisasi digunakan bersama dengan model *Data Mining* lainnya untuk memberikan pemahaman yang lebih jelas tentang pola atau hubungan yang ditemukan.

Data Mining memiliki dua tipe teknik mempelajari pola, yaitu: *Suppervised* dan *Unsupervised* (Oliveira, 2012: 32). *Suppervised* merupakan teknik pengelompokan data yang memiliki syarat bahwa dataset berisi ciri-ciri dan kebiasaan yang akan diprediksikan. Sebagai contoh, model yang diawasi dapat dilatih untuk mengidentifikasi pola-pola yang memungkinkan untuk mengklasifikasikan klien bank sebagai calon gagal bayar atau yang tidak membayar. Sedangkan *Unsuppervised* tidak memerlukan syarat seperti *Suppervised* karena *Unsuppervised* dapat dilatih untuk mengelompokkan pelanggan ke dalam kelompok yang tidak dikenal yang serupa.

Ketika melakukan *Data Mining* menggunakan metode yang dipilih ada kalanya perlu dilakukan untuk mengukur akurasi dari perhitungan yang dilakukan oleh metode tersebut. Pada bagian evaluasi untuk mengukur akurasi sebuah metode dan mengukur tingkat eror sebuah metode, maka dimanfaatkan perhitungan dari *Confusion Matrix* khususnya pada klasifikasi. Gambaran untuk *Confusion Matrix* dapat dilihat seperti pada tabel 2.1.

Tabel 2. 1
Gambaran *Confusion Matrix*

		Nilai yang sebenarnya	
		True	**False**
Nilai Prediksi	**True**	TP	FP
	False	FN	TN

Sumber: Adinugroho, 2018: 70

Hasil dari *Confusion Matrix* terdiri dari 4 istilah yaitu:

1. True Positive (TP)

   TP merupakan hasil prediksi benar dan nilai sebenarnya juga benar.

2. True Negative (TN)

   TN merupakan hasil prediksi salah dan nilai sebenarnya juga salah

3. False Positive (FP)

   FP merupakan hasil prediksi benar dan nilai sebenarnya salah

4. False Negative (FN)

   FN merupakan hasil prediksi salah dan nilai sebenarnya benar

Hasil dari penurunan *Confusion Matrix* adalah Akurasi, *Eror Rate*, *Sensitivity (Recall), Specifity, Precision, dan False Positive Rate.* Penjelasan dan perhitungan evaluasi adalah sebagai berikut (Adinugroho, 2018: 70):

1. Akurasi

   Akurasi merupakan hasil evaluasi dari perhitungan semua nilai prediksi yang benar dibagi dengan keseluruhan data. Nilai akurasi yang semakin mendekati 1.0 berarti semakin baik dan apabila mendekati 0 maka buruk. Akurasi dinyatakan dalam persamaan berikut:

$$\frac{TP + TN}{TP + TN + FN + FP}$$

2. *Eror Rate*

Eror Rate merupakan hasil evaluasi dari perhitungan semua nilai yang diprediksi salah dengan keseluruhan data. Nilai terbaik adalah mendekati 0 dan nilai paling buruk mendekati angka 1.0. *Eror Rate* dinyatakan dalam persamaan berikut:

$$\frac{FN + FP}{TP + TN + FN + FP}$$

3. *False Positive Rate (FPR)*

False Positive Rate merupakan hasil evaluasi dari perhitungan jumlah prediksi positif yang salah dibagi dengan keseluruhan dari jumlah kelas yang negatif. Nilai terbaik dari FPR adalah 0 dan nilai terburuknya adalah 1.0. FPR dinyatakan dalam persamaan berikut:

$$\frac{FP}{TN + FP}$$

4. *Sensitivity (Recall)*

Sensitivity (Recall) atau disebut juga *True Positive Rate* merupakan hasil evaluasi dari perhitungan prediksi positif yang benar dibagi dengan jumlah keseluruhan kelas yang positif. Nilai terbaik adalah mendekati 1.0 dan terburuknya mendekati 0. *Sensitivity* dinyatakan dalam persamaan berikut:

$$\frac{TP}{TP + FN}$$

5. *Specificity (True Negative Rate)*

Specificity (True Negative Rate) merupakan hasil evaluasi dari perhitungan keseluruhan jumlah prediksi salah yang benar dibagi dengan keseluruhan jumlah kelas yang salah. Nilai terbaik adalah mendekati 1.0 dan terburuknya adalah mendekati 0. *Specificity* dinyatakan dalam persamaan berikut:

$$\frac{TN}{TN + FP}$$

6. *Precision (Positive Predictive Value)*

Precision merupakan hasil evaluasi dari perhitungan jumlah keseluruhan nilai prediksi positif yang benar dibagi dengan jumlah keseluruhan prediksi kelas yang benar. Nilai terbaik adalah mendekati 1.0 dan terburuknya adalah mendekati 0. *Precision* dinyatakan dalam persamaan berikut:

$$\frac{TP}{TP + FP}$$

Berdasarkan pembahasan dari beberapa ahli dapat ditarik kesimpulan bahwa *Data Mining* dapat menjadi satu alat untuk melakukan segmentasi pelanggan terutama pada Institusi Pendidikan, karena menggunakan data yang tidak terstruktur untuk menemukan pola-pola sehingga dapat menjadi sebuah data terstruktur untuk menentukan klasifikasi karakteristik calon mahasiswa.

2.3. Segmentasi Target Pemasaran

Segmentasi merupakan hasil dari proses klasifikasi, menurut jurnal yang ditulis oleh Dicky Nofriansyah menjelaskan pengertian mengenai klasifikasi.

"*Metode Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui."* (Nofriansyah, 2016: 82)

Sedangkan segmentasi lebih diartikan pada kelompok-kelompok yang dipisahkan menurut aturan dan fungsi tertentu. Segmentasi pelanggan membagi pelanggan yang heterogen menjadi himpunan bagian yang homogen, bahkan hingga ke tingkat pelanggan yang unik (Buttle, 2015:64).

"*Segmentasi pasar adalah proses membagi pasar keseluruhan untuk suatu barang dan jasa ke dalam beberapa kelompok yang lebih kecil, dimana para anggota masing-masing kelompok mempunyai kesamaan terhadap faktor-faktor yang mempengaruhi permintaan."* (Mulyana, 2019:24)

Segmentasi dapat ditentukan melalui analisis terhadap perilaku pelanggan membeli atau menggunakan barang maupun jasa perusahaan, setelah mendapatkan data mengenai perilaku tersebut maka dilakukan pengelompokan pada perilaku yang memiliki kesamaan (Oliveira, 2012:74).

Pada penelitian ini taget pemasaran yang dimaksudkan adalah Calon Mahasiswa. Pada dasarnya kebutuhan target pasar sangatlah beragam (Mulyana, 2019:24) dan tentu hal tersebut juga berlaku pada sebuah Institusi pendidikan. Proses yang dilakukan untuk memilih dan mengelompokan pasar yang luas sehingga produk yang ditawarkan sesuai dengan kebutuhan kelompok yang dituju adalah proses segmentasi target pasar.

Segmen pasar yang homogen diidentifikasi dengan menggunakan faktor-faktor atau dasar-dasar segmentasi untuk menentukan karakteristik pasar sebagai berikut:

1. Faktor Demografis Geografis

   Faktor Demografis Geografis merupakan populasi, jaringan transportasi, iklim, tipe, komersial, pendirian ritel, media, persaingan, pola pertumbuhan, peraturan, biaya hidup. Segmentasi geografis ini membagi-bagi pasar berdasar unit-unit geografis, karena kebutuhan khusus yang hanya dibutuhkan pada daerah tersebut.

2. Faktor Demografis Pribadi

   Faktor Demografis Pribadi merupakan data umur, jenis kelamin, mobilitas, penghasilan, pekerjaan, status perkawinan, besarnya rumah tangga, agama, kesukuan, dan pendidikan. Segmentasi demografis ini memisahkan pasar ke dalam kelompok-kelompok yang didasarkan pada variabel-variabel tersebut.

Menurut Yudhohusodo dalam penelitian yang dilakukan Yulida, tingkat pendapatan seseorang dapat digolongkan dalam 4 golongan yaitu (Yulida, 2012,140):

1. Golongan yang berpenghasilan rendah (low income group) yaitu pendapatan rata-rata dari Rp.150.000 perbulan.
2. Golongan berpenghasilan sedang (Moderate income group) yaitu pendapatan rata-rata Rp.150.000 – Rp.450.000 perbulan.
3. Golongan berpenghasilan menengah (midle income group) yaitu pendapatan rata-rata yang diterima Rp.450.000 – Rp.900.000perbulan.
4. Golongan yang berpenghasilan tinggi (high income group) yaitu rata-rata pendapatan lebih dari Rp.900.000.

Pada tahun 1998, Upah Minimum Rata-Rata Nasional (UMR) berada di sekitar Rp 153.000 atau setara dengan Rp 3.600.000 pada tahun 2020. Istilah UMR dapat digolongkan pada kelas ekonomi sedang (menengah kebawah).

3. Faktor Psikografis dan keperilakukan

Faktor Psikografis dan Keperilakuan merupakan yaitu kepribadian, sikap, manfaat produk yang diinginkan, persepsi, loyalitas merek, pengalaman menggunakan produk, tingkat penggunaan produk, struktur pembelian (informal ke formal, otonomi ke bersama), daur hidup keluarga, kelas sosial, keinovatifan, kepemimpinan opini, kepentingan membeli.

Segmentasi pelanggan merupakan salah satu hasil yang dari teknik klasifikasi yang didapatkan dari proses *Data Mining*. Segmentasi pelanggan dapat membantu pengolahan data menentukan profil dari pelanggan tersebut, dan hal-hal yang membuat mereka menjadi puas. Setelah menemuka informasi seluruh pelanggan, maka segmentasi hadir untuk melakukan sleksi berdasarkan demografi dan kategori pelanggan.

Segmentasi pelanggan merupakan suatu konsep yang menggabungkan kebutuhan dan perilaku pelanggan yang sama. Segmentasi harus dapat diidentifikasi, dikuantifikasi, dialamatkan, dan dengan ukuran yang cukup agar layak ditangani (Srivastava, 2002 :18). Segmentasi merupakan proses yang membantu perusahaan untuk melihat lingkup yang lebih kecil dari pasaran yang luas agar perusahaan tidak hanya dapat lebih mengerti kebutuhan pelanggan namun akan membantu perusahaan untuk menggambarkan sebuah sistem marketing, produksi, dan pelayanan lebih efektif dari sebelumnya (Oliveira, 2012:82).

Berdasarkan beberapa pengertian diatas, maka dapat disimpulkan bahwa segmentasi merupakan hasil dari analisis yang berbentuk kelompok-kelompok sesuai kategori yang dibutuhkan untuk mencapai tujuan dari sebuah bisnis. Segmentasi ditentukan dengan memanfaatkan metode atau konsep *Data Mining*

2.4. Algoritma *Naïve Bayes*

Klasifikasi membentuk model data yang mampu untuk membedakan data dala kelas yang berbeda menggunakan aturan dan fungsi tertentu. Salah satu algoritma yang digunakan pada teknik klasifikasi adalah Algoritma *Naïve Bayes*. Algoritma tersebut memanfaatkan perhitungan probabilitas dan statistik (Nofriansyah, 2016: 87). Pada awalnya, teori tersebut diadopsi dari teori yang dikemukakan oleh Thomas Bayes yaitu perhitungan peluang yang akan terjadi berdasarkan pengalaman yang sudah terjadi (Adinugroho, 2018:72). Teori Bayes dituliskan dalam perhitungan seperti berikut:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Berdasarkan rumus tersebut dapat dijelaskan sebagai berikut:

1. P(A|B) merupakan peluang terjadinya kejadian A dengan syarat kejadian B telah terjadi
2. P(A|B) merupakan peluang terjadinya kejadian B dengan syarat kejadian A telah terjadi
3. P(A) merupakan peluang terjadinya kejadian A, tanpa pengaruh kejadian lain
4. P(B) merupakan peluang terjadinya kejadian B, tanpa pengaruh kejadian lain

Klasifikasi pada Bayes menjelaskan bahwa variabel B merupakan sekumpulan variabel pada data, sedangkan variabel A merupakan kelas pada data. Apabila kelas pada suatu data memiliki hubungan non-deterministik dengan variabel-variabel maka A dan B dapat dianggap variabel acak. Naïve dalam hal ini menyempurnakan rumus Bayes dengan asumsi variabel yang mempengaruhi bersifat mandiri (Adinugroho, 2018:74). Teori tersebut dimanfaatkan oleh Naïve untuk memperhitungkan sebuah peluang dengan variabel bebas. Adapun alur dari metode *Naïve Bayes* adalah sebagai berikut (Adinugroho, 2018:75):

1. Menghitung nilai peluang dari setiap kelas
2. Menghitung nilai peluang $P(B_i \mid A)$ atau peluang bersyarat tiap-tiap variabel terhadap kelas.
3. Menentukan label ke arah positif atau negatif dari peluang yang dihitung

Teori Naïve Bayes dinyatakan dalam perhitungan sebagai berikut:

$$P(A|B) = \frac{P(A)\pi_{i=1}^{d} P(B_i|A)}{P(B)}$$

Berdasarkan rumus tersebut terdapat perbedaan sedikit dengan teori Bayes yaitu pada $P(B_i|A)$ yang berarti peluang terjadinya setiap variabel B_i dengan syarat A. Peluang B bersifat konstan untuk setiap kejadian A, sehingga nilai peluang B dapat dihilangkan dari persamaan. Sedangkan untuk melakukan penentuan kelas dihitung dari nilai maksimal $P(A_k \mid B)$ untuk setiap kelas yang terdapat pada data (Adinugroho, 2018:74). Perhitungan tersebut dinyatakan dalam persamaan sebagai berikut:

$$C = \underset{k\in\{1,\dots K\}}{argmax} P(A_k)\pi_{i=1}^{d} P(B_i|A)$$

K dalam persamaan tersebut merupakan himpunan kelas yang terdapat pada data yang ada. Metode ini cenderung lebih mudah dipahami karena parameter yang digunakan dapat dipakai secara berulang, bahkan Ketika *dataset* bertambah besar.

Naïve Bayes merupakan algoritma yang mudah diimplementasikan pada banyak kasus dengan hasil yang baik, namun kelemahan algoritma tersebut tidak dapat menggambarkan sebuah keterkaitan antar variabel (Syarli & Ashari, 2016:74). Algoritma *Naïve Bayes* mampu menganalisis kelompok dengan kemungkinan tertinggi dari kelompok-kelompok yang telah dibuat dan kondisi yang diberikan. Menurut Yuda Septian Nugroho, Naïve Bayes merupakan algoritma yang mampu memberikan hasil dengan bukti hasil akurasi dan kecepatan yang tinggi ketika diaplikasikan pada database (Nugroho, 2013:3).

Berdasarkan penelitian dan penjelasan dari beberapa sumber, Naïve *Bayes* dapat menjadi salah satu algoritma yang cocok dengan penelitian ini dikarenakan tujuan dari algoritma tersebut sama dengan tujuan penelitian ini yaitu untuk menentukan sebuah probabilitas dan pengklasifikasian data yang dalam hal ini diimplementasikan pada data calon mahasiswa dan mahasiswa yang telah berhasil mendaftar pada STMIK LIKMI.

2.5. Strategi Pemasaran Segmen Target Pasar

Menurut KBBI, Strategi dapat diartikan sebagai suatu rencana cermat mengenai kegiatan untuk mencapai sasaran khusus. Secara tidak langsung pemasaran diartikan juga sebagai suatu proses perencanaan dan menjalankan konsep, harga, promosi dan distribusi sejumlah ide, barang, dan jasa untuk menciptakan pertukaran yang mampu memuaskan tujuan individu dan organisasi. Pemasaran merupakan suatu proses sosial dan managerial yang melibatkan kegiatan-kegiatan yang memungkinkan individu dankelompok mendapatkan kebutuhan dan keinginan melalui pertukaran dengan pihak lain dan untuk mengembangkan hubungan pertukaran (Khasanah, 2015:162).

Pemasaran tidak hanya diimplementasikan pada sebuah barang namun juga dapat dilakukan dalam sebuah jasa. Pemasaran dalam konteks jasa pendidikan adalah sebuah proses sosial dan managerial untuk mendapatkan hal yang dibutuhkan dan diinginkan melalui penciptaaan penawaran dan pertukaran produk yang bernilai dengan pihak lain dalam bidang pendidikan (Khasanah, 2015:165). Pemasaran pendidikan juga membutuhkan hal pendukung lainnya untuk dipadukan sehingga dapat menghasilkan sebuah strategi pemasaran yang dapat digunakan untuk memenangkan persaingan (Khasanah, 2015:166). Berikut merupakan hal/ unsur penting yang dapat menjadi pendukung sebuah strategi pemasaran tersebut:

1. *Product*

   Produk merupakan keseluruhan objek atau proses yang memberikan sejumlah nilai kepada konsumen. Produk jasa pendidikan merupakan produk yang dapat menawarkan reputasi, prospek, mutu pendidikan yang baik, prospek dan peluang yang cerah bagi para siswa untuk menentukan pilihan-pilihan yang diinginkannya.

2. *Price*

   Harga merupakan sejumlah uang yang harus dibayarkan oleh konsumen untuk mendapatkan suatu produk. Harga dalam jasa pendidikan merupakan seluruh biaya yang dikeluarkan untuk mendapatkan jasa pendidikan yang ditawarkan.

3. *Place*

Lokasi dalam hal ini merupakan tempat sebuah jasa tersebut, yang dalam hal ini adalah jasa pendidikan beroperasi. Lokasi yang strategis, nyaman dan mudah dijangkau akan menjadi daya tarik tersendiri.

4. *Promotion*

Promosi merupakan kegiatan mengkomunikasikan sebuah produk dipasaran yang berhubungan langsung dengan masyarakat. Promosi dilakukan untuk menginformasikan dan meyakinkan target pasar agar membeli jasa yang ditawarkan. Promosi penjualan juga dapat dilakukakan melalui pameran pendidikan, bazar pendidikan dan investasi, melakukan kontak langsung dengan target pasar dan juga melakukan kegiatan hubungan dengan masyarakat.

5. *People*

Orang dalam pemasaran yang dimaksud adalah setiap pihak yang berkaitan dalam menyampaikan dan memasarkan produk kepada target pasar, seperti: Bagian *Marketing, Customer service*, dan Tata usaha.

6. *Physical evidence*

Bukti fisik yang dimaksud dalam hal ini adalah lingkungan fisik tempat jasa diciptakan dan langsung berinteraksi dengan konsumennya. Terdapat dua macam bukti fisik, diantaranya yaitu: keputusan-keputusan yang dibuat oleh pemberi jasa mengenai desain dan tata letak gedung seperti desain kelas, gedung sekolah, perpustakaan, lapangan olahraga dan lain-lain. Selain itu terdapat bukti fisik lainnya seperti Bukti pendukung merupakan nilai tambah yang bila berdiri sendiri tidak akan berdiri sendiri dan memiliki peran yang sangat penting dalam proses jasa seperti raport, catatan siswa dan lain-lain.

7. *Process*

Proses merupakan sebuah prosedur/ mekanisme dalam rangkaian aktivitas dalam menyampaikan jasa dari produsen ke konsumen. Pada dunia pendidikan, proses merupakan kegiatan pendidikan tersebut dilaksanakan hingga menciptakan lulusan yang unggul.

Ketika target pasar telah ditentukan maka perlu adanya evaluasi mengenai peluang yang didapatkan dari setiap segmen dan menentukan target pasar yang akan menjadi target utama untuk meningkatkan pemasaran (Mulyana, 2019: 26). Setiap segmen pasar diperhitungkan laba, daya tarik jangka panjang, dan tujuang jangka panjang dari perusahaan. Pemilihan target pasar tersebut dapat mempertimbangkan beberapa strategi berikut:

1. *Undifferentiated Marketing*

   Undifferentiated Marketing merupakan usaha untuk meninjau pasar secara keseluruhan yang mendasarkan pada kesamaan-kesamaan kebutuhan konsumen dengan mencoba mengembangkan produk tunggal untuk memenuhi keinginan banyak orang.

2. *Differentiated marketing*

   Differentiated marketing merupakan usaha untuk mengidentifikasi kelompok-kelompok pasar tertentu dengan membagi pasar ke dalam dua kelompok atau lebih. Misal suatu perusahaan yang menawarkan berbagai macam merek dengan kemasan yang berbeda-beda untuk ditujukan pada segmen pasar yang berbeda.

3. *Concentrated* Marketing

   Concentrated Marketing merupakan usaha untuk memusatkan usaha pemasaran perusahaan pada satu atau beberapa kelompok pembeli saja dengan menempatkan posisinya pada segmen pasar yang terbesar yang paling menguntungkan. Misal penawaran sepeda motor jenis bebek dengan empat tak yang mempunyai potensi pasar yang tinggi.

Pada dunia pendidikan maupun dunia bisnis lainnya tetap dapat membutuhkan strategi pemasaran yang tepat. Strategi pemasaran dilakukan dengan langkah sebagai berikut (Khasana, 2015:168):

1. Identifikasi Pasar

Identifikasi pasar dalam hal ini merupakan suatu langkah yang mengidentifikasi dan menganalisis pasar untuk mengetahui kebutuhan target pasar, hal tersebut termasuk atribut yang dijadikan acuan untuk kepentingan konsumen pendidikan.

2. Segmentasi Pasar dan Posisi Produk

   Segmentasi merupakan suatu langkah yang mengelompokan sejumlah target pasar ke dalam kelompok-kelompok sesuai kebutuhan dan karakteristik dari target pasar tersebut. Sedangkan posisi produk yang dimaksud adalah sebuah proses memberikan citra pada produk sehingga produk atau jasa yang dipasarkan memiliki karakteristik dan perbedaan agar target pasar dapat membedakan produk yang ditawarkan dengan produk dari perusahaan lainnya.

3. Diferensiasi Produk

   Diferensiasi produk merupakan sebuah langkah memberikan keunikan pada produk yang ditawarkan. Diferensiasi dapat dilakukan dengan beberapa strategi diantaranya adalah sebagai berikut:

   a. Cara penawaran yang berbeda
   b. Keunggulan biaya yang sangat efisien
   c. Fokus pada konsumen yang diutamakan
   d. Komunikasi pemasaran yang sangat baik

2.6. *Data Mining* pada Institusi Pendidikan

Peran *Data Mining* pada sebuah Institusi pendidikan juga penting untuk membangun citra institusi yang berpengaruh pada kuantitas mahasiswa yang akan mendaftar. Oluseye dalam penelitiannya berpendapat bahwa pelayanan untuk membangun hubungan yang baik dengan mahasiswa memiliki peran yang sangat penting untuk meningkatkan kepuasan pelanggan di Perguruan Tinggi atau Universitas yang akan menghasilkan sebuah umpan balik dan bagi mahasiswa yang merasa puas akan secara alami merekomendasikan kampusnya kepada orang-orang lain disekitarnya (Oluseye,2014:58). Pada penelitian ini memberikan cara untuk Institusi mengerti kebutuhan mahasiswa agar merasa terpuaskan dengan pelayanan yang ada sehingga

pelayanan yang baik akan membantu seorang bagian pemasaran untuk melakukan promosi secara masal melalui rekomendasi mahasiswanya. Penelitian yang dilakukan Oluseye dapat dipahami juga bahwa sebuah rekomendasi menjadi salah satu yang mempengaruhi sebuah institusi menjadi pilihan bagi calon mahasiswa baru lainnya. Maka dari itu, pemilihan data atau atribut rekomendasi bisa menjadi salah satu pilihan dikarenakan adanya keterhubungan.

Data mengenai keberadaan keluarga atau saudara dari calon mahasiswa, kondisi finansial orang tua dan uang saku calon mahasiswa ternyata memberikan pola khusus mengenai penerimaan mahasiswa baru. Pada penelitian yang dilakukan Salim menjelaskan bahwa adanya hubungan mengenai demografi calon mahasiswa dengan karakteristik sebuah Kampus (Salim, 2017:2). Hal tersebut memberi pengaruh pada proses pendaftaran dan proses penerimaan mahasiswa baru. Pada penelitian lainnya yang dilakukan oleh Kahara dan kawan-kawan menjelaskan bahwa selain data finansial dan nilai dari calon mahasiswa, data mengenai mata pelajaran yang calon mahasiswa senangi, cita-cita dari calon mahasiswa dan bayangan pekerjaan yang didapatkan ketika lulus kuliah juga memberikan pola untuk memperlihatkan kelompok-kelompok dan memprediksi kemungkinan kelompok tersebut akan berhasil melakukan pendaftaran (Kahara, 2016:700).

Sedangkan penelitian yang dilakukan oleh Hananto menjelaskan bahwa peran *Data Mining* pada Institusi pendidikan sangatlah penting karena dengan adanya pengolahan data sebuah Institusi pendidikan mendapatkan pengetahuan baru mengenai pelanggannya yang dalam hal ini adalah calon mahasiswa. Ketika mengenal setiap mahasiswa maupun calon mahasiswa, setiap marketing dapat mengembangkan metode untuk melakukan promosi dan interaksi yang lebih sesuai target atau menyesuaikan dengan perilaku calon mahasiswa. Hasil pengolahan data juga dapat digunakan untuk menentukan produk/ program yang efektif untuk dilakukan terlebih dahulu dan juga dapat menentukan pelanggan yang memiliki potensi paling menguntungkan untuk dipertahankan. Selain itu melalui *Data Mining* dapat memprediksi calon mahasiswa yang akan batal mendaftar.

Berdasarkan dua penelitian tersebut dapat disimpulkan bahwa ada dua sudut pandang dalam pengolahan data menggunakan *Data Mining* yang diimplementasikan pada Perguruan tinggi sendiri yaitu dari sisi pelayanan pada mahasiswa dan pendekatan pada setiap calon mahasiswa. *Data Mining* memberikan cara untuk sebuah Institusi mengenal kemungkinan-kemungkinan yang terjadi melalui pengenalan akan perilaku calon mahasiswa melalui eksplorasi data.

2.7. Penelitian Terkait

Berikut merupakan rangkuman dari penelitian terdahulu yang ada kaitannya dengan penelitian ini.

Tabel 2. 2
Penelitian terdahulu

Studi kasus	**Metode, algoritma dan tools**	**Data Input**	**Data Output**	**Hasil penelitian**
"*Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques*" - Vera L'ucia Migu'eis Oliveira.				
Perusahaan bahan makanan	KDD berfokus pada *Clustering*, *Classification*, dan Assosiasi K-means	- Data transaksi pembelian, - Data pelanggan, dan - Data produksi	- Data produk yang paling menguntungkan selama satu periode dan - Data pelanggan yang menguntungkan	- Mengidentifikasi pelanggan yang akan beralih ke pesaing sehingga dapat dilakukan pendekatan/ *Followup* Kembali agar tidak hilang - meningkatkan produksi produk yang diprediksi meningkatkan keuntungan
Metode *Naive Bayes* Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)				
Penerimaan Mahasiswa Baru di Universitas x Sulawesi	*Classification* dan prediction menggunakan Naïve *Bayes*	-Data Nilai Ujian -Data Formulir pendaftaran calon mahasiswa	-Tabel hasil klasifikasi dari data pilihan jurusan dengan rata-rata nilai ujian saringan masuk - probabilitas kelulusan pilihan jurusan pertama Ketika mengikuti ujian saringan masuk	Pilihan pertama yang lebih tinggi yaitu 274 data (80%) dan Pilihan Kedua sebanyak 73 (17%) data serta yang tidak Lulus sebanyak 14 data. (3%)

Studi kasus	Metode, algoritma dan tools	Data Input	Data Output	Hasil penelitian
"*Customer* Relationship Management Approach and Student Satisfaction in Higher Education Marketing" oleh Ogunnaike Olaleke Oluseye, Borishade Taiye Tairat, Jeje Olumide Emmanuel				
Siswa Academic Staff Union of University (ASUU)	Identifikasi perilaku siswa menggunakan SPSS, Regresi *Data Mining*	-Survei kepada 72 siswa	-hipotesis hubungan antara kepuasan mahasiswa dengan manajemen hubungan orang tua dan	kesediaan mahasiswa untuk merekomendasikan universitasnya kepada orang lain meningkat ketika siklus hidup mahasiswa di universitas dikelola dengan baik dan manajemen hubungan orang tua yang kuat di universitas mengarahkan mahasiswa untuk merekomendasikan universitas mereka secara positif kepada orang lain
"*Data Mining* Model for Predicting Student Enrolment in STEM Courses in Higher Education Institutions" oleh Stephen Kahara Wanjau				
Fakultas Sains, Matematika, dan teknologi di salah satu Institusi pendidikan di Kenya	Prediksi Penerimaan Mahasiswa Baru menggunakan CRIPS-DM, alogoritma Naïve *Bayes* dan K-NN	- Data ketertarikan dan motivasi Mahasiswa - Data akademik calon mahasiswa di SMA - Data Konteks edukasi - Data Demografi	Data klasifikasi karakteristik calon mahasiswa yang mendaftar ke perguruan Tinggi dan memprediksi calon mahasiswa yang berhasil menjadi mahasiswa	5 atribut penting yang mempengaruhi seorang masuk ke sebuah perguruan tinggi adalah data Nilai akhir calon mahasiswa di SMA, Inspirasi dari guru, career flexibility, kepekaan sebelum memasuki Perguruan tinggi, dan level nilai Matematika.
"Predicting Student Enrollment Based on Student and College Characteristics" oleh Ahmad Slim				
Universitas New Mexico (UNM)	Menggunakan LR dan SVMs	- Data Demografi - Data Nilai di SMA - Data Saudara terdahulu di Kampus yang sama	Klasifikasi karakteristik seorang calon mahasiswa mau untuk mendaftar dan menjadi mahasiswa	Pada tahun 2017 diprediksi bahwa mahasiswa akan lebih banyak mendaftar pada musim gugur dibanding dengan musim panas dan musim semi.

Studi kasus	Metode, algoritma dan tools	Data Input	Data Output	Hasil penelitian
		- Data Financial anak dn orang tua - Data Lingkungan tempat tinggal - Data Biaya yang diberikan oleh Kampus - Data tanggal pendaftaran -Data Forrmulir pendaftaran		

Berdasarkan penelitian yang sudah ada, pada penelitian ini akan dilakukan penelitian pada institusi pendidikan yang menggunakan metode *Data Mining*, teknik *classfication* untuk menentukan profil calon mahasiswa yang berhasil menjadi mahasiswa berdasarkan data yang ada.

BAB III

OBJEK DAN METODE PENELITIAN

3.1. Profil Calon Mahasiswa/ Pendaftar di STMIK LIKMI

Penelitian ini akan dilakukan pada salah satu Institusi pendidikan dibidang Manajemen dan Informatika yang berlokai di Bandung yaitu STMIK LIKMI. Penelitian akan dilakukan untuk kepentingan marketing pada Institusi tersebut. STMIK LIKMI merupakan salah satu Institusi pendidikan yang berlokasi di Bandung dan sudah berdiri sejak 1985. Pada tahun 1985, dimana masih sangat sedikit masyarakat Indonesia yang mengerti komputer, beberapa dosen perguruan tinggi negeri, swasta dan profesional di bidang komputer bertekad untuk memperkenalkan ilmu komputer terapan kepada masyarakat dengan mendirikan Lembaga Ilmu Komputer Terapan Indonesia (LIKTI). LIKTI mempelopori sistem praktek langsung dalam belajar komputer. Seiring perkembangan pada tahun 1989, nama LIKTI diubah menjadi LIKMI (Lembaga Ilmu Komputer dan Manajemen Indonesia).

Marketing STMIK LIKMI atau yang mereka sebut dengan sebutan RAM (*Representatif of Administrasi and Marketing*) telah melakukan banyak upaya untuk menarik banyak calon mahasiswa melalui program-program beasiswa, iklan-iklan di media sosial, dan juga keikutsertaan di berbagai pameran pendidikan. Calon Mahasiswa STMIK LIKMI merupakan seluruh siswa SMA yang sudah lulus maupun yang akan lulus pada 1-2 tahun yang akan datang. Setiap siswa SMA yang berasal dari jurusan IPA maupun IPS juga dapat menjadi mahasiswa STMIK LIKMI. Calon Mahasiswa STMIK LIKMI berasal dari banyak sekolah dan kota yang terdapat di Indonesia. Beberapa sekolah juga sudah memiliki hubungan yang baik dengan STMIK LIKMI, sehingga memudahkan siswa SMA dari sekolah tersebut untuk mengenal dan mendaftar ke STMIK LIKMI.

Selama 2016 hingga 2020 calon mahasiswa yang melakukan pendaftaran banyak dijumpai oleh STMIK LIKMI pada saat pameran pendidikan, banyak siswa-siswa SMA yang tertarik pada promo Beasiswa maupun program studi di STMIK LIKMI. Pameran pendidikan yang diakukan oleh STMIK LIKMI rata-rata berasal dari sekolah swasta yang

berada di Bandung dan daerah sekitar Bandung. Calon mahasiswa yang mendaftar melalui jalur beasiswa maupun jalur ujian juga selalu banyak setiap tahunnya. Beasiswa yang ditawarkan kepada calon mahasiswa setiap tahunnya rata-rata tidak jauh berbeda yaitu: beasiswa penuh, beasiswa bebas uang pangkal, dan potongan uang pangkal.

Sedangkan tahun 2020 terjadi lonjakan yang cukup tinggi pada pendaftaran yang dilakukan namun ketika sudah diproses ternyata hanya sekitar 10% yang pada akhirnya serius dan melakukan proses akhir yaitu daftar ulang untuk menjadi mahasiswa. Hal tersebut cukup membuat kebingungan dan membutuhkan evaluasi untuk bagian RAM. Evaluasi tersebut dibutuhkan untuk menentukan strategi yang tepat dan memiliki peluang yang besar agar pada tahun berikutnya peningkatan mahasiswa dapat dirasakan.

3.2. Metode Pengumpulan data

Pengumpulan data dilakukan dengan melakukan Survei langsung pada STMIK LIKMI. Data yang dibutuhkan adalah data formulir pendaftaran calon mahasiswa STMIK LIKMI dari tahun 2016 hingga 2020. Data yang digunakan untuk menemukan pola adalah data pada tahun 2016-2019 dengan jumlah data sebanyak 2105 data. Sedangkan data untuk dilakukan pengujian adalah data pada tahun 2020 dengan jumlah data 765 data.

Data-data pendukung diperoleh melalui Survei menggunakan *google form* kepada mahasiswa STMIK LIKMI. Data pendukung tersebut berhubungan dengan demografi calon mahasiswa yang sudah menjadi mahasiswa dan juga alasan-alasan yang mendukung mereka untuk berkuliah di STMIK LIKMI. Pemilihan atribut tersebut didasarkan pada kebutuhan pengetahuan akan kondisi atau latar belakang dari mahasiswa yang berhasil mendaftar. Kebutuhan data mengenai finansial orang tua dan jumlah uang saku dari mahasiswa juga dibutuhkan dengan pertimbangan untuk memperkirakan kondisi ekonomi dari mahasiswa.

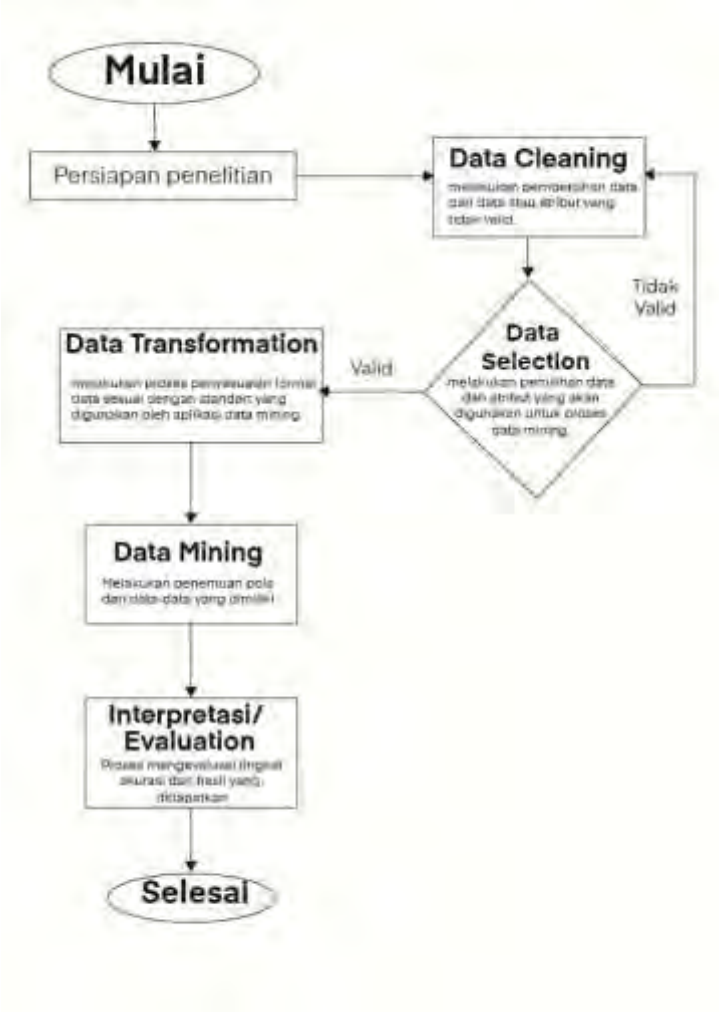
Berdasarkan pertimbangan kebutuhan data tersebut diharapkan bisa mendapatkan sebuah pola yang memetakan karakteristik calon mahasiswa yang berhasil menjadi mahasiswa. Pemilihan data yang telah dilakukan memberikan kesimpulan ada 13 atribut yang dijabarkan pada Tabel 3.1.

Tabel 3. 1
Atribut Data Penelitian

Atribut	Deskripsi
Asal Sekolah	Berisi informasi mengenai sekolah asal calon mahasiswa
Program pilihan	Berisi informasi mengenai program studi dan jurusan di LIKMI yang dipilih oleh calon mahasiswa ketika mendaftar
Jalur Pendaftaran	Berisi informasi mengenai jalur calon mahasiswa melakukan pendaftaran
Beasiswa yang diajukan	Berisi informasi mengenai beasiswa yang diajukan/ dipilih oleh calon mahasiswa ketika mendaftar
Beasiswa yang didapatkan	Berisi informasi mengenai hasil beasiswa yang didapatkan oleh calon mahasiswa
Daftar Ulang	Berisi informasi mengenai status calon mahasiswa melakukan daftar ulang atau tidak
Financial orangtua	Berisi informasi kisaran pendapatan orang tua calon mahasiswa yang berhasil menjadi mahasiswa di LIKMI
Jumlah Uang saku	Berisi informasi mengenai jumlah uang saku yang didapatkan oleh mahasiswa
Pihak yang merekomendasi	Berisi informasi mengenai pihak yang merekomendasikan LIKMI sebagai tempat kuliah
Alasan	Berisi informasi mengenai alasan dari mahasiswa memilih STMIK LIKMI untuk menjadi Kampus yang dipilih

3.3. Metode Penelitian

Desain penelitian yang digunakan menggunakan model *KDD.* Gambar 3.1 berikut merupakan gambaran alur penelitian ini dilakukan dari awal hingga akhir.



Gambar 3. 1
Model Penelitian

Tahapan pada KDD yang diImplementasikan pada penelitian ini adalah sebagai berikut:

1. *Data Cleanning*

   Data cleaning merupakan proses membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan penulisan.

2. *Data Selection*

   Pada tahap ini dilakukan penyesuaian kebutuhan data dengan tujuan bisnis perusahaan. Pemahaman akan hasil akhir dibutuhkan untuk mengerti kebutuhan

data dan atribut yang akan diolah oleh *Data Mining*. Data yang digunakan adalah data yang didapat dari formulir Pendaftaran Mahasiswa Baru Jalur Undangan dan juga data hasil proses pengajuan beasiswa. Data pendukung lainnya didapatkan dari hasil Survei pada mahasiswa STMIK LIKMI, data yang dibutuhkan tersebut merupakan data eknomi keluarga, data demografi seperti Hobi dan bidang minat, data alasan memilih STMIK LIKMI sebagai kampus dan juga data keberadaan anggota keluarga atau saudara di STMIK LIKMI sebelumnya.

3. Data Transformation

Pada penelitian ini dilakukan proses penyesuaian format data yang akan digunakan pada perangkat lunak *Data Mining*. Penelitian ini menggunakan aplikasi/ Software yang mendukung *Data Mining* yaitu DBFViewer dan Microsoft Excell, dalam melakukan transformasi data.

4. Data Mining

Pada tahap ini mencoba menemuka pola-pola dari data yang ada dan menentukan karakteristik dari data yang telah diolah. Algoritma yang digunakan untuk menemukan pola dan membentuk klasifikasi adalah Naïve *Bayes*. Klasifikasi ditentukan untuk menemukan kesamaan data dari data set mahasiswa Ketika pertama kali mendaftar dan Ketika sudah menjadi mahasiswa di STMIK LIKMI. Software yang digunakan adalah WEKA, pada alat bantu tersebut menawarkan banyak teknik *Data Mining* seperti: preprocessing data, klasifikasi, regresi, pengelompokan, aturan asosiasi, dan visualisasi, dengan antarmuka yang mudah. Pada penelitian ini akan menggunakan WEKA *Toolkit* 3.8.4

5. Interpretasion/ Evaluation

Setelah menemukan pola dan karakteristik yang ada, maka dapat dievaluasi akurasi dari data dan menentukan probabilitas untuk setiap kategori untuk mengetahui kemungkinan calon mahasiswa untuk menjadi mahasiswa.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Identifikasi Kebutuhan Bisnis

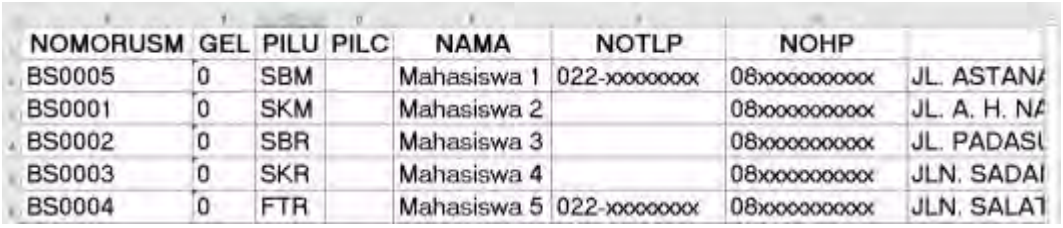
STMIK LIKMI selalu membuka Pendaftaran Mahasiswa Baru (PMB) setiap tahunnya, dengan presentasi yang cukup stabil setiap tahunnya. Namun pada tahun 2020, terjadi kesenjangan antara calon mahasiswa yang mendaftar dengan calon mahasiswa yang menjadi mahasiswa dalam hal ini dimaksudkan adalah calon mahasiswa yang melakukan Daftar Ulang. Kejadian ini memberikan sebuah hipotesa-hipotesa awal mengenai kesenjangan tersebut, salah satu faktor yang menjadi sorotan adalah perihal Komitmen dan kondisi dari calon mahasiswa.

Pada penelitian ini bertujuan untuk mencari karakteristik calon mahasiswa yang kemungkinannya lebih besar untuk berkomitmen atau serius menjadi mahasiswa dari STMIK LIKMI berdasarkan kondisi dari calon mahasiswa. Kondisi yang diperhatikan dalam hal ini adalah latar belakang dari calon mahasiswa dan kelas ekonomi dari calon mahasiswa. Sehingga dari data yang diperoleh dapat menghasilkan sebuah informasi mengenai target pasar calon mahasiswa untuk menyesuaikan strategi marketing, informasi mitra (sekolah) yang memberikan keuntungan bagi STMIK LIKMI, dan informasi karakteristik calon mahasiswa yang kemungkinannya lebih tinggi untuk memiliki komitmen.

Strategi untuk mendapatkan informasi tersebut adalah menggunakan *Data Mining* untuk menemukan klasifikasi/ segmentasi. Langkah pertama yang dilakukan untuk mendapatkan informasi tersebut adalah dengan melakukan permintaan data pendaftaran dan data daftar ulang STMIK LIKMI. Data yang diberikan dari STMIK LIKMI tentang PMB adalah data tahun 2016-2020 per Juni 2020 berjumlah 2866.

4.2. Data Cleanning

Pada tahap ini Data yang ada dibersihkan dari kesalahan penulisan dan melakukan pengumpulan serta pelengkapan data. Pada penelitian ini digunakan hanya data Formulir dan data Daftar Ulang calon mahasiswa,. Data yang digunakan adalah data dari tahun 2016 hingga 2020. Data yang digunakan untuk melakukan penemuan pola adalah data 2016 hingga 2019, sedangkan data pada tahun 2020 digunakan untuk pengujian akurasi dari pola yang ditemukan.



NOMORUSM	GEL	PILU	PILC	NAMA	NOTLP	NOHP	
BS0005	0	SBM		Mahasiswa 1	022-xxxxxxxx	08xxxxxxxxxx	JL. ASTANA
BS0001	0	SKM		Mahasiswa 2		08xxxxxxxxxx	JL. A. H. NA
BS0002	0	SBR		Mahasiswa 3		08xxxxxxxxxx	JL. PADASU
BS0003	0	SKR		Mahasiswa 4		08xxxxxxxxxx	JLN. SADAI
BS0004	0	FTR		Mahasiswa 5	022-xxxxxxxx	08xxxxxxxxxx	JLN. SALAT

Gambar 3. 2
Potongan Data dari Tabel Data Penerimaan Mahasiswa Baru
(Sumber: Data Base Penerimaan Mahasiswa Baru STMIK LIKMI)

Pada gambar 3.2 merupakan gambaran sekilas dari keadaan data Penerimaan mahasiswa baru yang sudah diperoleh berupa dokumen Excel. Dataset tersebut masih ada beberapa pencatatan data yang masih kosong/ Null, maka dapat dilihat bahwa kondisi dataset yang ada belum baik dan membutuhkan pembersihan. Data yang didapatkan dari STMIK LIKMI yaitu data Formulir dan Data Calon mahasiswa yang melakukan daftar ulang, data tersebut berisikan atribut seperti pada tabel 4.1.

Tabel 4. 1
Tabel Penjelasan Atribut

NO	Atribut	Keterangan
1	Nomor USM	Berisi Nomor pokok calon mahasiswa ketika mendaftar
2	Gelombang	Berisi Gelombang pendaftaran calon mahasiswa ketika mendaftar
3	Pilihan Utama	Berisi data pilihan program studi utama yang dipilih oleh calon mahasiswa
4	Pilihan Cadangan	Berisi data pilihan program studi cadangan yang dipilih oleh calon maha siswa
5	Nama	Berisi data Nama dari calon mahasiswa yang mendaftar
6	Nomor Telepon	Berisi data Nomor telepon dari calon mahasiswa yang mendaftar
7	Nomor HP	Berisi data Nomor HP dari calon mahasiswa yang mendaftar
8	Alamat 1	Berisi data Alamat yang tertera pada KTP calon mahasiswa

NO	Atribut	Keterangan
9	Alamat 2	Berisi data Alamat kedua sesuai KTP calon mahasiswa
10	Kota	Berisi data Kota tempat asal calon mahasiswa sesuai dengan KTP
11	Kode Pos 1	Berisi data Kode pos dari alamat tempat asal calon mahasiswa sesuai dengan KTP
12	Alamat di Bandung 1	Berisi data Alamat tempat Calon mahasiswa tinggal selama di Bandung
13	Alamat di Bandung 2	Berisi data Alamat tempat Calon mahasiswa tinggal selama di Bandung
14	Kode Pos 2	Berisi data Kode pos tempat calon mahasiswa tinggal selama di bandung
15	Tempat lahir	Berisi data tempat kelahiran dari calon mahasiswa
16	Tanggal Lahir	Berisi data tanggal calon mahasiswa lahir
17	Agama	Berisi data agama dari calon mahasiswa yang mendaftar
18	Golongan Darah	Berisi data golongan darah dari calon mahasiswa yang mendaftar
19	Jenis Kelamin	Berisi data jenis kelamin dari calon mahasiswwa yang mendaftar
20	Warga Negara	Berisi data status kewarganegaraan calon mahasiswa yang mendaftar
21	SMA asal	Berisi data asal SMA calon mahaiswa yang mendaftar
22	Sumbangan Sukarela	Berisi data sumbangan sukarela yang diberikan oleh calon mahasiswa
23	Tambahan Sumbangan	Berisi data tambahan sumbangan sukarela yang diberikan oleh calon mahasiswa
24	USM	Berisi data calon mahasiswa melakukan Ujian saringan masuk atau tidak ketika mendaftar, data ini menentukan jenis jalur pendaftaran yang dilakukan oleh calon mahasiswa
25	Beasiswa yang diajukan	Berisi data beasiswa yang diajukan oleh calon mahasiwa
26	Nilai Beasiswa	Berisi data nominal beasiswa yang diajukan oleh calon mahasiswa
27	Daftar Ulang	Berisi data calon mahasiswa yang melakukan daftar ulang setelah mendapatkan surat hasil penerimaan mahasiswa

Setelah melihat keseluruhan data, dapat dilihat bahwa ada beberapa atribut yang seluruh data yang tercatat kosong seperti atribut; Pertama-tama dilakukan filter data menggunakan ms.excell atribut yang sekiranya bsa dibuang dan dibersihkan karena tidak memiliki data pencatatan atau data yang tercatat memiliki isi yang sama. Atribut data yang tidak memiliki pencatatan data adalah Pilihan jurusan cadangan, nilai beasiswa dan sumbangan tambahan. Sedangkan untuk data yang isinya sama adalah status kewarganegaraan karena semua mahasiswa LIKMI berasal dari negara Indonesia menurut data yang ada.

Langkah selanjutnya adalah mencoba memasukan data ke dalam pemrosesan data di dalam aplikasi WEKA, beberapa atribut memiliki pola yang terlalu beragam/ Data Unik dan data bersifat Null sehingga tidak ditemukan pola dengan metode klasifikasi Naïve Bayes. Atribut data yang dimaksud seperti: Nomor USM, Nama, Kewarganegaraan, golongan darah, Nomor telepon, Nomor HP, Alamat 1, Alamat 2, Alamat Bandung 1, Alamat bandung 2, kodepos 1, kodepos 2, kodepos asal SMU, Sumbangan tambahan, Pilihan cadangan dan tanggal Lahir. Tabel 4.2 menjelaskan missing value, distinct dan unique dari atribut-atribut tersebut, dimana apabila missing value yang cukup tinggi dan unique dari atribut tinggi maka akan mempengaruhi kualitas atribut memburuk. Sehingga atribut tersebut dapat tidak dipakai atau dibersihkan untuk meningkatkan kualitas data.

Tabel 4. 2
Informasi Data atribut

Atribut	**Missing Value**	**Disticnt**	**Unique**
Nomor USM	0 (0%)	1675	1231 (43%)
Gelombang	0 (0%)	9	0 (0%)
Pilihan Utama	37 (1%)	28	2 (0%)
Pilihan cadangan	2869 (100%)	0	0 (0%)
Nama	1 (0%)	2800	2737 (95%)
Nomor Telepon	2305 (80%)	553	543 (19%)
Nomor HP	251(9%)	2525	2443 (85%)
Alamat 1	30 (1%)	2822	2805 (98%)
Alamat 2	1590 (55 %)	1226	1189 (41%)
Kota	87 (3%)	354	191 (7%)
Alamat bandung 1	1782 (62%)	1080	1073 (37%)
Alamat bandung 2	2562 (89%)	284	269 (9%)
Kodepos 1	1191 (42%)	934	669 (23%)
Kodepos 2	2222 (77%)	186	88 (3%)
Tempat lahir	140 (5%)	419	277 (10%)
Tanggal Lahir	94 (3%)	1671	989 (34%)
Agama	0 (0%)	7	1 (0%)
Golongan darah	2420 (84%)	5	0 (0%)
Jenis Kelamin	67 (2%)	2	0 (0%)
Kewarganegaraan	0 (0%)	1	0 (0%)
Asal SMU	35 (1%)	1172	918 (32%)
Sumbangan Sukarela	0 (0%)	18	7 (0%)
Sumbangan Tambahan	2869 (100%)	0	0 (0%)
USM	0 (0%)	2	0 (0%)
Beasiswa yang diajukan	0 (0%)	4	0 (0%)
Hasil Beasiswa	0 (0%)	7	0 (0%)
Daftar Ulang	0 (0%)	2	0 (0%)

Data yang memiliki kualitas buruk adalah data yang kurang lengkap (data hilang/ *Missing*), banyak noise (Outlier), dan data yang tidak konsisten sehingga menimbulkan ambigu (Adinugroho dan Arum, 2018 :28). Data yang memiliki kualitas buruk dapat dibersihkan sehingga menjadi data yang memiliki kualitas baik, tahapan tersebut yang biasa disebut juga tahap persiapan data. Langkah-langkah yang dilakukan adalah sebagai berikut:

1. Membersihkan atribut

   a. Atribut dengan tingkat *Distinct* (Kemiripan data) kurang dari 2

      Apabila kelompok data kurang dari 2 maka tidak akan memberikan pola apapun atau secara sederhana hasilnya tidak memberikan kemungkinan apapun (absolut). Data atribut dengan distinct kurang dari 2 adalah kewarganegaraan.

   b. Atribut dengan *Missing Value* yang tinggi

      Pada aplikasi Weka terdapat nilai *default* untuk pemisahan data uji dan data latih, dengan minimal data latih adalah 66%. Berarti dengan mempertimbangkan kebutuhan bisnis juga maka atribut yang dihapus adalah Pilihan cadangan, Nomor Telepon, Alamat Bandung 1, Alamat Bandung 2, , sumbangan sukarela, Golongan darah, Sumbangan Tambahan, Kode Pos 1, dan Kode Pos 2.

   c. Atribut dengan Unique yang tinggi

      Pada tahap ini agar data yang dimiliki tidak memberikan hasil yang ambigu ketika diproses maka dengan mempertimbangkan kebutuhan bisnis, atribut yang dihapus adalah Nomor USM, Alamat 2, tempat lahir, tgl lahir, Nama dan Nomor HP.

2. Membersihkan baris data kosong (tidak dapat dipakai)

   Setelah melihat lebih jelas dan detail bahwa atribut sudah sesuai dengan kebutuhan informasi, maka pembersihan dilakukan pada baris data yang hilang agar kualitas data lebih baik.

Tabel 4. 3
Verifikasi Kualitas Data

Atribut	Missing Value	Distinct	Unique
Gelombang	0	9	0
Pilihan Utama	37	28	2
Asal Kota	87	354	191
Agama	0	7	1
Jenis Kelamin	67	2	0
Asal SMU	35	1172	918
USM	0	2	0
Beasiswa yang diajukan	0	4	0
Hasil Beasiswa	0	7	0
Daftar Ulang	0	2	0

Berdasarkan data pada tabel 4.3 tersebut maka akan dilakukan pembersihan Kembali terhadap data bernilai null atau kosong pada data calon mahasiswa STMIK LIKMI sebagai berikut:

a. Data kosong pada atribut pilihan utama program studi sebanyak 37 data akan dihapus, maka jumlah data akan menjadi 2829 data.
b. Data kosong pada atribut asal kota sebanyak 87 data akan dihapus maka jumlah data akan menjadi 2742
c. Data kosong pada atribut jenis kelamin sebanyak 67 data akan dihapus maka jumlah data akan menjadi 2675 data.
d. Data kosong pada atribut asal SMU sebanyak 35 data akan dihapus maka jumlah data akan menjadi 2640 data.

4.3. Data Selection

Setelah melakukan pembersihan, pada tahap ini data akan disiapkan untuk dilakukan pengolahan menggunakan *Data Mining*. Data Calon Mahasiswa yang didapatkan dari STMIK LIKMI memiliki 28 atribut setelah digabungkan dan dilakukan normalisasi dan dilakukan pemilihan atribut sebanyak 11 atribut. Pada tahap ini ditambahkan atribut tambahan yaitu Angkatan untuk mengetahui perkembangan jumlan setiap tahunnya, lebih jelasnya dijelaskan pada Tabel 4.4.

Tabel 4. 4
Atribut yang terpilih dari data PMB

Atribut	**Keterangan**
Angkatan	Berisi informasi tahun Angkatan dari calon mahasiswa
Asal SMU	Berisi informasi asal SMU dari calon mahasiswa
Asal Kota	Berisi informasi asal kota dari calon mahasiswa
Gelombang	Berisi informasi gelombang ketika calon mahasiswa mendaftar
Agama	Berisi informasi agama dari calon mahasiswa
USM	Berisi informasi calon mahasiswa yang mengikuti jalur ujian dengan tidak
Beasiswa yang diajukan	Berisi informasi beasiswa yang diajukan oleh calon mahasiswa ketika mendaftar
Pilihan Utama	Berisi informasi pilihan program studi utama calon mahasiswa
Jenis Kelamin	Berisi informasi jenis kelamin calon mahasiswa
Hasil Beasiswa	Berisi informasi Hasil beasiswa yang didapatkan calon mahasiswa
DU	Berisi informasi calon mahasiswa yang mendaftar ulang dengan tidak

Sedangkan data yang didapatkan pada proses Survei akan digunakan untuk mendukung menjawab kebutuhan informasi yang dibutuhkan tim Marketing. Data yang didapatkan adalah sebanyak 86 data dengan 11 atribut. Jumlah yang tidak mendukung memberikan dampak data tidak dapat digabungkan untuk dianalisis. Namun data tersebut dapat digunakan untuk mengetahui beberapa hal khusus seperti kondisi perekonomian calon mahasiswa STMIK LIKMI dan juga alasan-alasan pemilihan kampus yang dapat digunakan sebagai bahan evaluasi STMIK LIKMI. Maka terpilih 4 atribut yang sesuai untuk mendukung informasi tersebut, atribut yang digunakan dijelaskan pada Tabel 4.5.

Tabel 4. 5
Atribut yang dipilih dari data pendukung

Atribut	**Keterangan**
Pendapatan Orang Tua	Berisi informasi besaran pendapatan orang tua dari calon mahasiswa

Atribut	Keterangan
Uang Saku	Berisi informasi besaran uang saku yang didapatkan oleh calon mahasiswa perbulannya
Informasi Kampus	Berisi informasi asal informasi calon mahasiswa mengetahui kampus STMIK LIKMI
Alasan	Berisi informasi alasan calon mahasiswa memilih STMIK LIKMI sebagai kampusnya

Atribut pada Tabel 4.5 diolah secara terpisah dan hasilnya hanya digunakan sebagai tambahan informasi (pendukung) dari karakteristik yang sudah ditemukan nantinya melalui pengolahan data PMB.

1. Analisis *Information Gain* Atribut

Sebelum melakukan penghapusan, pengujian pada atribut dilakukan dengan melakukan pemilihan data atribut berdasarkan fitur *selected attribute* untuk melihat merit dari setiap atribut. Tahapan ini juga memastikan bahwa atribut yang digunakan benar sesuai dengan tujuan bisnis dan menimbulkan pola. Manfaat dari proses ini adalah untuk mencegah *Overfitting*, dimana model yang dibentuk terlalu menyamai data latih.

Pertama-tama akan dilakukan penentuan rangking dari setiap atribut dengan menggunakan fitur pada aplikasi Weka yaitu Selected Attributes. Selected Attributes pada penelitian ini menggunakan Attributed Evaluator InfoGainAttributesEval untuk mengevaluasi kualitas atribut dengan menghitung keuntungan informasi terhadap class dan juga menggunakan Method Ranker untuk melihat urutan dari setiap atribut.

Hasil yang didapatkan adalah subset eror dan tidak dapat digunakan, jadi penghapusan beberapa atribut dilakukan hingga sistem dapat melakukan pemrosesan pada atribut yang sudah bersih dan memiliki kualitas cukup baik. Sehingga atribut dipilih dan disesuaikan dengan kebutuhan, pada tahap ini juga ditambahkan atribut Angkatan untuk mengetahui jumlah persentase calon mahasiswa pertahunnya. Hasil yang didapatkan melalui proses *Selected attribute* tersebut diurutkan berdasarkan rangking *Information Gain* pada tabel 4.5.

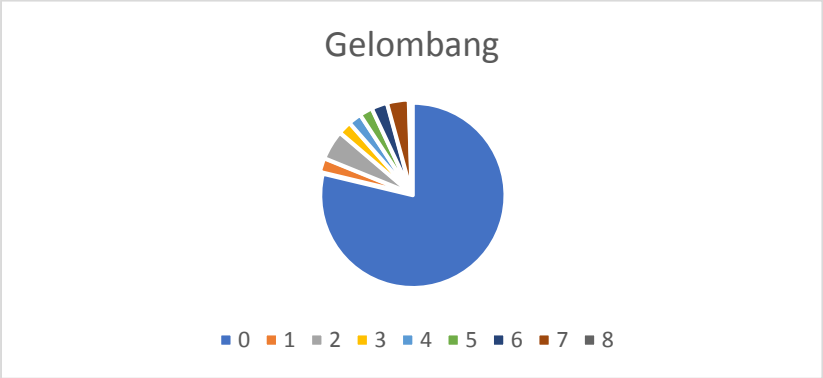
Tabel 4. 6
Information Gain Data PMB

Atribut	***Information Gain***
Asal SMU	0.3954
Asal Kota	0.264
Gelombang	0.1531
Agama	0.1224
USM	0.1114
Beasiswa yang diajukan	0.1106
Pilihan Utama	0.0561
Hasil Beasiswa	0.0379
Jenis Kelamin	0.0129

2. Analisis Deskriptif Atribut

Setelah melakukan pembersihan atribut tersebut, maka dihasilkan atribut tersisa adalah Gelombang, Pilihan Utama, Kota, Agama, Jenis Kelamin, SMU asal, Sumbangan Sukarela, USM, Beasiswa yang diajukan, Hasil Beasiswa, dan Daftar Ulang (DU) yang dijelaskan lebih detail dengan analisis deskriptif. Data analisis deskriptif dari setiap atribut yang ada pada data calon mahasiswa. Hasil yang didapatkan adalah sebagai berikut:

a. Analisis deskriptif atribut Gelombang



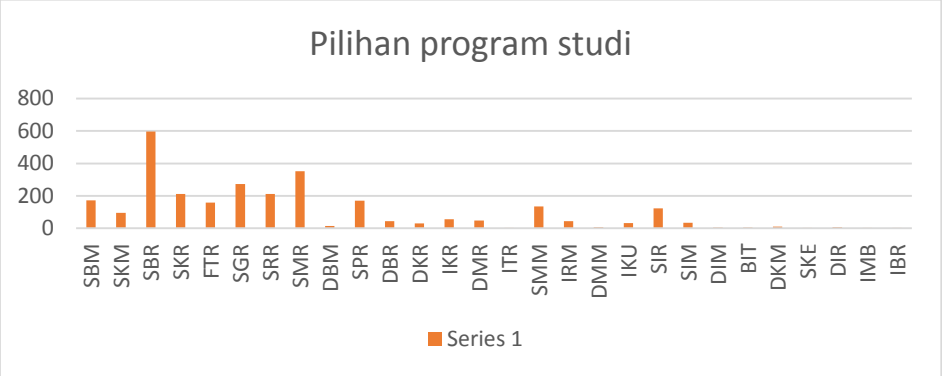
Gambar 4. 1
Analisis deskriptif atribut Gelombang

Berdasarkan Gambar 4.1 dapat dilakukan analisis sebagai berikut:

1) Data Gelombang terbanyak diraih pada gelombang 0 atau gelombang awal dengan jumlah data 2259.
2) Data Gelombang paling sedikit diraih pada gelombang 8 atau gelombang terakhir dengan jumlah data 14.

3) Data gelombang 1 berjumlah 69, gelombang 2 berjumlah 146, gelombang 3 berjumlah 66, gelombang 4 berjumlah 65, gelombang 5 berjumlah 64, gelombang 6 berjumlah 78, dan gelombang 7 berjumlah 108.

b. Analisis deskriptif atribut Pilihan utama (Jurusan/ program studi)



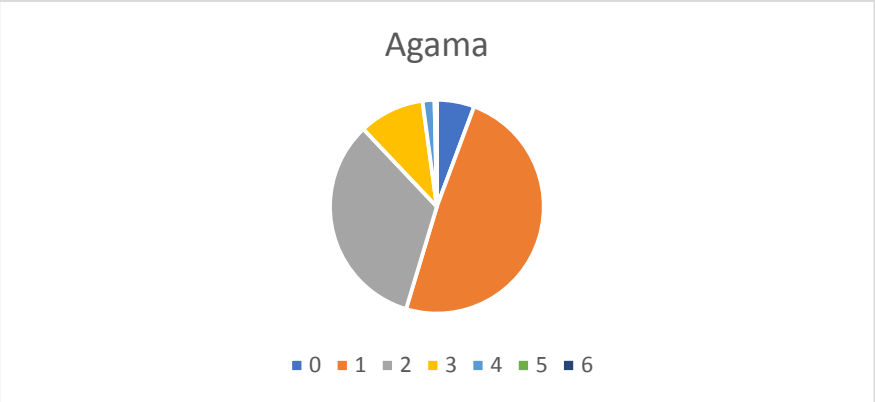
Gambar 4. 2
Analisis deskriptif data pilihan utama

Berdasarkan Gambar 4.2 dapat dilakukan analisis sebagai berikut:

1) Data Pilihan utama program studi di STMIK LIKMI yang terbanyak adalah SBR (Sarjana Sistem Informasi bidang keahlian Manajemen Bisnis Kelas Reguler) dengan jumlah data 596.
2) Data Pilihan utama program studi di STMIK LIKMI yang paling sedikit adalah IBR (Internasional Bisnis Kelas Reguler) dan SKE (Sarjana Sistem Informasi Bidang keahlian Komputerisasi Akuntansi)
3) Data Pilihan utama terdiri dari SBM (171 data), SKM (95 data), SKR (212 data), FTR (159 data), SGR (273 data), SRS (212 data), SMR (351 data), DBM (13 data), SPR (170 data), DBR (43 data), DKR (30 data), IKR (56 data), DMR (48 data), ITR (3 data), SMM (134 data), IRM (43 data), DMM(7 data), IKU (32 data), SIR (122 data), SIM (33 data), DIM (5 data), BIT (4 data), DKM (10 data), DIR (6 data), data IMB (2 data).
4) Dapat dilihat bahwa beberapa program studi yang cukup banyak peminatnya dapat dijadikan senjata ketika prmosi, salah satunya Jurusan

Sistem Informasi bidang keahlian manajemen bisnis dan manajemen Informatika. Sedangkan untuk jurusan IT yaitu bidang keahlian Grafis dan Multimedia.

c. Analisis deskriptif atribut Agama

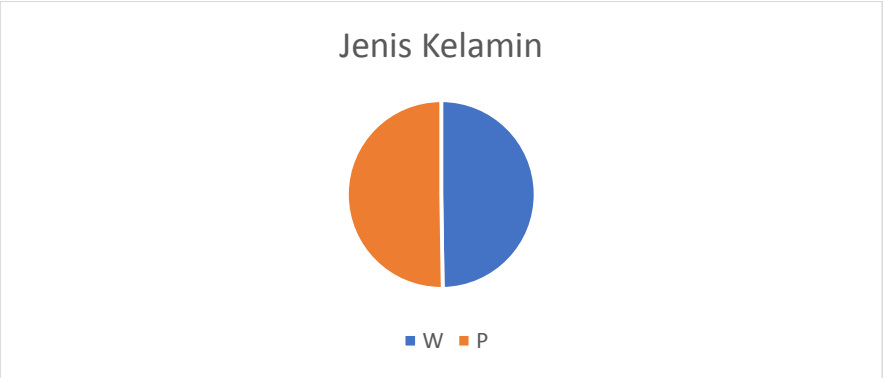


Gambar 4. 3
Analisis deskriptif atribut agama

Berdasarkan Gambar 4.3 dapat dilakukan analisis sebagai berikut:

1) Data agama yang paling banyak adalah Islam dengan jumlah data 1405.
2) Data agama yang paling sedikit adalah pilihan lainnya dengan jumlah data sebanyak 1 data.
3) Data agama Kristen berjumlah 955 data, Khatolik berjumlah 284, Budha 52, Hindu 9, dan tidak dicantumkan 163.
4) Namun apabila dilihat dari calon mahasiswa yang melakukan daftar ulang kembali lebih banyak berasal dari agama Kristen.

d. Analisis dekriptif Jenis Kelamin

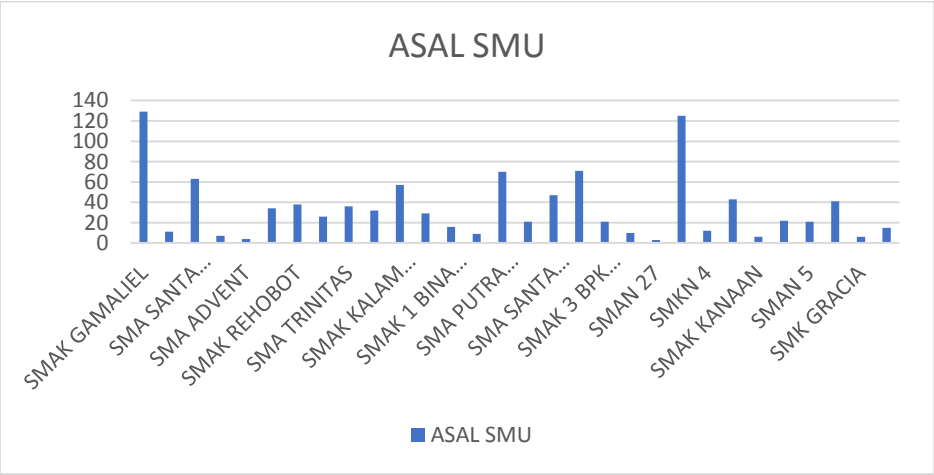


Gambar 4. 4
Analisis deskriptif jenis kelamin

Berdasarkan Gambar 4.4 dapat dilakukan analisis sebagai berikut:

1) Data jenis kelamin terbanyak yang mendaftar ke STMIK LIKMI adalah Wanita dengan jumlah data 1408.
2) Sedangkan Pria berjumlah 1394 data.
3) Dapat dilihat bahwa calon mahasiswa yang mendaftar ke Likmi cenderung seimbang walau lebih banyak wanita.

e. Analisis deskriptif atribut asal SMU

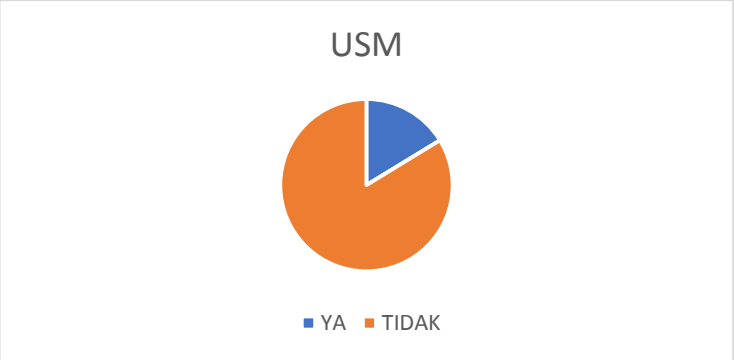


Gambar 4. 5
Analisis deskriptif asal SMU

Berdasarkan Gambar 4.5 dapat dilakukan analisis sebagai berikut:

1) Data asal SMU terbanyak diraih oleh SMAK Gamaliel dengan jumlah data 129.
2) Dari banyak SMU yang terdata ternyata hanya sekolah-sekolah yang sudah bermitra dengan STMIK LIKMI yang memiliki angka yang cukup tinggi, rata-rata sekolah yang berasal dari luar kota dan tidak pernah bermitra dengan STMIK LIKMI hanya memeiliki 1 pendaftar dan cenderung tidak melakukan daftar ulang.

f. Analisis deskriptif atribut USM (pilihan jalur USM atau NON-USM)

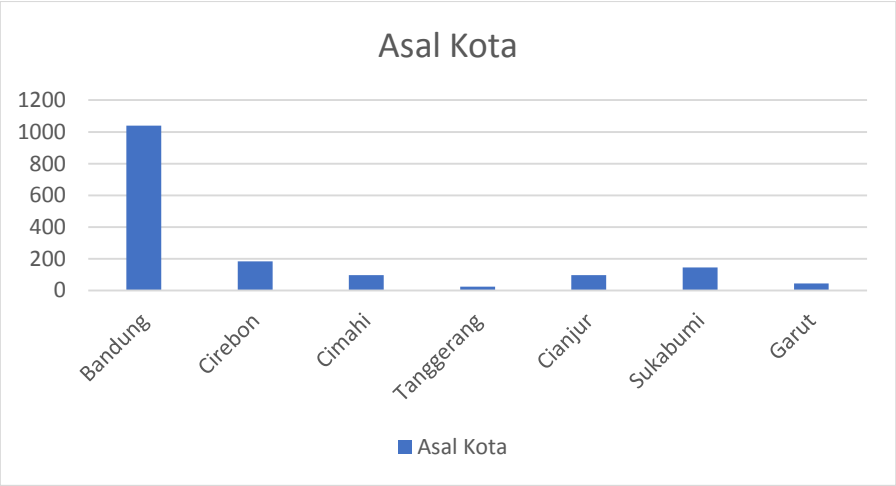


Gambar 4. 6
Analisis deskriptif atribut USM

Berdasarkan Gambar 4.6 dapat dilakukan analisis sebagai berikut:

1) Data pilihan jalur terbanyak adalah ketika jalur NON-USM (Tanpa Ujian Saringan Masuk) dengan jumlah data 2402.
2) Sedangkan untuk calon mahaiswa yang mengikuti USM (Ujian Saringan Masuk) adalah 467 data.
3) Dapat dilihat bahwwa Jalur Non-USM lebih banyak diikuti oleh calon mahasiswa ketimbang harus mengikuti ujian.

g. Analisis deskriptif Atribut Asal Kota



Gambar 4. 7
Analisis deskriptif atribut Asal kota

Berdasarkan Gambar 4.7 dapat dilakukan analisis sebagai berikut:

1) Asal Kota terbanyak berasal dari kota Bandung, Cirebon, Cimahi, Tanggerang, Cianjur, Sukabumi dan Garut..
2) Asal kota tertinggi adalah berasal dari Bandung dengan jumlah data 1034 data
3) Asal kota calon mahasiswa tersebar di seluruh Indonesia, namun dapat dilihat bahwa kota paling tinggi berada dekat dengan Bandung dan merupakan kota yang paling sering dikunjungi oleh STMIK LIKMI ketika melakukan pameran pendidikan di sekolah-sekolah yang sudah bermitra sejak lama.

h. Analisis deskripitif atribut beasiswa yang diajukan

Gambar 4. 8
Analisis deskriptif Beasiswa yang diajukan

Berdasarkan Gambar 4.8 dapat dilakukan analisis sebagai berikut:

1) Pilihan beasiswa yang diajukan calon mahasiswa paling banyak adalah pilihan 2 (beasiswa penuh) dengan jumlah data 1299.
2) Data beasiswa yang diajukan calon mahasiswa paling sedikit adalah pilihan 1 (Tidak mengajukan beasiswa) dengan jumlah data 62.
3) Data beasiswa yang diajukan calon mahasiswa pilihan 3 (Beasiswa Bebas Uang Pangkal) berjumlah 353 data sedangkan pilihan 4 (Beasiswa potongan uang pangkal) berjumlah 1155 data.

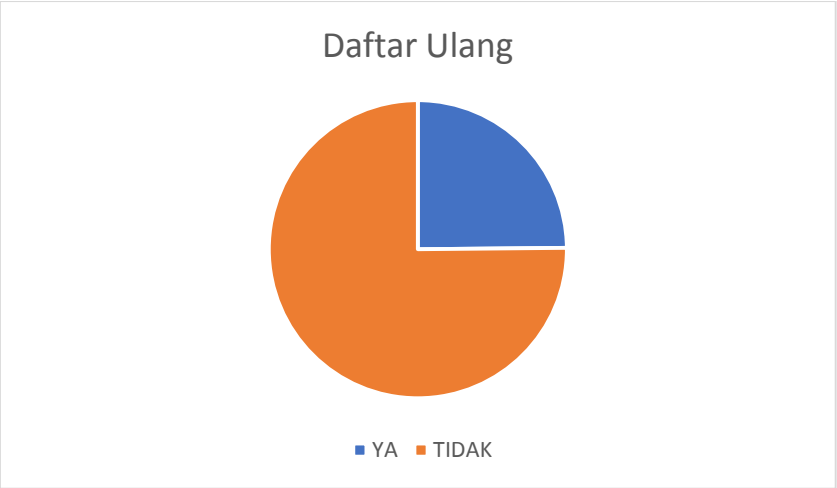
i. Analisis deskriptif atribut Hasil Beasiswa

Gambar 4. 9
Analisis deskriptif atribut Hasil Beasiswa

Berdasarkan Gambar 4.9 dapat dilakukan analisis sebagai berikut:

1) Hasil beasiswa yang paling banyak didapatkan adalah pilihan beasiswa 4 (Beasiswa Potongan Uang Pangkal) dengan jumlah data 1868 data. Menurut persentasenya lebih dari 70% calon mahasiswa mendapatkan beasiswa ini.
2) Hasil beasiswa yang paling sedikit didapatkan adalah pilihan beasiswa 0 (Tidak mengikuti program beasiswa) dengan jumlah data 2 data.
3) Dapat dilihat bahwa kebanyakan mahasiswa juga tidak masalah dengan beasiswa hanya potongan uang pangkal saja karena banyak yang melakukan daftar ulang walau hanya mendapatkan potongan uang pangkal saja.

j. Analaisis deskriptif atribut Daftar Ulang

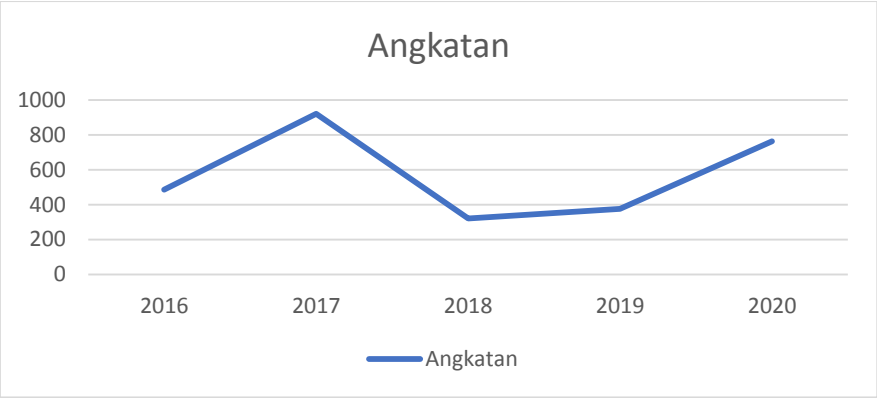


Gambar 4. 10
Analisis deskriptif atribut daftar ulang

Berdasarkan Gambar 4.10 dapat dilakukan analisis sebagai berikut:

1) Calon Mahasiswa yang melakukan daftar ulang setelah mendapatkan surat hasil penerimaan mahasiswa adalah 713 data.
2) Calon Mahasiswa yang batal mendaftar ulang setelah mendapatkan surat hasil penerimaan mahasiswa adalah 2156 data.
3) Dapat dilihat bahwa dari keseluruhan pendaftar hanya sekitar 30% yang menyelesaikan prosesnya hingga melakukan daftar ulang dan menjadi mahasiswa.

k. Analisis deskriptif atribut Angkatan



Gambar 4. 11
Analisis deskriptif atribut Angkatan

Berdasarkan Gambar 4.11 dapat dilakukan analisis sebagai berikut:

1) Pendaftaran yang dilakukan oleh calon mahasiswa paling tinggi terdapat pada tahun 2017 dengan jumlah 922 data tercatat
2) Pendaftaran yang dilakukan oleh calon mahasiswa paling rendah terdapat pada tahun 2018 dengan jumlah 321 data tercatat
3) Dapat dilihat bahwa jumlah pendaftar setiap tahunnya berubah-ubah naik dan turun. Namun apabila melihat jumlah mahasiswa yang ada, maka perbandinggannya sekitar 30% dari pendaftaran total telah menjadi mahasiswa

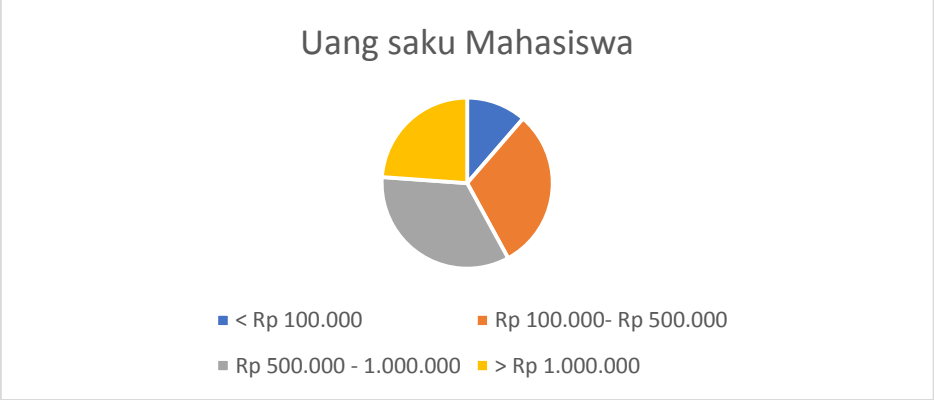
l. Analisis deskriptif atribut pendapatan orang tua

Gambar 4. 12
Analisis deskriptif atribut pendapatan orang tua

Berdasarkan Gambar 4.12 dapat dilakukan analisis sebagai berikut:

1) Pendapatan orang tua dari mahasiswa STMIK LIKMI paling tinggi berkisar antara Rp 3.000.000 – Rp 5.000.000 (sesuai dengan kisaran UMR di Kota Bandung) perbulannya dengan jumlah data 33 data dari 89 data.
2) Pendapatan orang tua dari mahasiswa STMIK LIKMI yang paling rendah berkisar antara kurang dari Rp 3.000.000 dengan jumlah yang cukup tipis dengan pendapatan di atas Rp 5.000.000 yaitu 27 data untuk pendapatan kurang dari Rp 3.000.000 dan 28 data untuk pendapatan lebih dari Rp 5.000.000
3) Dapat dilihat bahwa pendapatan orang tua dari mahasiswa STMIK LIKMI cenderung menengah dan ternyata yang memiliki pendapatan lebih dari UMR juga cukup banyak.

m. Analisis deskriptif atribut uang saku



Gambar 4. 13
Analisis Deskriptif Uang Saku Mahasiswa

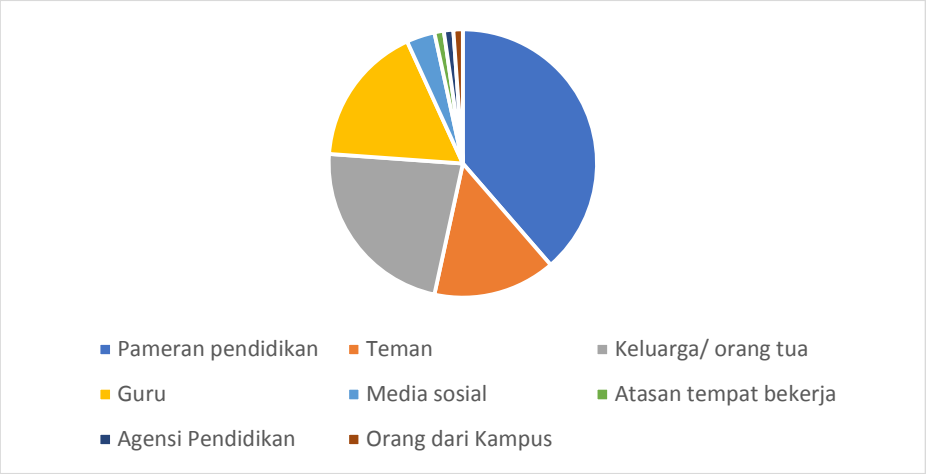
Berdasarkan Gambar 4.13 dapat dilakukan analisis sebagai berikut:

1) Uang saku dari Mahasiswa STMIK LIKMI rata-rata paling tinggi berkisar pada Rp 500.000 – Rp 1.000.000 perbulannya dengan jumlah data yaitu 30 data. Berbeda tipis dengan uang saku yang diatas Rp 1.000.000 perbulannya dengan jumlah data 27 data

2) Uang saku dari Mahasiswa STMIK LIKMI rata-rata paling rendah berkisar pada dibawah Rp 100.000 perbulannya dengan jumlah data yaitu 10 data

3) Dapat dilihat bahwa uang saku yang diterima mahasiswa berada menengah tinggi yaitu diatas Rp 500.000 perbulannya.

n. Analisis deskriptif atribut informasi kampus



Gambar 4. 14
Analisis deskriptif atribut Informasi kampus

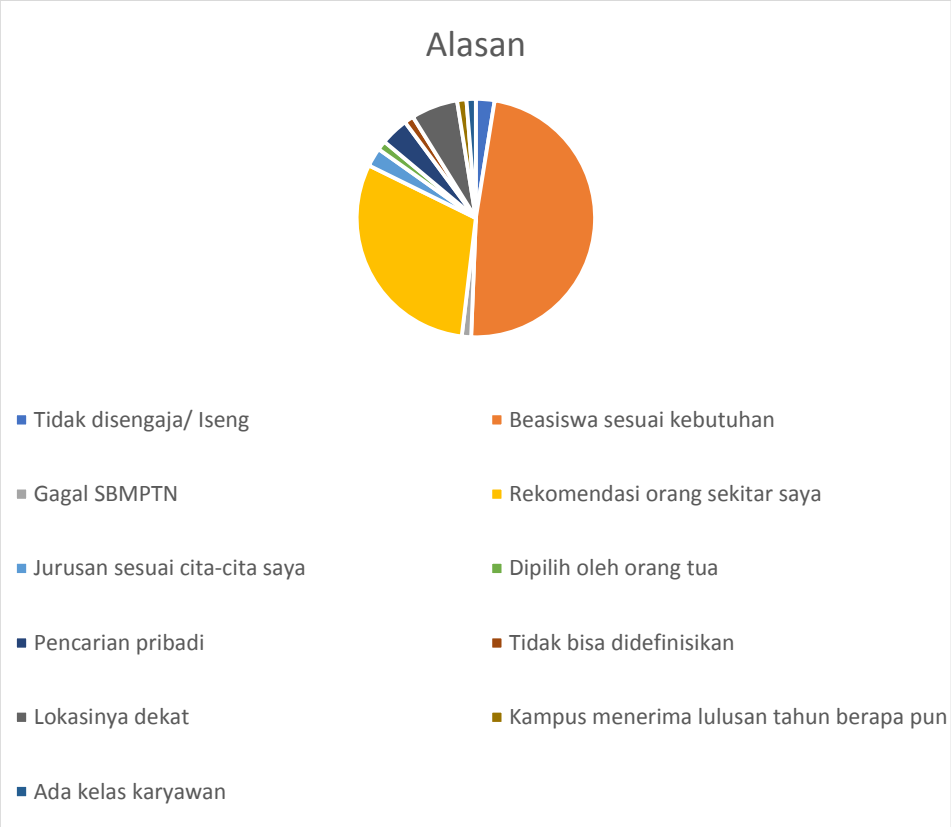
Berdasarkan Gambar 4.14 dapat dilakukan analisis sebagai berikut:

1) Mahasiswa STMIK LIKMI mengetahui kampus STMIK LIKMI paling banyak berasal dari Pameran pendidikan yaitu dengan jumlah data 33 dari 89 data.

2) Mahasiswa STMIK LIKMI mengetahui kampus STMIK LIKMI paling sedikit berasal dari Agensi pendidikan, atasan tempat bekerja, dan orang dari kampus yaitu dengan jumlah data 1 data.

3) Dapat dilihat bahwa hal yang paling berpengaruh untuk kampus dapat diketahui oleh calon mahasiswa lainnya adalah melalui pameran pendidikan, rekomendasi guru, teman dan orang tua atau keluarga.

o. Analisis deskriptif atribut Alasan



Gambar 4. 15
Analisis deskriptif atribut alasan

Berdasarkan Gambar 4.15 dapat dilakukan analisis sebagai berikut:

1) Alasan terbanyak Mahasiswa STMIK LIKMI memilih kampus STMIK LIKMI adalah Beasiswa yang sesuai kebutuhan dengan jumlah data 38 dari 89 data
2) Alasan yang paling jarang diberikan Mahasiswa STMIK LIKMI memilih kampus STMIK LIKMI adalah gagal SBMPTN, dipilihkan oleh orang tua, Tidak dapat didefinisikan, kampus yang bisa menerima dari tahun lulus berapa saja dan adanya kelas karyawan.
3) Dapat dilihat bahwa alasan yang cukup berpengaruh diantaranya adalah karena beasiswa yang sesuai, rekomendasi orang sekitar, jurusan dan program studi yang sesuai.

4.4. Data Transformation

Pada tahap ini data tersebut disesuaikan dengan format dari sistem WEKA sebagai alat untuk melakukan *Data Mining*. Data yang didapatkan berformat dbf sedangkan pengolahan dilakukan dalam format csv, maka perlu proses melakukan transformasi format data. Transformasi format data dilakukan menggunakan aplikasi bernama DBFViewer, pada aplikasi DBFViewer terdapat fitur untuk mengekstrak data ke dalam bentuk excel. Selain itu pada tahap ini dilakukan penyesuaian data dengan menggunakan *Microsoft excel* seperti melakukan sort dan filter.

Pada tahap ini juga dilakukan normalisasi data dari beberapa tabel, yaitu tabel data formulir dan data tabel Daftar Ulang, sedangkan data-data pendukung dianalisis terpisah. Pengkodean pada data yang memiliki format yang tidak sesuai seperti penggunaan angka untuk kode diubah menjadi huruf.

4.5. *Data Mining*

Pada tahap ini dilakukan klasifikasi menggunakan metode Naïve Bayes, metode ini dipilih karena sesuai dengan hasil yang ingin dicari. Pengukuran tingkat akurasi pada penelitian ini menggunakan Framework Weka ver. 3.8.4. *Class* yang menjadi acuan adalah yang melakukan daftar ulang dan tidak mendaftar ulang setelah mendapat hasil beasiswa atau surat hasil penerimaan mahasiswa. Kemungkian dari variable-variabel yang ada ditemukan dari korelasi antar variable. Perhitungan kemungkinan dengan metode Naïve Bayes addalah sebagai berikut:

$$P(A|B) = \frac{P(A)\pi_{i=1}^{d} P(B_i|A)}{P(B)}$$

Berdasarkan data yang ada dihitung peluang menggunakan rumus tersebut seperti berikut:

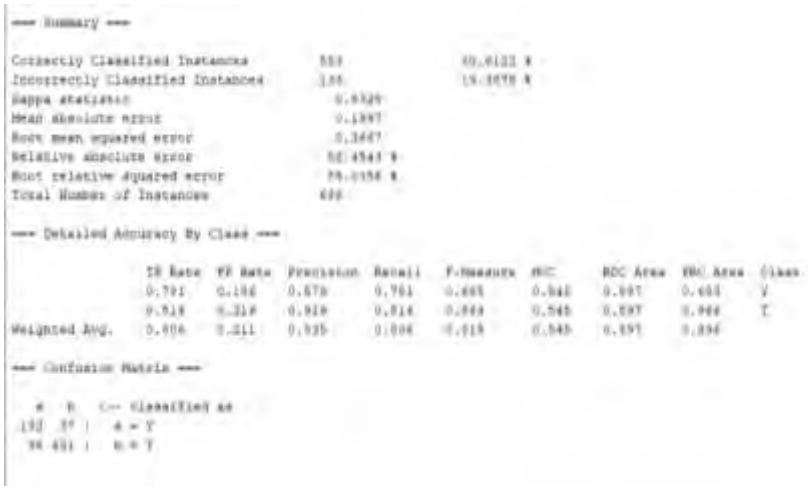
1. Perhitungan pada *class*, merupakan peluang antara yang melakukan setiap kelas terhadap banyak kelas data dan atribut data lainnya.
2. Perhitungan peluang bersyarat pada tiap variable terhadap *class.*

3. Perhitungan peluang untuk memberikan label pada data uji, apakah kalimat mengacu pada yang daftar ulang atau tidak daftar ulang

Setelah data CSV dimasukan ke dalam sistem WEKA, dilakukan klasifikasi dengan memilih *classifiy* dari *package Bayes* yaitu *Naïve Bayes*. *Test Option* menggunakan *Percentage Split* dikarenakan data pertahun yang banyak salah satunya adalah data tahun 2020 yaitu sekitar 26% dari keseluruhan data, maka akan dilakukan *Percentage Split* dengan 74% data latih dan 26% data uji.

Implementasi tahapan tersebut pada Weka dan memberikan hasil sebagai berikut:

1. Peluang Kelas DU



Gambar 4. 16
Classifier Summary

Pada Gambar 4.16 dapat dilihat bahwa nilai *error* yang diperoleh adalah 19.3878% yaitu sekitar 133 data dari 786 data uji dan data berhasilnya adalah 80.6122% yaitu sekitar 553 data dari 786 data uji. Setelah itu proses data model akan diuji cobakan dengan melakukan pengujian pada setiap atribut dengan atribut lainnya dan akan menghilangkan atribut yang sekiranya tidak dibutuhkan. Apabila dipisahkan perangkatan dengan *Test Option* yang sama maka tingkat akurasinya seperti pada tabel 4.7. Semakin tinggi akurasinya maka semakin terbentuk pola yang ada.

Tabel 4. 7
Classifier Summary Per Angkatan

Angkatan	Correctly Classified Instances	Incorrectly Classfied Instances:
2016	82.6446 %	17.3554 %
2017	90.566 %	9.434 %
2018	82.0513 %	17.9487 %
2019	73.0769 %	26.9231 %
2020	84.8485 %	15.1515 %

2. Model Data Hasil Klasifikasi Kelas DU

Nilai dari True Positive Rate yang diraih adalah 132 pola, sehingga kelas DU memiliki 132 pola untuk memperlihatkan karakteristik calon mahasiswa STMIK LIKMI. Hasil dari klasifikasi data uji memberikan margin data diantara -1 hingga 0 untuk data yang tidak tepat diprediksi dan 0 hingga 1 yang tepat diprediksi. Maka dapat dibaca bahwa nilai prediksi mendekati 0 merupakan peluang terbaik untuk Mahasiswa Tidak melakukan daftar ulang dan nilai akurasi mendekati 1 merupakan peluang terbaik untuk mahasiswa melakukan daftar ulang. Berikut model yang dihasilkan untuk kemungkinan paling besar Calon Mahasiswa melakukan daftar ulang:

Tabel 4. 8
Hasil Model Data Tahun 2016-2020

Pola (Gelombang-Program Studi-Asal Kota- Agama-Jenis-Asal SMU-USM-Beasiswa-Hasil Beasiswa)	Margin Prediction	Jumlah Data
Z - FTR - BANDUNG - B - P - SMAK 2 BPK PENABUR - T - D - D	0.940832	3
Z - FTR - BANDUNG - B - P - SMAK GAMALIEL - T - D - D	0.914862	3
Z - SBR - BANDUNG - B - P - SMAK GAMALIEL - T - D - D	0.669396	7
Z - SRR - BANDUNG - B - P - SMAK GAMALIEL - T - D – D	0.798064	5
Z - SMR - BANDUNG - B - P - SMAK GAMALIEL - T - D - D	0.486789	4

Setelah melakukan pengolahan lebih lanjut dengan mencoba kombinasi atribut lainnya, ternyata tidak semua atribut mempengaruhi Daftar Ulang. Salah satunya adalah Beasiswa dan Hasil beasiswa, setelah dicoba mengolah atribut Beasiswa, Hasil Beasiswa dan Daftar Ulang ternyata tidak ditemukan pola yang memiliki peluang calon mahasiswa melakukan daftar ulang dilihat dari *True Positive Rate* yang diraih adalah 0.

Berikut dilakukan proses kombinasi atribut dengan menghilangkan beberapa atribut untuk meningkatkan tingkat akurasi dan menemukan pola lainnya berdasarkan *Information Gain.* Penghapusan akan dimulai mengikuti urutan terkecil pada *Information Gain* yang telah dihitung sebelumnya pada tahap *Data Selection* seperti berikut:

1. Menghilangkan Atribut Jenis Kelamin

   Atribut yang digunakan: Gelombang, Agama, USM, Asal Kota, Asal SMU, Pilihan Utama, Beasiswa, Hasil Beasiswa

   Correctly Classified Instances: 81.4869 %

   Incorrectly Classfied Instances: 18.5131 %

   Hasil yang dikeluarkan oleh sistem, akurasi menjadi lebih besar maka Atribut Jenis kelamin akan dihapus.

2. Menghilangkan Atribut Hasil Beasiswa

   Atribut yang digunakan: Gelombang, Agama, USM, Asal Kota, Asal SMU, Pilihan Utama, Beasiswa

   Correctly Classified Instances: 81.637 %

   Incorrectly Classfied Instances: 18.3672 %

   Hasil yang dikeluarkan oleh sistem, akurasi menjadi lebih besar maka Atribut Hasil Beasiswa akan dihapus karena tidak mempengaruhi.

3. Menghilangkan Atribut Pilihan Utama

   Atribut yang digunakan: Gelombang, Agama, USM, Asal Kota, Asal SMU, Beasiswa

   Correctly Classified Instances: 82.3615 %

   Incorrectly Classfied Instances: 17.6385 %

   Hasil yang dikeluarkan oleh sistem, akurasi menjadi lebih besar maka Atribut Pilihan Utama akan dihapus.

4. Menghilangkan Atribut Beasiswa

Atribut yang digunakan:	Gelombang, Agama, USM, Asal Kota, Asal SMU, Pilihan Utama
Correctly Classified Instances:	82.9446 %
Incorrectly Classfied Instances:	17.0554 %

Hasil yang dikeluarkan oleh sistem, akurasi menjadi lebih besar maka Atribut Beasiswa akan dihapus.

5. Menghilangkan Atribut Asal USM

Atribut yang digunakan:	Gelombang, Agama, Pilihan Utama, Asal Sekolah, Asal Kota
Correctly Classified Instances:	81.6327 %
Incorrectly Classfied Instances:	18.3673 %

Hasil yang dikeluarkan oleh sistem, akurasi menjadi lebih kecil dengan sebelumnya maka Atribut USM akan tetap digunakan.

6. Menghilangkan Atribut Agama

Atribut yang digunakan:	Angkatan, Gelombang, Asal SMU, Pilihan Utama, Asal Kota
Correctly Classified Instances:	82.3615 %
Incorrectly Classfied Instances:	17.6385 %

Hasil yang dikeluarkan oleh sistem, akurasi menjadi lebih kecil dengan sebelumnya maka Atribut Agama akan tetap digunakan.

7. Menghilangkan Atribut Gelombang

Atribut yang digunakan:	Agama, Asal Kota, Asal SMU, Pilihan Utama
Correctly Classified Instances:	80.1749 %

Incorrectly Classfied Instances: 19.8251 %

Hasil yang dikeluarkan oleh sistem menjadi lebih kecil maka Atribut Gelombang tetap akan digunakan.

8. Menghilangkan Atribut Asal Kota

Atribut yang digunakan: Gelombang, Agama, Asal SMU, Pilihan Utama

Correctly Classified Instances: 82.7988 %

Incorrectly Classfied Instances: 17.2012 %

Hasil yang dikeluarkan oleh sistem menjadi lebih kecil maka Atribut Asal Kota akan tetap digunakan untuk meningkatkan akurasi.

9. Menghilangkan Atribut Asal SMU

Atribut yang digunakan: Angkatan, Gelombang, Agama, Pilihan Utama

Correctly Classified Instances: 82.2157 %

Incorrectly Classfied Instances: 17.7843 %

Hasil yang dikeluarkan oleh sistem menjadi lebih kecil maka Atribut Asal SMU tetap akan digunakan.

Berdasarkan perhitungan melalui fitur *Classifiy* dengan mengikuti urutan pada *Information Gain* maka atribut yang memberikan pola paling akurat untuk menentukan seorang calon mahasiswa melakukan daftar ulang adalah Gelombang, Agama, USM, Asal Kota, dan Asal SMU. Setelah dicoba kembali prosesnya maka dipilih beberapa model pola data yang memiliki data cukup tinggi seperti berikut:

Tabel 4. 9
Karakteristik Calon Mahasiswa 2016-2020

Pola (Gelombang- Asal Kota-Agama- Asal SMU-USM)	**Margin Prediction**	**Jumlah Data**
Z - BANDUNG - B - SMAK GAMALIEL - T	0.334553	93
Z - BANDUNG - B - SMAK KALAM KUDUS - T	0.374817	34
Z - BANDUNG - B - SMAK 3 BINA BAKTI - T	0.309332	23

Pola (Gelombang- Asal Kota-Agama- Asal SMU-USM)	Margin Prediction	Jumlah Data
Z - BANDUNG - C - SMA SANTA MARIA 1 - T	0.05598	21
Z - BANDUNG - B - SMAK 2 BPK PENABUR - T	0.490428	20
Z - CIMAHI - B - SMA SANTA MARIA 3 - T	0.560147	11
Z - BANDUNG - C - SMAK GAMALIEL - T	0.148736	9
Z - BANDUNG - B - SMAK BPPK - T	0.309332	7
Z - BANDUNG - D - SMA TALENTA - T	0.050531	6
B - BANDUNG - B - SMAK GAMALIEL - T	0.938474	6
Z - BANDUNG - B - SMAK SANTO PAULUS - T	0.639626	5
E - BANDUNG - B - SMAK BPPK - Y	0.988182	4
G - BANDUNG - B - SMAK KALAM KUDUS - Y	0.988211	4
B - BANDUNG - B - SMAK REHOBOTH - Y	0.991687	4
A - BANDUNG - B - SMAK REHOBOTH - Y	0.983312	4
E - BANDUNG - C - SMA SANTA MARIA 1 - Y	0.980054	3
B - KARAWANG - B - SMA YOS SUDARSO - T	0.483157	3
C - BANDUNG - B - SMAK GAMALIEL - Y	0.990796	3
D - BANDUNG - B - SMAK KALAM KUDUS - Y	0.992534	3
A - BANDUNG - B - SMAK KALAM KUDUS - Y	0.987358	3

Pada Tabel 4.9 dapat dilihat bahwa setiap tahunnya diantara 2016 hingga 2020 karakteristik yang ditemukan memiliki beberapa kemiripan yaitu sekolah dengan peluang tertinggi dan jumlah tertinggi adalah sekolah-sekolah yang sering dikunjungi untuk pameran pendidikan dan gelombang awal lebih efektif untuk membuat seseorang melakukan daftar ulang. Maka, kinerja pemasaran harus lebih difokuskan pada gelombang-gelombang awal pada sekolah-sekolah yang juga sudah lama bermitra dengan STMIK LIKMI.

Penelitian juga dilakukan dengan mengolah data per tahunnya untuk melihat efektivitas setiap tahunnya. Pada Tabel 4.10 hingga Tabel 4.14 merupakan beberapa pola yang didapatkan dengan jumlah prediksi benar dan jumlah data tertinggi sehingga dapat dilihat bahwa setiap tahunnya calon mahasiswa STMIK LIKMI memiliki pola-pola yang hampir mirip. Namun dapat dilihat bahwa adanya penurunan jumlah efektivitas pada

sekolah-sekolah yang bermitra dengan STMIK LIKMI dan munculnya sekolah-sekolah baru yang melakukan pendaftaran. Berdasarkan hal tersebut dapat dilihat bahwa pemasaran dilakukan semakin menyebar luas dan menarik banyak pendaftar, hal tersebut juga terjadi akibat penambahan sekolah-sekolah yang dikunjungi ketika pameran pendidikan.

Tabel 4. 10
Karakteristik Calon Mahasiswa Tahun 2016

Pola (Gelombang- Asal Kota- Agama- Asal SMU-USM)	Margin Prediction	Jumlah Data
Z - BANDUNG - B - SMAK GAMALIEL - T	0.576261	28
Z - BANDUNG - B - SMAK KALAM KUDUS - T	0.445048	9
Z - BANDUNG - B - SMAK 2 BPK PENABUR - T	0.238805	5
Z - BANDUNG - C - SMA SANTA MARIA 1 - T	0.256237	5
C - BANDUNG - B - SMAK REHOBOTH - Y	0.973759	3
Z - BANDUNG - B - SMAK YAHYA - T	0.445048	2
B - BANDUNG - B - SMAK REHOBOTH - Y	0.987279	2

Tabel 4.10 memperlihatkan pola karakteristik dari calon mahasiswa pada Tahun 2016 dengan jumlah data tertinggi. Apabila dilihat dari tingkat akurasinya maka semakin tinggi semakin tepat prediksinya dan memberi kemungkinan lebih tinggi untuk calon mahasiswa dengan karakteristik yang sama selanjutnya akan mendaftar ulang. Pada Tahun 2016 dapat dilihat bahwa SMAK Rehobot diprediksi cukup tinggi dan SMAK Gamaliel dengan jumlah calon mahasiswa yang tinggi diprediksi dapat cenderung ke daftar ulang dan dapat juga memiliki kecenderungan tidak mendaftar ulang. Kesimpulan dari pola yang ditemukan pada tahun 2016, karakteristik calon mahasiswa yang benar diprediksi daftar ulang adalah dengan pola melakukan pendaftaran pada gelombang awal, berasal dari kota Bandung, bergama Kristen, berasal dari sekolah Gamaliel dan rehobot, serta tidak mengikuti ujian.

Tabel 4. 11
Karakteristik Calon Mahasiswa Tahun 2017

Pola (Gelombang- Asal Kota- Agama- Asal SMU-USM)	Margin Prediction	Jumlah Data
Z -BANDUNG -B -SMAK GAMALIEL -T	0.506268	29
Z -BANDUNG -B -SMAK 3 BINA BAKTI -T	0.623138	9
Z -BANDUNG -B -SMAK 1 BINA BAKTI -T	0.235216	3
Z -BANDUNG -B -SMAK REHOBOTH -T	0.235216	2

Pada Tabel 4.11 merupakan pola yang ditemukan pada tahun 2017, kesimpulan dari pola yang didapatkan adalah pada tahun 2017 karakteristik calon mahasiswa yang benar diprediksi daftar ulang terdiri dari pola melakukan pendaftaran pada gelombang awal, berasal dari kota Bandung, bergama Kristen, berasal dari sekolah Gamaliel dan Bina Bakti, serta tidak mengikuti ujian.

Tabel 4. 12
Karakteristik Calon Mahasiswa Tahun 2018

Pola (Gelombang- Asal Kota- Agama- Asal SMU-USM)	Margin Prediction	Jumlah Data
Z - BANDUNG - B - SMAK GAMALIEL - T	0.232752	15
Z - BANDUNG - B - SMAK KALAM KUDUS - T	0.413503	7
Z - BANDUNG - B - SMAK REHOBOTH - T	0.092987	4
Z - BANDUNG - B - SMAK 2 BPK PENABUR - T	0.656367	4
Z - CIMAHI - C - SMA SANTA MARIA 3 - T	0.577009	4
Z - BANDUNG - B - SMAK PELITA BANGSA - T	0.092987	3

Pada Tabel 4.12 merupakan pola yang ditemukan pada tahun 2018, kesimpulan dari pola yang didapatkan adalah pada tahun 2018 karakteristik calon mahasiswa yang benar diprediksi daftar ulang terdiri dari pola melakukan pendaftaran pada gelombang awal, berasal dari kota Bandung, bergama Kristen, berasal dari sekolah SMAK 2 BPK Penabur dan SMAK Kalam Kudus, serta tidak mengikuti ujian. Kemungkinan besar pada tahun 2018 terjadi peningkatan kinerja pada gelombang awal melihat pola-pola yang dihasilkan paling akurat ada pada gelombang awal.

Tabel 4. 13
Karakteristik Calon Mahasiswa Tahun 2019

Pola (Gelombang- Asal Kota- Agama- Asal SMU-USM)	Margin Prediction	Jumlah Data
Z - BANDUNG - B - SMAK GAMALIEL - T	0.622095	12
Z - BANDUNG - B - SMAK REHOBOTH - T	0.708165	6
Z - BANDUNG - B - SMAK BPPK - T	0.322278	4
Z - BANDUNG - B - SMA TALENTA - T	0.490653	4
Z - BANDUNG - B - SMAK 3 BPK PENABUR - T	0.322278	3

Pada Tabel 4.13 merupakan pola yang ditemukan pada tahun 2019, kesimpulan dari pola yang didapatkan adalah pada tahun 2019 karakteristik calon mahasiswa yang benar diprediksi daftar ulang terdiri dari pola melakukan pendaftaran pada gelombang awal, berasal dari kota Bandung, bergama Kristen, berasal dari sekolah SMAK Rehobot dan SMAK Gamaliel, serta tidak mengikuti ujian.

Tabel 4. 14
Karakteristik Calon Mahasiswa Tahun 2020

Pola (Gelombang- Asal Kota- Agama- Asal SMU-USM)	**Margin Prediction**	**Jumlah Data**
B - BANDUNG - B - SMK GRACIA - T	0.867687	2
B - BANDUNG BARAT - B - SMKN 4 BANDUNG - T	0.742216	1
B - BANDUNG - B - SMA SANTA MARIA 2 - T	0.617935	1
B - BANDUNG - B - SMAK 3 BPK PENABUR - T	0.602628	1
B - BANDUNG - C - SMAK GAMALIEL - T	0.582123	1
A - BANDUNG - B - SMAK KALAM KUDUS - T	0.511665	1
B - BANDUNG - B - SMA TALENTA - T	0.239178	1

Pada Tabel 4.14 merupakan pola yang ditemukan pada tahun 2020, kesimpulan dari pola yang didapatkan adalah pada tahun 2020 karakteristik calon mahasiswa yang benar diprediksi daftar ulang terdiri dari pola melakukan pendaftaran pada gelombang awal, berasal dari kota Bandung, bergama Kristen, berasal dari sekolah SMK Gracia dan SMKN 4 Bandung, serta tidak mengikuti ujian. Pada tahun 2020 dapat dilihat bahwa munculnya sekolah-sekolah berbeda dari tahun sebelumnya, walau dengan jumlah yang kecil peningkatan kinerja pemasaran pada tahun 2020 cukup meningkat dengan memperbanyak kunjungan sekolah-sekolah atau secara tidak langsung dapat diartikan muncul peluang-peluang lainnya.

Berdasarkan data tersebut dapat dilihat bahwa atribut beasiswa yang diajukan, hasil beasiswa yang diberikan, program studi dan Jenis kelamin lebih berpengaruh untuk menarik banyak pendaftar mengisi formulir ketika dilakukan pemasaran. Namun untuk mempertahankan calon mahasiswa untuk tetap melakukan daftar ulang adalah atribut Gelombang, Asal Kota, Agama, Asal SMU, dan USM.

4.6. Data Evaluation/ Interpretasi

Berdasarkan *confusion matrix* data tersebut didapatkan hasil dari sistem seperti terdapat pada tabel 4.15.

Tabel 4. 15
Confusion Matrix

Correct Classification	**Classified as**	
	Y	**T**
Y	**TP** 132	**FP** 37
T	**FN** 96	**TN** 421

Berdasarkan Tabel 4.15 dijelaskan bahwa:

1. *True Positive Rate* (hasil prediksi yang nilainya sama-sama benar dengan sebenarnya) sebanyak 136 data
2. *False Negative Rate* (hasil prediksi salah, tetapi hasil yang sebenarnya bernilai Benar) sebanyak 33 data
3. *True Negative Rate* (hasil prediksi salah, tetapi hasil yang sebenarnya bernilai salah) sebanyak 436 data
4. *False Positive Rate* (hasil yang benar, tetapi hasil yang sebenarnya salah) sebanyak 81 data

Berikut merupakan perhitungan detail akurasi berdasarkan kelas DU yang dilakukan oleh sistem:

Tabel 4. 16
Evaluasi Perhitungan Sistem

Perhitungan Akurasi	**Rumus**	**Hasil**	**Keterangan**
Akurasi	$\frac{TP + TN}{TP + TN + FN + FP}$	0.806	Nilai akurasi dari klasifikasi cukup tinggi sekitar 80 % akurat
Eror Rate (ER)	$\frac{FN + FP}{TP + TN + FN + FP}$	0.194	Nilai eror cukup rendah karena hampir mendekati nilai 0
False Positive Rate	$\frac{FP}{TN + FP}$	0.080	Nilai prediksi dari kelas yang tidak terklasifikasi dengan benar 0.080 yang artinya sangat baik karena hampir mendekati nilai 0
Sensitivity (recall)	$\frac{TP}{TP + FN}$	0.579	Nilai prediksi dari kelas yang terklasifikasi dengan benar cukup bagus karena hampir menyentuh angka 1

Perhitungan Akurasi	Rumus	Hasil	Keterangan
Specificity (True Negative Rate)	$\frac{TN}{TN + FP}$	0.919	Hasil specificity sangat tinggi karena hampir menyentuh angka 1.
Precision (Positive Predictive Value)	$\frac{TP}{TP + FP}$	0.781	Nilai precision atau nilai dari jumlah kelas yang terklasifikasi dengan benar sangat tinggi hampir mendekati 1

Berdasarkan Tabel 4.16 tersebut memberikan kesimpulan bahwa tingkat akurasi dari perhitungan yang dilakukan oleh sistem sudah sangat baik karena hampir semuanya menunjukan hasil yang mendekati angka sempurna. Sehingga hasil klasifikasi dapat dipercaya.

4.7. Hasil Rekomendasi

Hasil penelitian ini diharapkan dapat menjadi beberapa pertimbangan untuk memperbaiki strategi marketing dan proses penerimaan mahasiswa baru. Berikut merupakan beberapa hasil dan rekomendasi dari penelitian:

1. Karakteristik Calon Mahasiswa
    a. Kelas Ekonomi Calon Mahasiswa yang mendaftar hingga berrhasil menjadi mahasiswa berkisar di kelas Sedang menengah dikarenakan pendapatan orang tua Mahasiswa STMIK LIKMI berkisar Rp 3.000.000 – Rp 5.000.000 (UMR). Sehingga ketika dalam menentukan target pasar untuk melakukan promosi dapat memprioritaskan yang kelas menengah kebawah.
    b. Sekolah yang memberikan keuntungan paling tinggi adalah SMAK Gamaliel Bandung dan beberapa sekolah mitra, sedangkan sekolah yang tidak begitu dikenal cenderung memiliki peluang yang lebih kecil. Setiap Sekolah memiliki karakternya masing-masing yang mempengaruhi pemilihan program studi.
    c. Beasiswa menjadi salah satu daya tarik untuk mempengaruhi seseorang mengisi formulir, namun tidak terlalu mempengaruhi proses Daftar Ulang. Hal ini dapat dilihat bahwa beberapa calon mahasiswa yang sudah

mendapatkan beasiswa penuh sekalipun ada peluang untuk batal menjadi mahasiswa. Begitu juga sebaliknya beberapa calon mahasiswa yang mendapatkan beasiswa yang tidak sesuai dengan yang diajukan tetap memiliki peluang yang sama untuk menjadi mahasiswa.

d. Calon Mahasiswa yang berasal dari luar kota cenderung lebih kecil kemungkinannya untuk melakukan pendaftaran ulang, hal ini dapat dilihat dari lingkungan asal sekolah. Asal sekolah dapat memberikan pengetahuan mengenai latar belakang ekonomi dari calon mahasiswa, bagi calon mahasiswa yang kondisi ekonominya tidak sesuai dengan margin pemasaran STMIK LIKMI (menengah keatas) kesulitan untuk melakukan pembayaran daftar ulang walau mendapatkan beasiswa.

e. Karakteristik calon mahasiswa STMIK LIKMI yang paling berpeluang tinggi untuk menjadi mahasiswa adalah yang berasal dari sekolah mitra (SMAK Gamaliel, SMAK Rehobot, SMAK Kalam Kudus, SMA Santa Maria, dan SMAK Bina Bakti) di daerah bandung pada gelombang jalur undangan dan mayoritas beragama kristen. Lingkungan yang ada tercipta dengan adanya bantuan rekomendasi dari mahasiswa STMIK LIKMI yang sudah lulus maupun yang masih menempuh studi, sehingga banyak calon mahasiswa yang berasal dari keluarga mahasiswa STMIK LIKMI.

2. Beasiswa

Salah satu yang menjadi keunggulan STMIK LIKMI dibanding kampus lain adalah Beasiswa. Mahasiswa STMIK LIKMI yang berada di kelas ekonomi menengah keatas tetap merasa bersyukur ketika mendapatkan beasiswa dari STMIK LIKMI. Beberapa calon mahasiswa yang gagal mendapatkan beasiswa penuh diberikan beasiswa yang lebih kecil dan cenderung tidak melakukan daftar ulang. Sehingga peranan beasiswa juga menarik banyak calon mahasiswa namun untuk mencapai proses daftar ulang ditentukan oleh hasil yang didapatkan dan alasan pendukung lainnya seperti kualitas

kampus. Berdasarkan karakteristik yang disimpulkan, beberapa beasiswa yang disarankan adalah sebagai berikut:

a. Beasiswa Keluarga (Family Scholarship), beasiswa yang diberikan untuk calon mahasiswa yang memiliki keluarga di STMIK LIKMI. Beasiswa ini juga dapat membantu STMIK LIKMI untuk membuat mahasiswanya merekomendasikan STMIK LIKMI kepada keluarga atau kerabat.
b. Beasiswa Potongan Uang Pangkal, beasiswa yang diberikan untuk membantu meringankan biaya studi dan menyasar calon mahasiswa yang sesuai dengan margin ekonomi.
c. Beasiswa Fasilitas (Beasiswa dengan fasilitas asrama), beasiswa yang diberikan untuk menunjang mahasiswa yang berasal dari luar kota untuk memberikan rasa percaya bagi orang tua yang takut atau kebingungan ketika melepaskan anaknya berkuliah ke luar kota atau bagi mahasiswa yang membutuhkan fasilitas tempat tinggal.
d. Beasiswa Prestasi, beasiswa yang diberikan untuk menjaring mahasiswa berprestasi agar dapat membantu prestasi Kampus sehingga meningkatkan citra kualitas pendidikan.

3. Strategi Pemasaran

Saat ini STMIK LIKMI cukup mampu menarik perhatian banyak kalangan dari yang mudah dijumpai hingga hanya dijumpai melalui media sosial. Strategi pemasaran yang dapat diberikan adalah selain mempertahankan senjata promosi menggunakan beasiswa sebagai tarik, STMIK LIKMI dapat meningkatkan mutu, layanan, dan sarana-prasarana untuk meningkatkan *Loyality* dari Mahasiswa. Sehingga citra kampus dan promosi secara verbal dapat tersampaikan melalui rekomendasi Mahasiswa maupun keluarga Mahasiswa kepada orang-orang disekitarnya. Terciptanya lingkungan/ gaya hidup dari mahasiswa-mahasiswa STMIK LIKMI (prestasi, tingkah laku, dll) yang lebih baik juga akan memberikan

citra yang baik pada kampus. Beberapa langkah nyata yang dapat digunakan adalah sebagai berikut:

a. Membangun asrama/ tempat tinggal sementara

Hal ini dapat membantu kesulitan orang tua calon mahasiswa yang berasal dari luar kota untuk mencari tempat tinggal yang aman dan nyaman bagi putra/ putrinya. Beasiswa yang diberikan juga dapat diversifikasi kembali jenisnya dengan mempromosiokan adanya asrama mahasiswa.

b. Smart Campus

STMIK LIKMI yang sudah memiliki citra sebagai *Digital Campus* sejak 1985, citra kampus dapat ditingkatkan lagi dengan banyaknya kemajuan teknologi yang ada. Salah satu konsep yang dapat digunakan adalah *Smart Campus,* yaitu dengan memanfaatkan teknologi seperti cloud computing, sensor pada bangunan (*smart window/ door*), sistem administrasi kemahasiswaan online, sistem absensi *face recognition* dan teknologi lainnya.

c. Sistem Akademik yang terpercaya

Kampus yang memiliki citra pendidikan yang baik tentunya akan mempermudah sistem pemasaran. Citra pendidikan tersebut dapat terus ditingkatkan, beberapa cara yang dapat disarankan adalah sebagai berikut:

1) Meningkatkan model pendidikan kampus menjadi Institut atau universitas, agar dapat menjaring banyak calon mahasiswa ke berbagai bidang dan jurusan.
2) Mendukung mahasiswa berprestasi di tingkat kota/ nasional
3) Melakukan pelatihan bagia setiap dosen atau tim pengajar agar memberikan pengajaran yang tidak hanya dapat menyampaikan materi namun memberikan kesan bagi mahasiswa
4) Berinvestasi pada penelitian dan pengembangan materi pembelajaran disetiap mata kuliah agar meningkatkan standart pendidikan yang sesuai dengan perkembangan zaman

BAB V

KESIMPULAN

5.1. Kesimpulan

Kesimpulan dari penelitian ini adalah sebagai berikut:

1. Beberapa Calon Mahasiswa yang saat ini sudah menjadi mahasiswa diberikan Survei melalui kuisioner, memberikan hasil bahwa mahasiswa STMIK LIKMI merupakan mahasiswa yang berada dikalangan ekonomi kelas sedang menengah. Mahasiswa STMIK LIKMI lebih banyak mengetahui tentang kampus melalui pameran pendidikan dan rekomendasi orang terdekat. Sedangkan alasan mereka memutuskan untuk memilih STMIK LIKMI sebagai kampus mereka adalah karena pilihan beasiswa yang sesuai kebutuhan mereka dan adanya dorongan melalui rekomendasi orang terdekat mereka.
2. Karakteristik calon mahasiswa untuk melakukan daftar ulang dipengaruhi oleh atribut Gelombang, Asal Kota, Asal SMU, Agama dan USM.
3. Strategi yang dapat diterapkan diantaranya:
    a. Memasang promosi lebih tinggi untuk daerah yang berada di sekitar Bandung dan cukup dekat dengan Bandung.
    b. Memberikan Beasiswa sebagai senjata promosi untuk menarik banyak orang untuk mendaftar. Beberapa beasiswa yang dapat digunakan adalah beasiswa keluarga, beasiswa prestasi, beasiswa fasilitas, dan beasiswa potongan uang pangkal.
    c. Mengembangkan citra *Digital Campus* menjadi *Smart Campus* dengan menggunakan teknologi yang ada.
    d. Menyediakan dan mengembangkan fasilitas-fasilitas kebutuhan dasar mahasiswa seperti: Asrama mahasiswa, *Food Court*, Gedung, dll untuk membangun citra yang baik.
    e. Membentuk sistem akademik yang semakin dipercaya oleh masyarakat sehingga membentuk citra yang baik dari mulut ke mulut. Beberapa langkah

yang dapat dilakukan seperti meningkatkan model kampus menjadi institut atau universitas, mendukung mahasiswa berprestasi di jenjang kota/ nasional, dan mengembangkan materi pembelajaran sesuai dengan perkembangan zaman.

5.2. Saran

Penelitian ini memberikan beberapa saran bagi perusahaan untuk mengembangkan bisnisnya terutama pada bagian pemasaran agar lebih efektif dan efisien. Adapun saran yang diberikan melalui penelitian ini bagi peneliti selanjutnya agar dapat mengembangkan penelitian lebih sempurna lagi adalah sebagai berikut:

1. Melakukan pengolahan data dengan metode selain Naïve Bayes untuk meningkatkan tingkat akurasi, seperti menggunakan metode C-45
2. Melakukan penelitian lebih lanjut pada karakteristik calon mahasiswa yang berpeluang sangat kecil untuk daftar ulang agar dapat melihat strategi pemasaran dari sudut pandang lain
3. Menambahkan atribut-atribut lainnya untuk menambah pengetahuan mengenai karakteristik calon mahasiswa lainnya seperti atribut sumber mengetahui STMIK LIKMI.

DAFTAR PUSTAKA

Arndt, Dirk & Wendy Gersten. 2001. "*Data Management in Analytical Customer Relationship Management*". Germany.

Buttle, Francis and Stan Maklan. 2015. "*Customer Relationship Management: Concepts and technologies".* Butterworth-Heinemann, an imprint of Elsevier. London.

Bygstad, Bendik dan Wanda Presthus. "*Social Media as CRM? How Two Scandinavian Airline Companies Used Facebook during the "Ash Crisis" in 2010*". Scandinavian.

Dewi, Sari. 2016. "*Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan*". AMIK BSI. Pontianak.

Fayyad, Usama, Gregory Piatetsky-Shapiro, dan Padhraic Smyth. 1996. "*From Data Mining to Knowledge Discovery in Databases*". American Association for Artificial Intelligence.

Gunawan, Agus dkk. 2013. "*Analytical Customer Relationship Management untuk Restoran*". Universitas Khatolik Parahyangan. Bandung.

Hananto, Valentinus Robu, dkk. 2017. "*Perancangan Analytical CRM untuk Mendukung Segmentasi Pelanggan di Institusi Pendidikan*". Institut Bisnis dan Informatika Stikom. Surabaya

Kahara, Stephen Wanjau, dkk. 2016. "*Data Mining Model for Predicting Student Enrolment in STEM Courses in Higher Education Institutions*". Murang'a University of Technology Murang'a. Kenya.

Khasana, Afidatun. 2015. "*Pemasaran Jasa Pendidikan Sebagai Strategi Peningkatan Mutu Di Sd Alam Baturraden*". UIN Sunan Kalijaga Yogyakarta.

Mulyana, Mumuh. 2019. "*Segmentasi Pasar, Penargetan Pasar dan Pemosisian*". Universitas Terbuka.

Nofriansyah, Dicky. 2016. "*Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi)*". STMIK Triguna Dharma. Medan.

Oliveira, Vera Lucia Migueis. 2012. "*Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques*". Universidade do Porto. Prancis

Oluseye, Ogunnaike Olaleke, dkk. 2014. "*Customer Relationship Management Approach and Student Satisfaction in Higher Education Marketing".* Nigeria.

Payne, Adrian. 2005. "*Handbook of CRM: Achieving Excellent in Customer Management".* Butterworth-Heinemann, an imprint of Elsevier. London.

Reinhold, Olaf. 2011. "*Analytical Social CRM: Concept and Tool Support*". Bled, Slovenia

Nugroho, Yuda Septian. 2013. "*Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro". Universitas Dian Nuswantoro*

Slim, Ahmad, dkk. 2017. "*Predicting Student Enrollment Based on Student and College Characteristics*". University of New Mexico.

Srivastara, Jaideep. 2002. "*A Case for Analytical Customer Relationship Management*". *University of Minnesota*, Minneapolis.

Syarli, dan Ashari, Asrul M. 2016. "*Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi*)". Makasar.

Turban, E., McLean, E., and Wetherbe, J., (2006). *Information Technology for Management: Transforming Organisations in the Digital Economy, Fifth Edition*, New York.: John Wiley

Yaakub, Mohd Ridzwan. 2013. "*Integration of Sentiment Analysis into Customer Relational Model: The Importance of Feature Ontology and Synonym". Queensland University of Technology.* Australia.

Yulida, Roza. 2012. "*Kontribusi Usahatani Lahan Pekarangan Terhadap Ekonomi Rumah Tangga Petani Di Kecamatan Kerinci Kabupaten Pelalawan". Indonesian Journal of Agricultural Economics* (IJAE).

https://www.destinationcrm.com/Articles/Web-Exclusives/Viewpoints/An-Introduction-to-Sentiment-Analysis-the-Next-Great-*Customer*-Experience-Tool--134619.aspx , diakes pada tanggal 25 Februari, Jam 08.47 WIB.

https://www.bigdataframework.org/data-types-structured-vs-unstructured-data/ , diakses pada tanggal 26 Februari 2020, Jam 08.35 WIB.

https://kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+Internet+di+Indonesia+63+Juta+Orang/0/berita_satker, diakses pada tanggal 26 Februari 2020, Jam 09.06 WIB.

LAMPIRAN

Lampiran 1 Hasil Data Mining Tahun 2016-2020

Angkatan	Asal Kota	Agama	Sekolah	USM	Predic
Z	BANDUNG	B	SMAK GAMALIEL	T	0.334553
Z	BANDUNG	B	SMAK KALAM KUDUS	T	0.374817
Z	BANDUNG	B	SMAK 3 BINA BAKTI	T	0.309332
Z	BANDUNG	C	SMA SANTA MARIA 1	T	0.05598
Z	BANDUNG	B	SMAK 2 BPK PENABUR	T	0.490428
Z	CIMAHI	B	SMA SANTA MARIA 3	T	0.560147
Z	BANDUNG	C	SMAK GAMALIEL	T	0.148736
Z	BANDUNG	B	SMAK BPPK	T	0.309332
Z	BANDUNG	D	SMA TALENTA	T	0.050531
B	BANDUNG	B	SMAK GAMALIEL	T	0.938474
Z	BANDUNG	B	SMAK SANTO PAULUS	T	0.639626
E	BANDUNG	B	SMAK BPPK	Y	0.988182
G	BANDUNG	B	SMAK KALAM KUDUS	Y	0.988211
B	BANDUNG	B	SMAK REHOBOTH	Y	0.991687
A	BANDUNG	B	SMAK REHOBOTH	Y	0.983312
E	BANDUNG	C	SMA SANTA MARIA 1	Y	0.980054
B	KARAWANG	B	SMA YOS SUDARSO	T	0.483157
C	BANDUNG	B	SMAK GAMALIEL	Y	0.990796
D	BANDUNG	B	SMAK KALAM KUDUS	Y	0.992534
A	BANDUNG	B	SMAK KALAM KUDUS	Y	0.987358
Z	BANDUNG	B	SMA SANTA MARIA 1	T	0.248805
F	BANDUNG	B	SMA SANTA MARIA 3	Y	0.995107
C	BANDUNG	B	SMA TRINITAS	Y	0.975228
B	BANDUNG	C	SMA TRINITAS	Y	0.972488
D	BANDUNG	B	SMAK 1 BINA BAKTI	Y	0.983239
D	BANDUNG	C	SMAK BPK PENABUR HOLIS	T	0.542868
B	BANDUNG	B	SMAK BPPK	Y	0.992706
F	BANDUNG	B	SMAK GAMALIEL	Y	0.990464
G	BANDUNG	B	SMAK GAMALIEL	T	0.887429
B	BANDUNG	B	SMAK PELITA BANGSA	Y	0.984171

Angkatan	Asal Kota	Agama	Sekolah	USM	Predic
B	BANDUNG	B	SMAK SANTO PAULUS	Y	0.996954
B	BANDUNG	A	SMK BPI	T	0.735607
F	BANDUNG	A	SMK PASUNDAN 2 BANDUNG	Y	0.539755
C	BANDUNG	B	SMK PELITA BANGSA	Y	0.97584
F	BANDUNG	D	CSC BINTANG HARAPAN	Y	0.981361
D	BANDUNG	B	MUTIARA NUSANTARA	Y	0.985616
E	BANDUNG	B	PAKET C	Y	0.993417
G	BANDUNG	B	PKBM RAMA PUTRA	Y	0.988602
B	BANDUNG	B	SMA EL SHADAI	T	0.894008
B	BANDUNG	B	SMA MARDI YUANA CIANJUR	Y	0.975893
C	TANGERANG	B	SMA MUTIARA BANGSA	Y	0.71508
E	SUKABUMI	B	SMA PEMBINA BALAM RIAU	Y	0.627593
F	BANDUNG	B	SMA SANTA MARIA 1	Y	0.988508
E	BANDUNG	B	SMA SANTA MARIA 1	Y	0.986535
A	BANDUNG	Z	SMA SANTA MARIA 1	Y	0.943991
B	BANDUNG	B	SMA SANTA MARIA 2	Y	0.975893
G	BANDUNG	B	SMA SANTA MARIA 2	Y	0.955174
F	KAB. BANDUNG BARAT	B	SMA SANTA MARIA 3	Y	0.998613
B	CIMAHI	C	SMA SANTA MARIA 3	Y	0.994202
E	CIMAHI	B	SMA SANTA MARIA 3	Y	0.993666
E	BANDUNG	Z	SMA SANTA MARIA 3	Y	0.980496
G	CIMAHI	C	SMA SANTA MARIA 3	T	0.904549
Z	BANDUNG BARAT	B	SMA SANTA MARIA 3	T	0.475289
D	BANDUNG	B	SMA SANTO ANDREAS	Y	0.985616
F	BANDUNG	Z	SMA SANTA ANGELA	Y	0.943782
B	BANDUNG	C	SMA SUMATRA 40	Y	0.964381
F	BANDUNG	B	SMA TALENTA	Y	0.984527
E	BANDUNG	C	SMA TALENTA	Y	0.973184
A	BANDUNG	Z	SMA TALENTA	T	0.474306
B	BANDUNG	A	SMA TARUNA BAKTI	Y	0.905092

Angkatan	Asal Kota	Agama	Sekolah	USM	Predic
C	BANDUNG	B	SMA TIMIKA	Y	0.983828
D	BANDUNG	D	SMA TRINITAS	Y	0.97548
A	KARAWANG	B	SMA YOS SUDARSO	Y	0.857737
G	BANDUNG	B	SMAK 1 BINA BAKTI	T	0.781443
F	BANDUNG	B	SMAK 2 BPK PENABUR	Y	0.993452
B	BANDUNG	B	SMAK 3 BPK PENABUR	T	0.871962
A	BANDUNG	B	SMAK 3 BPK PENABUR	T	0.757648
E	BANDUNG	B	SMAK GAMALIEL	Y	0.988825
A	BANDUNG	B	SMAK GAMALIEL	Y	0.986147
H	BANDUNG	B	SMAK GAMALIEL	Y	0.975032
B	BANDUNG	C	SMAK KALAM KUDUS	Y	0.990665
E	BANDUNG	B	SMAK KALAM KUDUS	Y	0.989804
E	BANDUNG	C	SMAK REHOBOTH	Y	0.980054
B	BANDUNG	B	SMAK REHOBOTH	T	0.926245
C	BANDUNG	B	SMAK SANTO PAULUS	Y	0.995932
G	BANDUNG	B	SMAK SANTO PAULUS	Y	0.994285
D	BANDUNG	B	SMAK TRIMULIA	Y	0.957459
B	BANDUNG	A	SMAN 1	Y	0.438599
C	BANDUNG	A	SMAN 11 BANDUNG	Y	0.87517
E	BANDUNG	C	SMAN 15	Y	0.933575
A	BANDUNG	B	SMAN 27	Y	0.951989
D	CIMAHI	B	SMK BUNDA SATRIA	T	0.863232
B	BANDUNG	B	SMK KARTINI	Y	0.987873
C	BANDUNG	B	SMK MEDIKACOM	Y	0.991881
Z	BANDUNG	B	SMK SANDHY PUTRA TELKOM	T	0.064304
A	CIMAHI	B	SMK SANGKURIANG 1 CIMAHI	Y	0.960062
F	BANDUNG	B	SMKN 1	Y	0.907362
F	BANDUNG	A	SMKN 1 BANDUNG	Y	0.657088
C	BANDUNG	B	SMKN 2 CIMAHI	Y	0.884769
G	CIMAHI	B	SMKN 2 CIMAHI	Y	0.826541
F	CIMAHI	A	SMKN 2 CIMAHI	Y	0.272719

Angkatan	Asal Kota	Agama	Sekolah	USM	Predic
B	BANDUNG	C	SMKN 7	T	0.600634
C	BANDUNG BARAT	B	SMTK BETHEL	Y	0.977537
B	BANDUNG	A	TEMASEK INDEPENDENTSCHOOL	Y	0.905092
E	CIREBON	B	SMAN 1 BABAKAN	Y	0.66989
E	BANDUNG	A	SMAN 1 CILEUNYI	Y	0.850308
G	BANDUNG	A	SMAN 1 MANGGAR	Y	0.828799
A	BANDUNG	A	SMAN 1 MARGAASIH	T	0.040971
A	SUBANG	C	SMAN 1 SUBANG	Y	0.526266
H	GARUT	A	SMAN 13	Y	0.005417
A	BANDUNG	A	SMAN 2 TEBO	T	0.040971
B	BANDUNG	A	SMAN 20	Y	0.818784
B	KAB. BANDUNG BARAT	B	SMAN 5	T	0.575702
G	BANDUNG	A	SMAN BANGGAI LAUT	T	0.075982
B	SUBANG	A	SMK ANGKASA 1 KALIJATI	Y	0.083396

Lampiran 2 Hasil *Data Mining* Tahun 2016

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
Z	BANDUNG	B	SMAK GAMALIEL	T	0.576261
B	BANDUNG	B	SMK TELKOM SANDHY PUTRA	Y	0.980979
Z	BANDUNG	B	SMAK KALAM KUDUS	T	0.445048
C	BANDUNG	B	SMK PELITA BANGSA	Y	0.960895
Z	BANDUNG	B	SMAK 2 BPK PENABUR	T	0.238805
Z	BANDUNG	C	SMA SANTA MARIA 1	T	0.256237
C	BANDUNG	B	SMAK REHOBOTH	Y	0.973759
C	BANDUNG	B	SMK YADIKA KALIJATI	Y	0.92329
Z	BANDUNG	B	SMAK YAHYA	T	0.445048
A	BANDUNG	C	YPS SEJATI	Y	0.875095
C	SUMEDANG	A	SMU NEGRI CIKERUH	Y	0.79671

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
E	BANDUNG	C	SMK BUDAYA BANGSA	Y	0.764874
C	BANDUNG	A	SMA ALMASOEM	Y	0.71606
F	PALEMBANG	B	SMA BUKIT ASAM TANJUNGENM	Y	0.643822
C	TANGERANG	B	SMA MUTIARA BANGSA	Y	0.641777
F	BANDUNG	A	SMA AL IHSAN BELEENDAH	Y	0.583658
B	BANDUNG	B	SMAK REHOBOTH	Y	0.987279
G	BANDUNG	A	SMAN 1 MANGGAR	Y	0.576201
D	BANDUNG	B	SMAK PELITA BANGSA	Y	0.880809
E	BANDUNG	B	SMAK KALAM KUDUS	Y	0.957705
C	CIANJUR	A	SMAN 1 SUKANAGARA	Y	0.335254
A	BANDUNG	C	SMAK KALAM KUDUS	Y	0.935534
Z	BANDUNG	B	SMAK 3 BPK PENABUR	T	0.322708
F	BANDUNG	B	SMA TALENTA	T	0.937942
B	SUKABUMI	A	SMK KOMPUTER AL BASHIY	Y	0.134631
Z	BANDUNG	B	SMKN 1 GUNUNGSITOLI	T	0.131175

Lampiran 3 Hasil Data Mining Tahun 2017

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
Z	BANDUNG	B	SMAK GAMALIEL	T	0.506268
Z	BANDUNG	B	SMAK 3 BINA BAKTI	T	0.623138
Z	BANDUNG	B	SMAK 1 BINA BAKTI	T	0.235216
Z	BANDUNG	B	SMAK REHOBOTH	T	0.235216
B	BANDUNG	C	SMA TRINITAS	Y	0.999541
Z	CIMAHI	C	SMK FARMASI BPK PENABUR	T	0.526491
Z	BANDUNG	B	SMAK SANTO PAULUS	T	0.235216
B	BANDUNG	B	SMAK REHOBOTH	Y	0.996355
B	BANDUNG	B	SMAK KALAM KUDUS	Y	0.998176
B	BANDUNG	B	SMAK GAMALIEL	Y	0.998069
C	BANDUNG	B	SMAK GAMALIEL	Y	0.997747

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
C	BANDUNG	B	SMAK BPK PENABUR HOLIS	Y	0.991516
A	BANDUNG	B	SMAK BPK PENABUR HOLIS	Y	0.974761
Z	BANDUNG	B	SMAK 3 BPK PENABUR	T	0.235216
B	BANDUNG	A	SMA TARUNA BAKTI	Y	0.806198
F	BANDUNG	C	SMA TALENTA	Y	0.998929
Z	BANDUNG BARAT	C	SMA SANTA MARIA 1	T	0.012711
G	BANDUNG	B	PKBM RAMA PUTRA	Y	0.998297
C	BANDUNG	D	PKBM MINDA UTAMA	Y	0.997753
G	BANDUNG	B	TELKOM UNIVERSITY	Y	0.996598
B	BANDUNG	B	SEKOLAH BINA PERSADA	Y	0.996355
C	BANDUNG	C	SMAN 14	Y	0.990403
A	BANDUNG	B	SEKOLAH BINA PERSADA	Y	0.9873
G	TARUTUNG	B	SMAN 1	Y	0.92306
G	KALIMANTAN TIMUR	B	SMK PUTRA BANGSA	T	0.920665
D	BANDUNG	A	SMK BINA SARANA CENDEKIA	Y	0.919888
F	BANDUNG	B	SMKN 1	Y	0.913065
B	BANDUNG	A	SMAN 9	Y	0.898165
G	KAB. BANDUNG	A	SMAN 1 MARGAHAYU	Y	0.854643
F	BANDUNG	A	SMKN 1 BANDUNG	Y	0.845947
D	INDRAMAYU	A	SMK MUHAMMADIYAH 1	Y	0.04473

Lampiran 4 Hasil Data Mining Tahun 2018

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
Z	BANDUNG	B	SMAK GAMALIEL	T	0.232752
Z	BANDUNG	B	SMAK KALAM KUDUS	T	0.413503
Z	BANDUNG	B	SMAK REHOBOTH	T	0.092987
B	BANDUNG	B	HOME SCHOOLING	Y	0.951768
Z	BANDUNG	B	SMAK 2 BPK PENABU	T	0.656367
Z	CIMAHI	C	SMA SANTA MARIA 3	T	0.577009
Z	BANDUNG	B	SMAK PELITA BANGS	T	0.092987

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
D	BANDUNG	B	SMAK KALAM KUDUS	Y	0.967585
E	BANDUNG	B	SMAK BPPK	Y	0.947499
B	BANDUNG	C	SMA ADVENT	Y	0.948864
G	BANDUNG	B	SMAK KALAM KUDUS	Y	0.880504
Z	BANDUNG	C	SMAK KALAM KUDUS	T	0.388349
B	BANDUNG	B	SMK PASUNDAN JATINANGOR	Y	0.905808
D	BANDUNG	B	SMAK GAMALIEL	Y	0.951768
E	BANDUNG	C	SMA PANGUDI LUHU	Y	0.866437
A	BANDUNG	B	SMAK GAMALIEL	Y	0.920885
F	BANDUNG	B	SMAN 2 MEDAN	Y	0.861962
H	BANDUNG	B	SMAK GAMALIEL	Y	0.861962
D	CIMAHI	B	SMK BUNDA SATRIA	T	0.815118
G	BANDUNG	Z	SMAN 2 CIREBON	Y	0.785423
C	BANDUNG	B	SMAK BASIC	Y	0.741749
G	BANDUNG	B	SMAN 6 BANDUNG	Y	0.679776
G	BANDUNG	B	SMK GUNA DHARMA	Y	0.679776
B	BANDUNG	C	SMAK GAMALIEL	T	0.818313
D	BANDUNG	C	SMAK BPK PENABUR HOLIS	T	0.666886
C	BANDUNG	B	SMAK 2 BPK PENABU	Y	0.951768
D	BANDUNG	B	SMAK 1 BINA BAKTI	Y	0.905808
B	BANDUNG	B	SMA TRINITAS	Y	0.905808
E	BANDUNG BARAT	C	SMA SANTA MARIA 3	Y	0.929931
H	BANDUNG	B	SMA SANTA MARIA 3	Y	0.894654
Z	BANDUNG BARAT	C	SMA SANTA MARIA 3	T	0.197204
B	BOGOR	B	SMA MARDI YUANA CIANJUR	Y	0.956394
B	BANDUNG	B	SMA MARDI YUANA CIANJUR	Y	0.905808
D	CIANJUR	A	SMK BELA NUSANTA CIANJUR	Y	0.08774

Lampiran 5 Hasil Data Mining Tahun 2019

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
Z	BANDUNG	B	SMAK GAMALIEL	T	0.622095
Z	BANDUNG	B	SMAK REHOBOTH	T	0.708165
Z	BANDUNG	B	SMAK BPPK	T	0.322278
Z	BANDUNG	B	SMA TALENTA	T	0.490653
Z	BANDUNG	B	SMAK 3 BPK PENABUR	T	0.322278
Z	BANDUNG	B	SMK GRACIA	T	0.592014
B	BANDUNG	B	SMAK PELITA BANGSA	Y	0.97438
Z	BANDUNG	B	SMAK 3 BINA BAKTI	T	0.708165
C	BANDUNG	C	SMA SANTA MARIA 1	Y	0.930274
E	BANDUNG	Z	SMK GRACIA	T	0.520648
G	BANDUNG	B	SMAK SANTO PAULUS	Y	0.928672
D	BANDUNG	B	SMAK GAMALIEL	Y	0.992166
E	BANDUNG	B	SMAK GAMALIEL	Y	0.98246
Z	CIMAHI	B	SMAK GAMALIEL	T	0.667095
E	BANDUNG	B	SMAK BPPK	Y	0.961814
G	BANDUNG	B	SMAK BPPK	Y	0.862257
A	CIMAHI	A	SMAK BPPK	Y	0.460957
G	BANDUNG	B	SMAK 1 BINA BAKTI	T	0.260403
F	BANDUNG	B	SMA TALENTA	Y	0.981302
D	BANDUNG	B	SMA SANTA MARIA 3	Y	0.996163
A	CIMAHI	C	SMA SANTA MARIA 3	Y	0.946431
G	BANDUNG	B	SMA SANTA MARIA 2	Y	0.800264
Z	BANDUNG	C	SMA SANTA MARIA 1	T	0.355393
D	BANDUNG	B	MUTIARA NUSANTARA	Y	0.982846
D	BANDUNG	A	SMAN 13	Y	0.968614
E	BANDUNG	B	SMKN 1 KATAPANG	Y	0.961814

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
F	BANDUNG	A	SMK ANGKASA 1 MARGAHAYU	Y	0.72931
F	BANDUNG	A	SMAN 1	Y	0.672712
A	BANDUNG BARAT	C	SMK PASUNDAN 2 BANDUNG	Y	0.527159
F	BANDUNG	A	SMKN RAJAPOLAH	T	0.52619
Z	CIMAHI	B	SMAN 4 CIMAHI	T	0.389544
G	BANDUNG	A	SMKN 5 BANDUNG	Y	0.100187
Z	BANDUNG	C	SMK DHARMA BHAKTI	T	0.025024

Lampiran 6 Hasil *Data Mining* Tahun 2020

Gelombang	Asal Kota	Agama	ASAL SMU	USM	predic
B	BANDUNG	B	SMK GRACIA	T	0.867687
B	BANDUNG BARAT	B	SMKN 4 BANDUNG	T	0.742216
B	BANDUNG	B	SMA SANTA MARIA 2	T	0.617935
B	BANDUNG	B	SMAK 3 BPK PENABUR	T	0.602628
B	BANDUNG	C	SMAK GAMALIEL	T	0.582123
A	BANDUNG	B	SMAK KALAM KUDUS	T	0.511665
B	BANDUNG	B	SMA TALENTA	T	0.239178