

# **PENERAPAN DATA MINING DALAM PEMBUATAN STRATEGI PROMOSI MAHASISWA BARU INSTITUT TEKNOLOGI TELKOM SURABAYA**

**Haniyah Al Kalimah<sup>\*1</sup>**

<sup>1</sup>Sistem Informasi, Fakultas Teknologi Informasi dan Bisnis, Institut Teknologi Telkom Surabaya, Jl. Ketintang No.156, Ketintang, Kec. Gayungan, Surabaya, 60231, Indonesia  
alliciaputri@student.itelkom-sby.ac.id

## ***Abstrak***

Tingkat persaingan perguruan tinggi mengalami peningkatan yang sangat pesat setiap tahun. Penerimaan mahasiswa baru merupakan aktivitas penting bagi perguruan tinggi untuk menjaring mahasiswa baru. Dalam proses mendapatkan mahasiswa baru setiap perguruan tinggi memiliki strategi promosi yang berbeda. Penentuan strategi promosi yang tepat dapat menekan biaya promosi mencapai sasaran promosi yang tepat. Penelitian ini bertujuan untuk mengidentifikasi pola karakteristik profil mahasiswa dan calon mahasiswa sebagai rekomendasi strategi promosi. Pengolahan Data Mining menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) dengan model klasterisasi dan klasifikasi yang memanfaatkan algoritma *K-Means Clustering* dan *Decision Tree*. Implementasi pengolahan data menggunakan *RapidMiner Studio 9.10.1*. Atribut yang digunakan dalam model klasterisasi adalah Nim, KotaAsal, Asal Sekolah, Prodi, Jurusan Sekolah, Jenis Kelamin, Jalur Seleksi, Tahun pendaftaran. Pengolahan data menggunakan algoritma *K-Means Clustering* menghasilkan 3 buah *cluster*, *cluster\_1* berjumlah 91 *items* (4,56%) dominasi mahasiswa paling banyak dari program studi Sistem Informasi. *Cluster\_2* berjumlah 1683 *items* (84,48%) dominasi mahasiswa paling banyak dari Kota Surabaya dan Kab. Sidoarjo, *Cluster\_3* berjumlah 218 *items* (10,96%) dominasi mahasiswa paling banyak dari daerah Kawasan Indonesia Timur dengan Prodi Informatika. Nilai *Davies Bouldin Index* (DBI) sebesar 0,477. Atribut yang digunakan dalam model klasifikasi adalah No.Peserta, Status Payment Biaya UP3, Kab./Kota Asal, Kab./Kota Asal sekolah, Jenis Sekolah, Kategori Sekolah, Sumber Informasi, Penghasilan orang tua, Pilihan Prodi. Data set yang digunakan sebanyak 3.760 *record* dengan *mode test percentage split* yaitu 70% sebanyak 2.637 sebagai data *training* dan 30% sebanyak 1.113 sebagai data *testing*. Pengolahan data menggunakan *Decision Tree* menghasilkan nilai *accuracy* pengujian sebesar 98,86%, nilai *precision* sebesar 93,33%, nilai *recall* sebesar 32,56% dan AUC sebesar 0.992. Atribut yang paling berpengaruh paling tinggi berdasarkan *Decision Tree* menghasilkan sebuah pohon keputusan untuk memudahkan pengambilan keputusan strategi yang tepat.

***Kata Kunci: Data mining, Strategi Promosi, CRISP-DM, K-Means, Decision Tree***

## **1. Pendahuluan (Introduction)**

Institut Teknologi Telkom Surabaya merupakan perguruan tinggi swasta dibawah naungan Yayasan Pendidikan Telkom yang berfokus di bidang maritim, transportasi, dan logistik. Institut Teknologi Telkom Surabaya membawa warna baru dalam dunia Pendidikan Teknologi di Surabaya, hal ini menimbulkan persaingan antar perguruan tinggi swasta di Surabaya semakin ketat. Persaingan yang ketat antar perguruan tinggi memungkinkan semua institusi Pendidikan menjalankan institusinya secara profesional (Oktaviarna Tensao et al., n.d.-a). Penerimaan Mahasiswa Baru (PMB) merupakan program rutin yang dijalankan oleh setiap lembaga pendidikan, seperti perguruan tinggi untuk menerima mahasiswa baru setiap tahunnya. Sebagai perguruan tinggi yang mengelola keuangannya sendiri, setiap perguruan harus aktif mempromosikan keunggulan masing-masing untuk menarik mahasiswa baru. Semakin banyak mahasiswa baru yang diterima, maka semakin tinggi pendapatan keuangan perguruan tinggi agar kegiatan belajar mengajar dapat terlaksana dengan optimal (Abriyanto & Damastuti, 2019).

Institut Teknologi Telkom Surabaya mengharapkan jumlah mahasiswa baru setiap tahunnya memenuhi target, sehingga pihak admisi memiliki target mahasiswa baru pada tiap tahunnya. Kegiatan promosi yang dilakukan oleh pihak admisi Institut Teknologi Telkom Surabaya saat ini secara *online* seperti sosialisasi dan periklanan, yaitu periklanan nonpersonal yang dapat diidentifikasi dengan menggunakan berbagai media untuk mempengaruhi calon mahasiswa, selain itu juga melakukan promosi langsung (*offline*) merupakan bentuk pemasaran individu dengan mengunjungi sekolah secara langsung (*campus expo*) dan mengundang sekolah secara langsung (*open house*). Mahasiswa yang mendaftar di Institut Teknologi Telkom Surabaya berasal dari berbagai sekoalah negeri maupun swasta yang tersebar diberbagai wilayah. Maka dari itu persebaran lokasi sosialisasi yang dilakukan meningkat setiap tahunnya, sehingga pihak admisi perlu mempertimbangkan serta menganalisa untuk melakukan pemasaran kepada calon mahasiswa di sekolah daerah lain agar hasilnya lebih maksimal. Disamping itu promosi periklanan sedang dimaksimalkan untuk memperoleh mahasiswa yang mendaftar, tetapi iklan dijadikan sebagai kesadaran merek (*Brand Awareness*) dimana minat calon mahasiswa masih kurang sehingga derlukan tindakan lanjut denga cara sosialisasi ke sekolah dan daerah calon mahasiswa. Permasalahan yang terjadi setelah dilakukan berbagai cara sosialisasi dan promosi untuk menarik minat mahasiswa namun hasil yang didapat tidak sesuai dengan apa yang diharapkan. Oleh karena itu, pihak admisi membutuhkan strategi pemasaran yang baik untuk memenuhi target mahasiswa baru. Maka diperlukan pengetahuan baru dalam menentukan strategi promosi untuk mendapatkan target mahasiswa baru dengan tepat sasaran.

Perkembangan ilmu pengetahuan saat ini sangat maju dan dapat dilihat di segala bidang kehidupan, dari berbagai bidang seperti industri, ekonomi dan Pendidikan. Pemanfaatan teknologi dalam bidang Pendidikan khususnya di perguruan tinggi dapat menghasilkan informasi yang berlimpah berdasarkan data historis, sehingga data akan bertambah secara terus menerus misalnya data pendaftaran mahasiswa baru (Abriyanto & Damastuti, 2019). Oleh karena itu, diperlukan suatu teknik atau perangkat yang membantu dalam mentransformasikan data dengan jumlah yang besar tersebut menjadi informasi yang bermanfaat, yaitu penerapan data mining untuk strategi promosi penerimaan mahasiswa baru, yang dapat dijadikan sebagai dasar atau pedoman untuk mendukung pengambilan keputusan, terutama dalam keputusan pembuatan strategi promosi dalam penerimaan mahasiswa baru. Data mining adalah proses menggunakan teknik statistik matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang berguna dan informasi terkait dari basis data besar (Mirza, 2019). Teknik data mining yang dapat digunakan diantaranya; klasifikasi, klusterisasi, asosiasi dan berbagai teknik yang tepat dapat digunakan. Penggunaan teknik data mining bertujuan untuk mempercepat proses pengambilan keputusan untuk mengubah informasi yang terkandung dalam data menjadi (*knowledge*) yang baru (Mirza, 2019).

Penelitian ini bertujuan untuk mengimplementasikan teknik data mining klusterisasi dan klasifikasi, menggunakan algoritma *K-Means Clustering* dan algoritma *Decision Tree*. *K-Means* merupakan metode *clustering* data *non-hierarchical* yang dapat mengelompokkan data kedalam beberapa *cluster* berdasarkan kesamaan data, sehingga data dengan karakteristik yang sama dikelompokkan dalam satu *cluster* dan data karakteristik yang berbeda dikelompokkan dalam *cluster* lain (Oktaviarna Tensao et al., n.d.-b). *Decision Tree* merupakan model prediksi yang paling umum digunakan dalam data mining karena menggunakan struktur hirarki dan pohon. Pada konsep pohon keputusan didefinisikan sebagai pengganti data yang menjadi aturan keputusan untuk mengubah data menjadi *Decision Tree* dan aturan-aturan keputusan (Inda Sari & Nasir, n.d.). Pengolahan data bertujuan untuk mengetahui karakteristik profil mahasiswa dan calon mahasiswa sebagai penunjang untuk menentukan strategi promosi mahasiswa baru.

Tahapan pengolahan data mining dalam penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining* CRISP-DM, merupakan standarisasi proses data mining sebagai

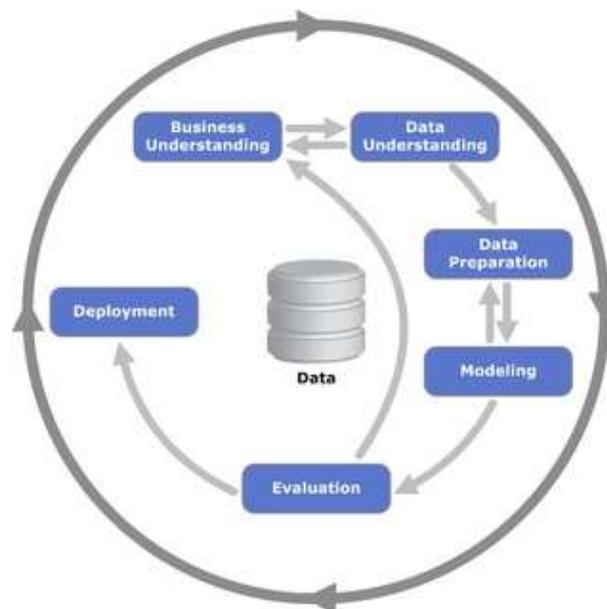
strategi pemecahan masalah secara umum dari sebuah penelitian dengan melalui proses *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment*. Untuk membantu pengolahan data mining proses implementasi menggunakan *software RapidMiner Studio 9.10.1*.

Dengan pengolahan data mining, data calon mahasiswa baru dan data mahasiswa yang tersimpan dapat digunakan untuk mendapatkan karakteristik profil sebagai rekomendasi strategi promosi yang tepat sasaran.

## **2. Metode Penelitian (Methods)**

### **2.1 CRISP-DM**

Penelitian ini mengacu pada frame work data mining yaitu metode CRISP- DM (*Cross Industry Standard Process for Data Mining*). Berikut tahapan yang dilakukan dalam menganalisis data :



Gambar 2.1 Alur CRISPDPM

CRISP-DM merupakan strategi pemecahan masalah secara umum dari bisnis atau suatu penelitian dengan menyediakan standar proses data mining yang memiliki enam tahapan. Fase pertama dimulai dari *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment* (Sri et al., 2021). Dalam siklus pengembangannya, CRISP-DM diakui sebagai metode data mining terlengkap untuk memenuhi kebutuhan proyek industri karena metode ini paling banyak digunakan dalam proyek analitik, data mining dan sains data. Gambar diatas adalah alur proses dalam metode CRISP-DM yang biasa digunakan dalam data mining. Setip fase memiliki tugas dan tanggung jawab masing-masing. Panah pada gambar menunjukkan hubungan antar langkah-langkah tersebut, Lingkaran luar dari dari setiap fase juga mewakili siklus data mining. Tujuan metode CRISP-DM adalah untuk membuat proyek data mining lebih mudah dikelola (Kurniawan & Yasir, n.d.).

#### **2.1.1 Business Understanding**

*Business Understanding* adalah tahap pertama dalam proses CRISP-DM yang juga dapat disebut sebagai tahap pemahaman bisnis (penelitian). Pada tahap ini fokus pada analisis kebutuhan bisnis dengan mendefinisikan masalah dalam data mining kemudian merumuskan tujuan dan menyiapkan strategi untuk mencapai tujuan tersebut (Kusumawardana, n.d.).

Penelitian ini bertujuan untuk melakukan analisis terhadap model data mining yaitu klusterisasi dan klasifikasi dengan kinerja algoritma *K-Means* dan *Decision Tree* terhadap data set untuk menentukan cluster dan mengidentifikasi pola-pola yang terbentuk dari data masa lampau. Sehingga dapat menjadi rekomendasi tim admisi dalam menyusun strategi promosi.

Tujuan bisnis pada penelitian ini untuk meningkatkan jumlah aplikasi yang mendaftar di Institut Teknologi Telkom Surabaya dan juga meningkatkan calon mahasiswa yang melakukan registrasi penerimaan mahasiswa baru.

Penilaian Situasi Institut Teknologi Telkom Surabaya merupakan perguruan tinggi swasta dibawah naungan Yayasan Pendidikan Telkom yangberfokus pada bidang maritim, transportasi dan logistik. Institut Teknologi Telkom Surabaya memiliki 2 (dua) fakultas dan 11 (sebelas) program studi diantaranya :

- 1.) Fakultas Teknologi Telkom Informasi dan Bisnis yang memiliki 6 (enam) program studi antara lain; Sistem Informasi, Sains Data, Teknologi Informasi, Rekayasa Perangkat Lunak, Informatika, dan Bisnis Digital.
- 2.) Fakultas Teknologi Elektronika dan Industri Cerdas yang memiliki 5 (lima) program studi antara lain; Teknik Telekomunikasi, Teknik Industri, Teknik Komputer dan Teknik Elektro.

Institut Teknologi Telkom Surabaya memiliki unit bagian khusus untuk mempromosikan kampus yang bertujuan menarik calon mahasiswa yaitu unita bagian Admisi. Berdasarkan hasil wawancara dengan pihak admisi yaitu dengan Ibu Pramaditya Arismawati selaku staff admisi dan data, diketahui bahwa upaya strategi promosi yang dilakukan dilakukan secara *offline* dan *online*. Promosi *offline* dilakukan dengan mengadakan kunjunganke sekolah-sekolah atau *campus expo*, selain itu juga mengadakan kegiatan *open house* atau mengundang sekolah-sekolah datang di Institut Teknologi Telkom Surabaya. Kegiatan promosi langsung yang dilakukan oleh tim admisi dioptimalkan pada wilayah Jawa Timur khususnya kota Surabaya, sedangkan promosi yang dilakukan untuk wilayah diluar Jawa Timur dilakukan dengan pendekatan secara online yang menggunakan empat mediayaitu Google adds, Facebook, Instagram, dan Tiktok. Kegiatan promosi online diantaranya memasang iklan Instagram, mengadakan *sharing session* untuk mengenalkan prodi-prodi. Namun dari upaya yang sudah dilakukan oleh tim admisi strategi promosi yang jalankan kurang maksimal karena daritahun ke tahun setiap kepengurusan belum ada fix program.

Tujuan data mining dalam penelitian ini yaitu untuk klusterisasi karakteristik profil mahasiswa dan mengidentifikasi pola klasifikasi profil calon mahasiswa yang melakukan pendaftaran berdasarkan status pembayaran UP3 sebagai rekomendasi dalam menentukan strategi promosi mahasiswa baru.

### **2.1.2 Data Understanding**

Pada tahap pemahaman data (*Data Understanding*) terdiri dari beberapa proses diantaranya pengumpulan data awal, deskripsi data, eksplorasi data dan verifikasi kualitas data. Proses pengumpulan bertujuan untuk mempelajari struktur database, selanjutnya menentukan tabel-tabel yang dipergunakan pada proses data mining. Proses deskripsi data bertujuan untuk menganalisis sumber data awal yang didapatkan. Proses eksplorasi data bertujuan untuk menentukan sumber data yang digunakan. Proses verifikasi kualitas data bertujuan untuk memastikan kualitas data yang digunakan pada proses data mining tidak ada kesalahan (Niqotaini, 2021).

#### **2.1.4 Modelling**

Pemodelan adalah proses yang secara langsung melibatkan teknik datamining dan penentuan algoritma yang digunakan. Teknik pemodelan dalam penelitian ini menggunakan model clustering dan classification.

#### **2.1.5 Evaluation**

Evaluasi merupakan fase lanjutan terhadap tujuan data mining. Evaluasi dilakukan secara menyeluruh agar hasil dari tahap modeling sesuai dengan tujuan yang ingin dicapai dalam tahap pemahaman bisnis.

#### **2.1.6 Deployment**

Tahapan ini memberikan informasi dari data yang sudah diolah melalui tahapan-tahapan sebelumnya. Informasi yang diberikan berupa bentuk laporan dan visualisasi data yang sesuai dengan pemahaman bisnis.

### **3. Hasil dan Pembahasan (Results and Discussions)**

Pengolahan data dalam penelitian ini dilakukan dengan model klusterisasi dan klasifikasi mengacu pada tahapan yang sesuai dengan pengembangan tahapan CRISP-DM yang telah dijelaskan sebelumnya, yaitu terdiri enam tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*.

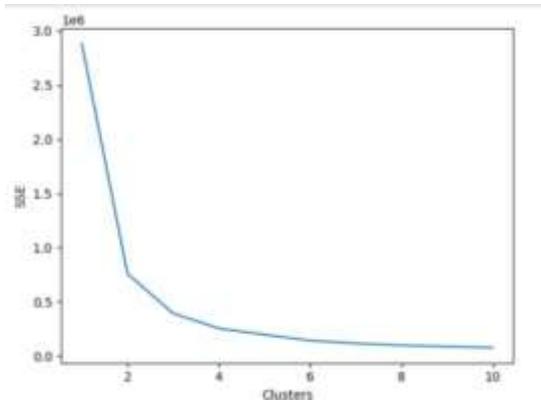
#### **3.1 Analisa Alur Pengolahan Data Model Klusterisasi**

Pengolahan data model klusterisasi pada penelitian dilakukan untuk mengolah data mahasiswa yang bertujuan untuk mengetahui karakteristik profil mahasiswa. Pengolahan data menggunakan data primer, yang diperoleh dengan melakukan permintaan data kepada Pusat Teknologi Informasi (PUTI) Institut Teknologi Telkom Surabaya. Data set yang diperoleh sebanyak 1992 data merupakan data mahasiswa Angkatan 2020, 2021, dan 2022 yang berisi data profil mahasiswa terdiri dari No.Peserta, Kota asal, Provinsi asal, Asal Sekolah, Jurusan Sekolah, Program Studi, Status mahasiswa, Pekerjaan Ortu, Tempat, Tanggal lahir, Tahun Angkatan, Nilai SMA/K, NIM, Jenis Kelamin, Beasiswa, Status pembayaran, Penghasilan orang tua, Kota Sekolah, Provinsi Sekolah, Usia, Jalur seleksi, tahun lulus, dan Tahun pendaftaran. Data-data akan dikelompokkan ke dalam kluster dengan algoritma *K-Means Clustering*.

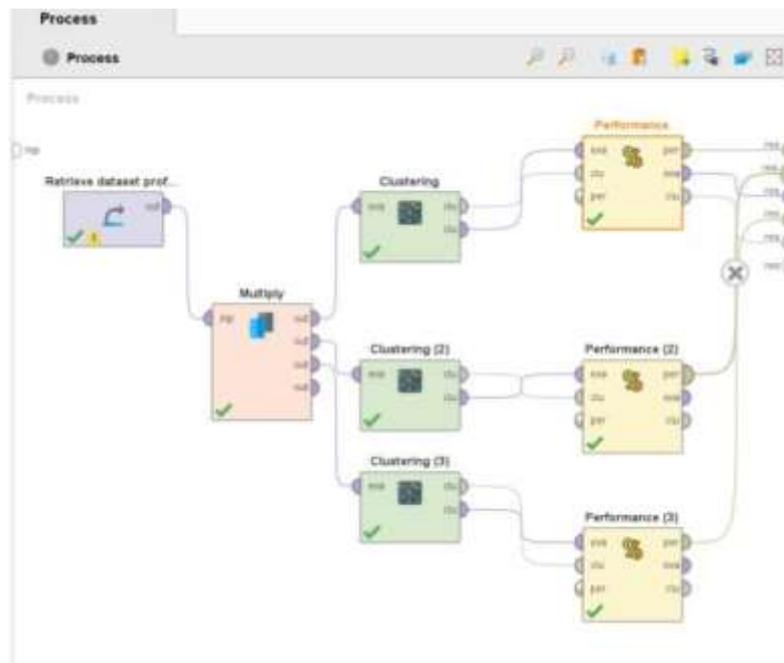
##### **3.1.1 Pengujian Dengan Algoritma K-Means Clustering**

Model *Clustering* yang digunakan adalah *non-hierarchical clustering* dengan algoritma *K-Means*. Tahap awal yang dilakukan yaitu menentukan jumlah K kluster terlebih dahulu. Untuk menentukan *cluster* yang optimal penelitian ini menggunakan metode Elbow. Metode Elbow merupakan metode yang digunakan untuk memberikan informasi untuk menentukan jumlah cluster terbaik dengan melihat persentase hasil membandingkan jumlah cluster (k) yang membentuk siku pada suatu titik. Dalam penelitian ini optimasi *cluster* dieksekusi dengan *Google Colaboratory* yang menggunakan bahasa pemrograman *Python*. Dari proses tersebut maka menghasilkan jumlah

*cluster* yang optimal seperti yang pada gambar dibawah ini. Setelah menentukan optimasi cluster Langkah selanjutnya yaitu mengimport dataset kedalam *machine mining*. *Machine mining* yang digunakan untuk implementasi proses algoritma *K-Means clustering* adalah RapidMiner 10.1. Dari proses *data preparation* maka dataset baru yang siap dieksekusi oleh tools Rapidminer.



**Gambar 3.1** Grafik SSE

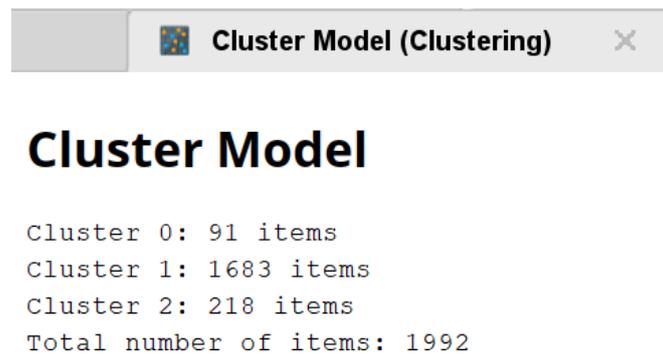


**Gambar 3.2** Hasil Model Clustering

Setelah proses *running* maka untuk menentukan jumlah cluster terbaik dapat dilihat hasil pada nilai *DBI (Davies Bouldin Index)* yang paling kecil, semakin kecil nilai *DBI* maka semakin optimal *cluster* yang dihasilkan. Berdasarkan proses diatas nilai *DBI* terkecil dengan nilai 0.477 pada *cluster* yang berjumlah 3 *cluster*.

Setelah proses *running* maka untuk menentukan jumlah cluster terbaik dapat dilihat hasil pada nilai *DBI (Davies Bouldin Index)* yang paling kecil, semakin kecil nilai *DBI* maka semakin optimal *cluster* yang dihasilkan. Berdasarkan proses diatas nilai *DBI* terkecil dengan nilai 0.477 pada

*cluster* yang berjumlah 3 *cluster*. Berikut hasil 3 *cluster* yang diperoleh.



**Gambar 3.3** Hasil Model Clustering

Dengan menggunakan pemodelan algoritma *K-Means clustering* pada gambar diatas maka dihasilkan jumlah *cluster* sebanyak 3 dengan dataset sebanyak 1.992 *record*. Dari gambar diatas dapat didefinisikan jumlah nilai *k* pada *cluster\_0* : 91 items, *cluster\_1* : 1683 items, *cluster\_2* : 218 items. Dari hasil *clustering* yang telah dilakukan maka dapat disimpulkan bahwa:

*Cluster 1* Program studi yang diminati adalah Sistem Informasi, Informatika, Bisnis Digital, Teknik Industri, Teknik Telekomunikasi, Teknik Elektro, Rekayasa Perangkat Lunak, Teknologi Informasi, Sains Data, Teknik Komputer, Teknik Logistik tetapi untuk program studi yang dominan adalah Sistem Informasi dengan dominan asal sekolah SMAN dengan jurusan MIPA, Jalur seleksi dominan jalur raport dan kota asal didominasi pada wilayah Kawasan Indonesia Timur.

*Cluster 2* Program studi yang diminati adalah Sistem Informasi, Teknik Industri, Teknologi Informasi, Informatika, Bisnis Digital, Teknik Telekomunikasi, Teknik Komputer, Rekayasa Perangkat Lunak, Teknik Elektro, Sains Data, Teknik Logistik, tetapi untuk program studi yang dominan adalah Sistem Informasi dengan dominan asal sekolah SMAN dengan jurusan MIPA, Jalur seleksi dominan jalur raport dan kota asal yang dominan berasal dari Kota Surabaya, Kab. Sidoarjo, dan Kab. Gresik.

*Cluster 3* Program Studi yang diminati adalah Informatika, Sistem Informasi, Teknik Industri, Bisnis Digital, Teknik Elektro, Teknik Telekomunikasi, Rekayasa Perangkat Lunak, Teknologi Informasi, Sains Data, Teknik Komputer tetapi untuk program studi yang paling dominan adalah Informatika dengan dominan asal sekolah SMAN dengan jurusan MIPA, Jalur seleksi dominan jalur raport dan kota asal yang dominan berasal dari Kota Jayapura, Kota Kupang, dan Kota Madiun.

Evaluasi memakai metode *Davies Bouldin index* untuk memilih jumlah *cluster* terbaik dengan cara melihat presentase hasil perbandingan antar jumlah *cluster*. Semakin kecil nilai semakin bagus pembagian *clusternya*.

Tabel 3.1. Hasil Davies Bouldin Index

| Penentuan jumlah <i>cluster</i> terbaik | Hasil <i>Davies Bouldin Index</i> |
|---|-----------------------------------|
| 3 <i>Cluster</i>                        | 0.477                             |
| 4 <i>Cluster</i>                        | 0.535                             |
| 5 <i>Cluster</i>                        | 0.544                             |

Dari tabel diatas dapat disimpulkan bahwa jumlah *cluster* optimal berjumlah 3 *cluster*. Dari hasil dengan hasil penelitian yang diperoleh 3 *cluster*, maka dilakukan analisis *cluster* yang berisikan hasil pengelompokkan berdasarkan kedekatan jarak.

### **3.2 Analisa Alur Pegolahan Data Model Klasifikasi**

Pengolahan data model klasifikasi pada penelitian dilakukan untuk menganalisis kinerja algoritma *Decision Tree* serta mengidentifikasi pola-pola yang terbentuk dari data masa lampau untuk penelusuran potensi dan minat calon mahasiswa baru. Sehingga tim admisi dapat menerapkan strategi promosi untuk menarik minat mahasiswa baru. Pengolahan data menggunakan data primer, yang diperoleh dengan melakukan wawancara kepada bagian Admisi. Dataset yang diolah dalam model klasifikasi merupakan data calon mahasiswa baru tahun 2022 dan 2023 yang berikan data profil terdiri dari No. Peserta, PIN, Status, Jalur, Nama Lengkap, Asal Sekolah, Jenis Sekolah, Kategori Sekolah, Kab./Kota Sekolah, Provinsi Sekolah, Email, No.Telp, No.HP, Provinsi, Kabupaten, Kecamatan, Pilihan prodi 1, Pilihan prodi 2, Pilihan prodi 3, Biaya, Tanggal Daftar Akun, Kode Diskon, Nama Wali, No.HP Wali, Hubungan Wali, Status Seleksi, Status Aktifasi Akun, Alamat Rumah, Penghasilan Orang Tua, Cerita Berita Acara, Status Payment Token, PIN Price, Tanggal Aktivasi Akun, Tanggal Request PIN, Tanggal Bayar Pin, Tanggal Kelulusan, NPSN, Sumber Informasi.

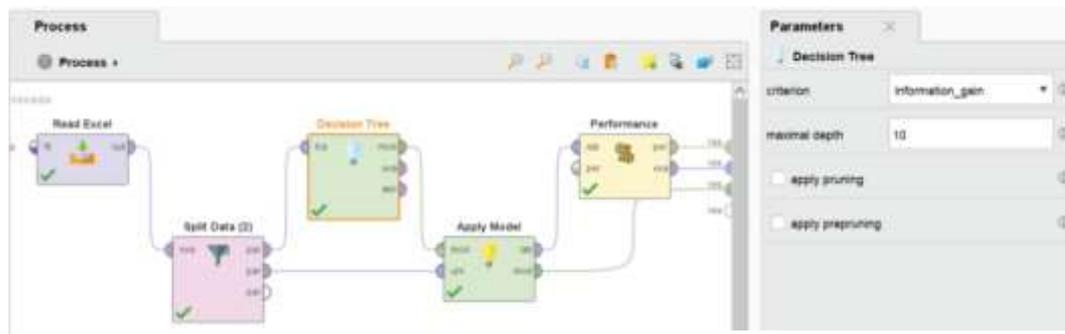
#### **3.1.2 Pengujian Dengan Algoritma *Decision Tree***

Setelah melewati proses pembersihan data maka menghasilkan dataset terbaru dari data awal sebanyak 27.606 record dimana data yang tidak lengkap dibuang sehingga yang tersisa sebanyak 3.766 record yang siap diolah dengan model klasifikasi menggunakan algoritma *Decision Tree*. Penelitian ini data dibagi menjadi 2(dua) bagian, dengan ratio 70% digunakan sebagai data training dan 30% untuk data testing dari 3.766 data.

Tabel 3.2. Perbandingan Hasil Ratio

| <b>Ratio Split Data</b> | <b>Accucary</b> |
|-------------------------|-----------------|
| 0.7 / 0.3               | 98.86 %         |
| 0.8 / 0.2               | 98.67 %         |
| 0.9 / 0.1               | 98.64 %         |

Berdasarkan tabel diatas nilai *Accuracy* tertinggi dari ratio 0.7 : 0.3 atau 70% banding 30% dengan nilai sebesar 98.86%. Setelah itu dilanjut ke dalam proses klasifikasi menggunakan algoritma *Decision Tree*.



Gambar 3.4 Model *Decision Tree*



Gambar 3.5 Hasil Model *Decision Tree*

Berdasarkan gambar diatas dapat dijelaskan bahwa node pertama merupakan Pilihan Prodi, Node kedua adalah sumber informasi, Node ketiga penghasilan orang tua, Node keempat adalah jenis Sekolah dan Node kelima adalah kategori sekolah. Jika pilihan prodi S1 Sistem Informasi, sumber informasi berasal dari tiktok, penghasilan orang tua golongan I Kategori sekolah Negeri maka status pembayaran UP3 “SUDAH BAYAR”, namun jika kategori sekolah swasta dengan jenis sekolah SMA maka status pembayaran “SUDAH BAYAR” tetapi masih ada yang “BELUM BAYAR”, Jika kategori sekolah SMK maka status pembayaran UP3 “SUDAH BAYAR”.

accuracy: 98.86%

|                   | true SUDAH BAYAR | true BELUM BAYAR | class precision |
|-------------------|------------------|------------------|-----------------|
| pred. SUDAH BAYAR | 2592             | 29               | 98.89%          |
| pred. BELUM BAYAR | 1                | 14               | 93.33%          |
| class recall      | 99.96%           | 32.56%           |                 |

Gambar 3.6 Nilai *Accuracy*

Proses algoritma Decision Tree menghasilkan nilai *accuracy* sebesar 98,86% .Hasil pengujian dengan nilai *accuracy* 98,86% mendapatkan prediksi sebagai berikut :

- a). Prediksi SUDAH BAYAR – True SUDAH BAYAR (TP) = 2.592
- b). Prediksi SUDAH BAYAR – True BELUM BAYAR (TN) = 29
- c). Prediksi BELUM BAYAR – True SUDAH BAYAR (FP) = 1
- d). Prediksi BELUM BAYAR – True BELUM BAYAR (FN) = 14

precision: 93.33% (positive class: BELUM BAYAR )

|                   | true SUDAH BAYAR | true BELUM BAYAR | class precision |
|-------------------|------------------|------------------|-----------------|
| pred. SUDAH BAYAR | 2592             | 29               | 98.89%          |
| pred. BELUM BAYAR | 1                | 14               | 93.33%          |
| class recall      | 99.96%           | 32.56%           |                 |

Gambar 3.7 Nilai Precision

Nilai Presisi untuk prediksi label status pembayaran UP3 sebesar 93,33%.

recall: 32.56% (positive class: BELUM BAYAR )

|                   | true SUDAH BAYAR | true BELUM BAYAR | class precision |
|-------------------|------------------|------------------|-----------------|
| pred. SUDAH BAYAR | 2592             | 29               | 98.89%          |
| pred. BELUM BAYAR | 1                | 14               | 93.33%          |
| class recall      | 99.96%           | 32.56%           |                 |

Gambar 3.8 Nilai Recall

Nilai Recall untuk prediksi label status pembayaran UP3 sebesar 32,56 %.



Gambar 3.9 Nilai AUC

Untuk nilai AUC *Decision Tree* berdasarkan tahap *modelling* yang telah dilakukan adalah sebesar 0.992.

#### 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan pengolahan data mahasiswa dapat diambil kesimpulan bahwa klasterisasi menggunakan algoritma *K-Means* menghasilkan 3 cluster yang optimal, yaitu *cluster\_1* dengan jumlah 91 mahasiswa. Program studi yang paling dominan adalah Sistem Informasi, dominan asal sekolah SMAN dengan jurusan MIPA, Jalur seleksi dominan jalur raport dan kota asal pada wilayah Kawasan Indonesia Timur. *Cluster\_2* dengan jumlah 1683 mahasiswa. Program studi yang dominan adalah Sistem Informasi, dominan asal sekolah SMAN dengan jurusan MIPA, Jalur seleksi dominan jalur raport dan kota asal yang dominan berasal dari Kota Surabaya, Kab. Sidoarjo, dan Kab. Gresik. *Cluster\_3* dengan jumlah 218 mahasiswa. Program Studi yang dominan adalah Informatika, paling dominan adalah Informatika dengan dominan asal sekolah SMAN dengan jurusan MIPA, Jalur seleksi dominan jalur raport dan kota asal yang dominan berasal dari Kota Jayapura, Kota Kupang, dan Kota Madiun. Dengan nilai Davies\_Bouldin sebesar 0,477. Rekomendasi Strategi promosi yaitu dengan, Melakukan promosi yang lebih banyak dan gencar terhadap daerah yang masih sedikit jumlah mahasiswanya. Melakukan promosi dengan memperbanyak jumlah sekolah potensial di daerah Surabaya dan Sidoarjo. Melakukan promosi ke sekolah yang linier dengan program studi di Institut Teknologi Telkom Surabaya. Melakukan promosi dengan periklanan melalui media social pada daerah yang jauh dijangkau.

Pengolahan data dengan model klasifikasi menggunakan algoritma *Decision Tree* untuk prediksi calon mahasiswa yang membayar UP3. dengan *mode test percentage split* yaitu 70% sebanyak 2.637 sebagai data *training* dan 30% Data set yang digunakan sebanyak 3.771 *record* sebanyak 1.113 sebagai data *testing*. Pengolahan data menggunakan *Decision Tree* menghasilkan nilai *accuracy* pengujian sebesar 98,86%, nilai *precision* sebesar 93,33%, nilai *recall* sebesar 32,56% dan AUC sebesar 0.992. *Decision Tree* menghasilkan sebuah pohon keputusan untuk memudahkan pengambilan keputusan strategi yang tepat. Rekomendasi strategi promosi yaitu dengan, Menjalin kerjasama dengan sekolah yang memiliki program jurusan yang linear dengan program studi di Institut Teknologi Telkom Surabaya, Meningkatkan jumlah sekolah SMA khususnya kategori Swasta di wilayah Jawa Timur, Meningkatkan kualitas konten yang diunggah ke media sosial, Konsisten dalam mengunggah konten dan memasang iklan berlangganan di semua platform media sosial, Membuat program Member Get Member untuk keluarga / Teman untuk menginformasikan kepada mahasiswa/keluarganya, apabila membawa temanya untuk masuk kuliah di Institut Teknologi Telkom Surabaya akan mendapatkan intensif dari pihak Admisi, Menerapkan biaya Pendidikan yang dapat dijangkau oleh Masyarakat ekonomi menengah ke bawah. Memberikan potongan tambahan untuk calon mahasiswa dari penghasilan orang tua golongan I. Memberikan banyak pilihan metode pembayaran biaya kuliah yang dapat dicicil oleh sesuai dengan kemampuan calon mahasiswa.

#### Ucapan Terima Kasih (Acknowledgement)

Penulis juga menyadari penulisan ini tidak bisa terselesaikan tanpa pihak-pihak yang mendukung baik secara moril dan juga materil. Maka penulis menyampaikan dengan penuh hormat mengucapkan terimakasih kepada:

1. Kedua orang tua terkasih yang telah memberikan segala doa dan dukungan yang terbaik.
2. Ibu Rokhmatul Insani, S.T., M.T selaku dosen pembimbing I atas bimbingan, masukan waktu, dan tenaga yang diberikan dan selalu sabar membimbing penulis hingga penyelesaian penulisan ini.
3. Bapak Mochammad Nizar Palefi Ma'ady, S.Kom., M.Kom., M.IM., selaku dosen pembimbing II yang memberikan arahan, masukan, waktu dan tenaga yang diberikan selama bimbingan dan selalu sabar membimbing penulis hingga penyelesaian penulisan ini.
4. Teman-teman seperjuangan yang selalu memberikan semangat, dukungan, membantu dan mengisi hari-hari penulis selama masa perkuliahan menjadi menyenangkan.
5. Penulis berterimakasih kepada diri sendiri yang telah berjuang baik secara fisik, fikiran dan mental

yang mampu bertahan sejauh ini dalam menjalani hari-hari yang berat dalam kehidupan yang keras ini sehingga dapat menyelesaikan tugas akhir, meskipun ini bukan akhir dari segalanya justru malah akan menjadi awal dari segalanya, penulis akan tetap melanjutkan perjuangan perjalanan hidup hingga merajut kisah-kisah yang penuh makna.

#### **Daftar Pustaka**

- Abriyanto, A., & Damastuti, N. (2019). Segmentasi Mahasiswa Dengan “Unsupervised” Algoritma Guna Membangun Strategi Marketing Penerimaan Mahasiswa. *Jurnal Insand Comtech*, 4(2).
- Inda Sari, P., & Nasir, M. (n.d.). Implementasi Model Decision Tree Menggunakan Algoritma ID3 Untuk Menetapkan Strategi Pemasaran Universitas Bina Darma Palembang. *Bina Darma Conference on Computer Science*.
- Kurniawan, D., & Yasir, D. M. (n.d.). *Optimization Sentiment Analysis Using CRISP-DM And Naive Bayes Methods Implemented On Social Media*.
- Kusumawardana, D. (n.d.). *Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*.
- Mirza, A. H. (2019). Penerapan Algoritma Naive Bayes Classifier Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru. *Journal of Information Systems and Informatics*, 1(1). <http://journal-isi.org/index.php/isi>
- Niqotaini, Z. (2021). *Perbandingan Algoritma Klasifikasi Data Mining Untuk Penelusuran Minat Calon Mahasiswa Baru*. 15. <https://journal.uniku.ac.id/index.php/ilkom>
- Oktaviarna Tensao, O., Nyoman Yudi Anggara Wijaya, I., Queena Fredlina, K., Studi Teknik Informatika, P., Primakara, S., Tukad Badung No, J., & Denpasar, R. (n.d.-a). *Analisa Data Mining Dengan Algoritma K-Means Clustering untuk Menentukan Strategi Promosi Mahasiswa Baru STMIK Primakara*.
- Oktaviarna Tensao, O., Nyoman Yudi Anggara Wijaya, I., Queena Fredlina, K., Studi Teknik Informatika, P., Primakara, S., Tukad Badung No, J., & Denpasar, R. (n.d.-b). *INFORMASI (Jurnal Informatika dan Sistem Informasi) Analisa Data Mining dengan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Pada STMIK Primakara*.
- Sri, N., Zulfa, L., & Hadiana, A. (2021). *Kajian Data Mining Menggunakan Algoritma K-Means Dan K-Medoids Dalam Strategi Promosi (Studi Kasus: Universitas Islam Al-Ihya Kuningan)* (Vol. 2, Issue 2).