PREDIKSI SELEKSI PENERIMA BEASISWA KIP KULIAH MENGGUNAKAN ALGORITMA *NAÏVE BAYES* (STUDI KASUS POLITEKNIK TEDC BANDUNG)

Muhammad Andry Juniawan, Ari Sudrajat
Teknik Informatika, Politeknik TEDC Bandung
Jl. Pesantren KM.2, Cimahi, Jawa Barat.
m.andry.juniawan@gmail.com

ABSTRAK

Pendidikan tinggi berperan penting dalam meningkatkan kualitas sumber daya manusia di Indonesia. Namun, keterbatasan ekonomi seringkali menghalangi lulusan SMA/SMK untuk melanjutkan pendidikan ke perguruan tinggi. Pemerintah Indonesia meluncurkan program Beasiswa KIP Kuliah sebagai bentuk bantuan pendidikan bagi siswa dari keluarga kurang mampu untuk mengakses pendidikan tinggi secara merata dan mendukung pembangunan sumber daya manusia unggul. Di Politeknik TEDC Bandung, proses seleksi penerima beasiswa KIP Kuliah saat ini masih dilakukan secara manual yang dinilai kurang efisien dan berisiko menimbulkan ketidakakuratan dalam penentuan kelayakan. Penelitian ini menggunakan algoritma *Naïve Bayes Classifier* untuk memprediksi status kelayakan penerima beasiswa KIP Kuliah dengan tujuan mendapatkan proses seleksi yang lebih cepat, akurat, dan objektif. Pengujian dilakukan dengan menggunakan teknik *Cross Validation* pada data pendaftar dari tahun 2020, 2021 dan 2024. Hasil penelitian dan pengujian Algoritma *Naïve Bayes* dalam memprediksi kelayakan calon mahasiswa beasiswa KIP Kuliah di Politeknik TEDC Bandung mendapatkan hasil berupa tingkat akurasi sebesar 76,39%, *precision* 25,00%, *recall* 5,54%, Dari total 322 data yang diuji, sebanyak 254 data berhasil diprediksi dengan benar, sementara 68 data diprediksi secara keliru.

Kata Kunci : *KIP Kuliah, Naïve Bayes, Klasifikasi*

1. PENDAHULUAN

Pendidikan tinggi memiliki peran penting dalam peningkatan kualitas sumber daya manusia di Indonesia. Namun, tidak semua siswa lulusan SMA/SMK mampu melanjutkan ke perguruan tinggi karena keterbatasan ekonomi. Untuk menjawab tantangan tersebut, pemerintah Indonesia meluncurkan program Beasiswa Kartu Indonesia Pintar Kuliah (KIP Kuliah). KIP Kuliah merupakan bentuk bantuan pendidikan bagi siswa dari keluarga kurang mampu agar tetap dapat melanjutkan pendidikan hingga jenjang perguruan tinggi [1]

Program KIP Kuliah bertujuan untuk mewujudkan pemerataan akses pendidikan tinggi sekaligus mendukung pembangunan sumber daya manusia yang unggul dan berdaya saing. Beasiswa ini mencakup biaya pendidikan dan bantuan biaya hidup. Di Politeknik TEDC Bandung, pendaftar jalur KIP Kuliah mengalami peningkatan setiap tahunnya. Namun demikian, proses seleksi penerima beasiswa ini masih dilakukan secara administratif dan manual, yang dinilai kurang efisien dan berisiko menimbulkan ketidaktepatan dalam penentuan kelayakan penerima. Selain itu pendaftar KIP Kuliah selalu melebihi kuota yang telah ditetapkan.

Proses seleksi yang akurat dan adil sangat diperlukan untuk memastikan bantuan tepat sasaran. Oleh karena itu, dibutuhkan sebuah sistem atau metode yang mampu membantu pihak kampus dalam melakukan seleksi awal berdasarkan data-data pendaftar. Metode tersebut harus mampu mengolah data dalam jumlah besar, mengenali pola, serta menghasilkan keputusan prediktif secara cepat dan tepat. Salah satu pendekatan yang dapat digunakan untuk keperluan ini adalah teknik *Data Mining*.

Data Mining adalah proses penggalian informasi atau pengetahuan tersembunyi dari kumpulan data besar untuk membantu pengambilan keputusan penting. *Data Mining* merupakan salah satu tahapan penting dalam proses *Knowledge Discovery in Database* (KDD), yaitu sebuah prosedur sistematis yang digunakan untuk mengekstraksi dan menemukan informasi berharga dari kumpulan data yang sangat besar. Melalui *data mining*, pola-pola tersembunyi dapat diidentifikasi untuk mendukung pengambilan keputusan yang lebih efektif dan efisien [2]. Dalam *data mining* terdapat berbagai metode seperti *classification*, *clustering*, *regression*, dan lainnya Salah satu metode populer dalam proses klasifikasi adalah *Naïve Bayes Classifier*.

Metode *Naïve Bayes Classifier* merupakan teknik klasifikasi yang dikenal memiliki tingkat akurasi yang sangat baik. Proses perhitungannya menggunakan Teorema *Bayes*, sehingga mampu berjalan dengan cepat dan sederhana, namun tetap memberikan hasil klasifikasi yang akurat. Keunggulan metode ini terletak pada kemampuannya dalam bekerja secara efisien meskipun dengan jumlah data yang besar dan kompleks. *Naïve Bayes* sering digunakan dalam berbagai kasus prediksi, seperti pengelompokan kelayakan penerima bantuan, klasifikasi email spam, dan lain sebagainya [3]

Beberapa penelitian terdahulu juga telah menunjukkan efektivitas algoritma ini. Cendikia Fitri Nuril Halizah dan Meivi Kartikasari menggunakan *Naïve Bayes* untuk Penentuan Pemberian Beasiswa KIP Kuliah dan mendapatkan akurasi 98,89% [4].

Angga Pebdika, dkk juga menerapkan metode yang sama dalam penentuan penerima Program Indonesia Pintar (KIP) Kuliah dengan hasil akurasi yang juga tinggi [5].

Berdasarkan hal tersebut, penelitian ini akan mengkaji bagaimana algoritma *Naïve Bayes* dapat digunakan untuk memprediksi status penerimaan beasiswa KIP Kuliah bagi calon mahasiswa di Politeknik TEDC Bandung. Harapannya, hasil dari penelitian ini dapat memberikan manfaat bagi institusi dalam mendukung proses seleksi yang lebih adil, akurat, dan efisien. Oleh karena itu, penulis menuangkan penelitian ini dalam bentuk Tugas Akhir dengan judul "Prediksi Seleksi Penerimaan Beasiswa Kip Kuliah Menggunakan Algoritma *Naïve Bayes* (Studi Kasus: Politeknik TEDC Bandung)".

Berdasarkan latar belakang di atas, terdapat beberapa permasalahan yang menjadi fokus dalam penelitian ini, yaitu: Bagaimana hasil nilai probabilitas untuk kelas "Layak" dan "Tidak Layak" dalam penentuan kelayakan penerima Beasiswa KIP Kuliah bagi calon mahasiswa di Politeknik TEDC Bandung, Bagaimana hasil akurasi algoritma *Naïve Bayes* dalam memprediksi data Beasiswa KIP Kuliah bagi calon mahasiswa di Politeknik TEDC Bandung?

Adapun tujuan dari penelitian ini adalah sebagai berikut: Mendapatkan nilai probabilitas untuk kelas "Layak" dan "Tidak Layak" dalam penentuan kelayakan penerima Beasiswa KIP Kuliah bagi calon mahasiswa Politeknik TEDC Bandung, Mengukur dan menganalisis tingkat akurasi algoritma *Naïve Bayes* dalam memprediksi data penerima beasiswa.

Untuk menyelesaikan permasalahan tersebut, langkah-langkah yang akan dilakukan dalam penelitian ini meliputi: Pengumpulan data penerima beasiswa KIP Kuliah dari Politeknik TEDC Bandung sebagai studi kasus, Melakukan proses *preprocessing* data agar layak untuk digunakan dalam proses klasifikasi, Menerapkan algoritma *Naïve Bayes* untuk membangun model prediksi kelayakan penerima beasiswa, Melakukan evaluasi dan pengujian model dengan menggunakan data uji untuk melihat tingkat akurasi dan performa model, Menarik kesimpulan dan memberikan rekomendasi berdasarkan hasil penelitian.

2. TINJAUAN PUSTAKA

2.1. KIP Kuliah

Beasiswa KIP Kuliah termasuk salah satu program beasiswa yang sangat diminati dengan tingkat penerimaan yang cukup tinggi. Program ini dirancang khusus sebagai bentuk layanan pendidikan yang bertujuan membantu mahasiswa kurang mampu untuk melanjutkan pendidikan ke jenjang yang telah mereka pilih. Tujuan utama dari Beasiswa KIP Kuliah adalah memberikan dukungan finansial kepada mahasiswa yang menghadapi keterbatasan ekonomi, sehingga mereka dapat memenuhi kebutuhan dasar dan menjalani masa studinya dengan baik. Dengan bantuan ini, mahasiswa diharapkan mampu mengatasi berbagai hambatan yang mungkin menghalangi kelangsungan pendidikan mereka. Selain itu, mahasiswa penerima KIP Kuliah idealnya menunjukkan sikap, penampilan, dan perilaku yang mencerminkan identitas serta nilai yang dipegang oleh penerima beasiswa tersebut dalam kehidupan sehari-hari [6]

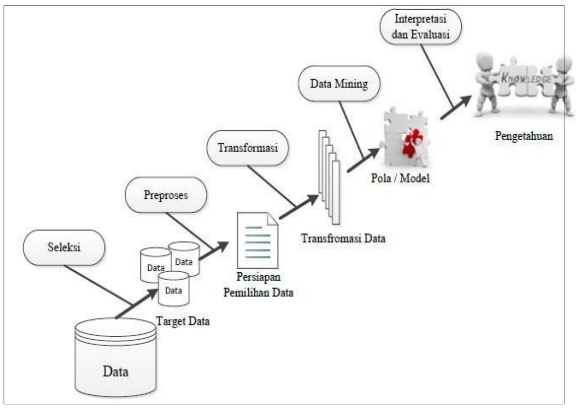
2.2. Beasiswa

Beasiswa merupakan bantuan finansial yang diberikan kepada individu, seperti pelajar atau mahasiswa, untuk mendukung kelangsungan pendidikan yang sedang ditempuh. Pemberian beasiswa bisa berasal dari lembaga pemerintah, perusahaan, atau yayasan. Beasiswa dapat diberikan secara cuma-cuma atau dengan adanya ikatan dinas, yaitu kewajiban kerja setelah menyelesaikan pendidikan, yang durasinya bervariasi tergantung pada instansi pemberi beasiswa tersebut. Selain itu, beasiswa juga sering diberikan kepada kelompok, misalnya sebagai hadiah dalam *event* perlombaan yang diselenggarakan oleh lembaga pendidikan [7].

2.3. *Data Mining*

Data mining merupakan proses menggali informasi penting dari kumpulan data. Informasi tersebut diperoleh melalui prosedur yang kompleks seperti penerapan kecerdasan buatan, metode statistik, matematika, *machine learning*, dan teknik lainnya. Proses ini akan mengenali serta mengambil informasi bernilai dari *database* besar. *Data mining* sebagai disiplin ilmu berkembang di bidang kecerdasan buatan (AI) dan rekayasa pengetahuan (KE). Ilmu ini berawal dari *machine learning* dan statistika, namun juga mencakup berbagai bidang ilmu komputer dan disiplin lain seperti biologi, lingkungan, keuangan, jaringan, dan lain-lain [8]

Dalam proses data mining yang biasa dikenal dengan Knowledge Discovery Database (KDD), terdapat proses terlihat pada Gambar 1.



Gambar 1. Proses *Knowledge Discovery Database*

Menurut Juna Eska dalam penelitiannya, Tahapan data mining dibagi menjadi enam bagian yaitu [9] :

a. Pembersihan Data (*Data Cleaning*)
Sebelum melakukan proses *data mining*, terlebih dahulu harus dilakukan pembersihan pada data yang akan menjadi fokus dalam KDD. Proses ini meliputi penghapusan data yang terduplikasi, pengecekan data yang tidak konsisten, dan perbaikan kesalahan seperti typo atau kesalahan penulisan. Selain itu, dilakukan juga *enrichment*, yaitu menambahkan informasi tambahan yang relevan dan berguna, termasuk data eksternal, untuk memperkaya data yang ada.
b. Integrasi Data (*Data Integration*)
Integrasi data adalah proses menggabungkan data dari berbagai sumber *database* menjadi satu *database* terpadu. Data yang diperlukan untuk *mining* seringkali berasal dari lebih dari satu *database* atau file teks. Penggabungan dilakukan dengan mengacu pada atribut kunci yang dapat mengidentifikasi entitas unik, seperti nama, jenis produk, atau nomor pelanggan. Proses ini harus dilakukan dengan hati-hati karena jika terjadi kesalahan, hasil analisis bisa salah atau menyesatkan. Contohnya, integrasi berdasarkan jenis produk yang salah dapat menciptakan korelasi palsu antar kategori produk yang berbeda.
c. Seleksi Data (*Data Selection*)
Tidak semua data dalam *database* digunakan dalam analisis, sehingga hanya data yang relevan dengan tujuan analisis yang diambil. Misalnya, dalam penelitian kecenderungan pembelian dalam *market basket analysis*, nama pelanggan tidak perlu dimasukkan, cukup menggunakan ID pelanggan saja.
d. Transformasi data (*Data Transformation*)
Data diubah atau dikombinasikan ke format yang sesuai untuk proses *data mining*. Beberapa metode *mining* memerlukan format khusus, seperti metode asosiasi dan *clustering* yang hanya menerima data kategorikal. Oleh karena itu, data *numerik kontinu* harus dibagi ke dalam *interval-interval* tertentu. Proses ini disebut transformasi data.
e. Proses *Mining*
Ini adalah tahap inti di mana metode yang dipilih diterapkan untuk menemukan pola, pengetahuan tersembunyi, dan informasi yang bernilai dari data yang sudah diproses.
f. Evaluasi Pola (*Pattern Evaluation*)
Pada tahap ini, pola-pola ataupun model prediksi yang diperoleh dari proses *mining* dievaluasi untuk menentukan apakah pola tersebut benar-benar menarik dan bermanfaat serta memenuhi hipotesis awal.
g. Presentasi Pengetahuan (*Knowledge Presentation*)
Tahap terakhir ini berfokus pada penyajian dan visualisasi pengetahuan yang diperoleh agar dapat dipahami oleh pengguna, termasuk yang bukan ahli data *mining*. Penyajian yang jelas dan visualisasi yang efektif membantu dalam mengkomunikasikan hasil sehingga dapat digunakan untuk pengambilan keputusan atau tindakan selanjutnya.

2.4. *Klasifikasi*

Klasifikasi merupakan salah satu teknik utama dalam *data mining* yang berfungsi untuk mengelompokkan data ke dalam kelas tertentu berdasarkan pola yang ditemukan dari data berlabel (*training data*). Proses klasifikasi menggunakan pendekatan *supervised learning*, dimana model dilatih dengan data yang sudah memiliki kelas target untuk mempelajari hubungan antar atribut input dan kelas tersebut. Model klasifikasi kemudian digunakan untuk memprediksi kelas dari data baru yang belum memiliki label, dengan tujuan menghasilkan prediksi yang akurat [10]

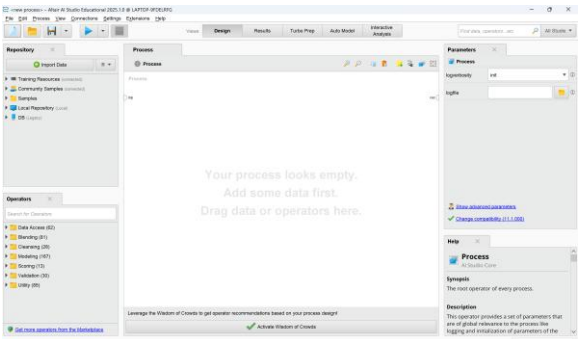
2.5. Algoritma *Naïve Bayes*

Naive Bayes adalah metode klasifikasi statistik yang bekerja dengan prinsip *probabilitas Bayes* dan asumsi bahwa setiap atribut dalam data bersifat independen atau tidak saling bergantung satu sama lain. Proses kerja algoritma ini dimulai dengan penggunaan data pelatihan untuk menghitung probabilitas dari setiap fitur dalam kaitannya dengan kelas yang ada. Kemudian, untuk data uji, *probabilitas* tiap fitur dikalikan bersama dengan *probabilitas* kelas untuk menentukan kelas dengan *probabilitas* tertinggi sebagai hasil klasifikasi [11].

2.6. *RapidMiner*

RapidMiner adalah sebuah platform perangkat lunak data science yang menyediakan berbagai alat untuk analisis data, mulai dari persiapan data, transformasi, visualisasi hingga pemodelan dan *deployment model machine learning*. *RapidMiner* dikembangkan menggunakan bahasa pemrograman *Java* dan memiliki tampilan antarmuka grafis (GUI) yang intuitif, sehingga pengguna dapat merancang proses *data mining* secara *drag-and-drop* tanpa perlu menulis kode pemrograman khusus [12]

Adapun Tampilan *RapidMiner* terdapat pada dibawah ini.



Gambar 2. Tampilan Beranda *RapidMiner*

2.7. *Cross Validation*

Cross Validation adalah suatu teknik dalam data mining yang digunakan untuk mendapatkan tingkat

akurasi terbaik dengan membagi data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Salah satu jenis *Cross Validation* adalah *K-Fold Cross Validation*, yang berfungsi untuk mengevaluasi kinerja suatu algoritma dengan cara membagi data secara acak menjadi K kelompok. Pada metode *K-Fold Cross Validation*, dataset dibagi menjadi beberapa bagian secara acak, kemudian proses pelatihan dan pengujian dilakukan sebanyak K kali, di mana pada setiap eksperimen, satu partisi digunakan sebagai data pengujian sementara partisi lainnya digunakan sebagai data pelatihan [13]

2.8. Dataset

Dataset merupakan kumpulan data yang tersimpan dalam media digital (memori), dimana setiap data memiliki nilai tertentu. Setiap nilai tersebut berkaitan dengan variabel yang berbeda dan setiap variabel memiliki fungsi atau tujuan sesuai dengan jenis data yang dikumpulkan. Dataset ini bisa berbentuk dokumen atau file [14]

2.9. Penelitian Terkait

Terdapat beberapa penelitian yang telah dilakukan oleh banyak peneliti yang berkaitan dengan latar belakang masalah penelitian dan tugas akhir ini, seperti yang dijelaskan di bawah ini:

a. Penelitian yang dilakukan oleh Imamsyah Priyanto, dkk pada tahun 2024 yang berjudul "Penerapan Algoritma Metode *Naïve Bayes* untuk Penentuan Penerimaan Bantuan Program Indonesia Pintar (PIP) (Studi Kasus SMP PGRI 1 Cilacap)". Penelitian ini mengkaji penerapan metode *data mining Naïve Bayes* untuk memprediksi kelayakan penerima bantuan Program Indonesia Pintar (PIP) di SMP PGRI 1 Cilacap. Hasil menunjukkan metode *Naïve Bayes* mampu memprediksi penerima dengan akurasi 88,89% dan recall 85,71%, lebih akurat dibandingkan penentuan manual oleh sekolah [15]
b. Penelitian yang dilakukan oleh Irma Purnama Oktavia dan Novita Lestari pada tahun 2024 yang berjudul "Implementasi Algoritma *Naïve Bayes* dengan Metode Klasifikasi dalam Menentukan Siswa Penerima Bantuan Program Indonesia Pintar (Studi Kasus: SMPN 3 Cihampelas)". Penelitian ini mengkaji penggunaan algoritma *Naïve Bayes* untuk mengklasifikasikan kelayakan siswa penerima bantuan Program Indonesia Pintar (PIP) di SMPN 3 Cihampelas. Pengujian dengan *RapidMiner* dan teknik *Cross Validation* menghasilkan akurasi 97,24%, *precision* 72,73%, *recall* 88,89%, dan AUC 0,886, menunjukkan klasifikasi yang baik dan efektivitas metode dalam mendukung keputusan penerima bantuan di sekolah tersebut [16]
c. Penelitian yang dilakukan oleh Nurhidayati, dkk pada tahun 2023 yang berjudul "Implementasi Algoritma *Naive Bayes* Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi)". Penelitian ini menerapkan algoritma *Naive Bayes* untuk mengklasifikasikan calon penerima beasiswa Bidikmisi di Universitas Hamzanwadi. Pengujian dengan *k-fold cross validation* menghasilkan akurasi tertinggi 91,43% dan AUC 0,996, menunjukkan klasifikasi yang sangat baik dan valid sebagai dasar pengambilan keputusan beasiswa. Sistem ini diharapkan meningkatkan efektivitas dan efisiensi seleksi di universitas tersebut [17]
d. Penelitian yang dilakukan oleh Cendikia Fitri Nuril Halizah dan Meivi Kartikasari pada tahun 2024 yang berjudul "Implementasi Algoritma *Naïve Bayes* dalam Penentuan Pemberian Beasiswa KIP Kuliah (Studi Kasus STIKI Malang)". Penelitian ini menerapkan algoritma *Naïve Bayes* untuk mengklasifikasikan kelayakan calon penerima beasiswa KIP Kuliah di STIKI Malang. Pengujian pada 100 data calon penerima menghasilkan akurasi sebesar 98,89%, menunjukkan efektivitas algoritma ini dalam mendukung pengambilan keputusan pemberian beasiswa sesuai kriteria yang ditetapkan [4]
e. Penelitian yang dilakukan oleh Angga Pebdika, dkk pada tahun 2023 yang berjudul "Klasifikasi Menggunakan Metode *Naïve Bayes* untuk Menentukan Calon Penerima PIP" membahas tentang penerapan algoritma *Naïve Bayes* dalam proses klasifikasi tingkat kelayakan calon penerima Program Indonesia Pintar (PIP). Hasil pengujian menunjukkan akurasi sebesar 88,89%, *precision* 90%, *recall* 97,67%, dan nilai AUC 0,807, yang membuktikan bahwa metode ini efektif dalam mendukung proses seleksi penerima bantuan secara lebih objektif, tepat sasaran, dan efisien dibandingkan metode manual [5]

3. METODE PENELITIAN

3.1. Teknik Pengumpulan Data

Metode pengumpulan data pada penelitian ini meliputi:

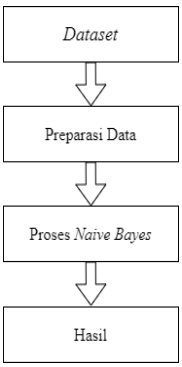
a. Data Primer
   Data primer dalam penelitian ini merupakan data yang diperoleh secara langsung dari sumber utama, yaitu Politeknik TEDC Bandung. Data tersebut berupa informasi pendaftar beasiswa KIP Kuliah pada rentang tahun 2020, 2021 dan 2024. Pengumpulan data primer dilakukan melalui dokumentasi administratif yang tersedia di bagian kemahasiswaan Politeknik TEDC Bandung.
b. Data Sekunder
   Data sekunder pada penelitian ini diperoleh dari berbagai referensi ilmiah dan sumber pendukung lain yang relevan. Sumber-sumber tersebut meliputi jurnal, artikel, buku ilmiah, serta sumber daring terpercaya yang membahas topik *data mining*, algoritma *Naïve Bayes*, seleksi beasiswa, dan studi tentang KIP Kuliah. Data sekunder ini berfungsi sebagai landasan teori, rujukan

metodologi, serta bahan perbandingan hasil penelitian.

c. Observasi
Observasi dilakukan dengan mengamati secara langsung proses administrasi dan pemilahan data pendaftar KIP Kuliah di Politeknik TEDC Bandung. Langkah ini meliputi pemantauan tahapan seleksi manual, pengumpulan dokumen, serta wawancara singkat dengan staf bagian kemahasiswaan terkait prosedur, kendala, dan kebutuhan pengembangan sistem prediksi berbasis data mining.

3.2. Metode yang Digunakan

Penelitian ini menggunakan metode klasifikasi dengan algoritma *Naïve Bayes*.



Gambar 3. Metode yang digunakan

Gambar diatas merupakan alur proses pengolahan data pada penelitian ini dengan penjelasan sebagai berikut dibawah ini

a. Dataset
Pengumpulan dataset yang berisikan data mahasiswa calon penerima beasiswa KIP Kuliah untuk dianalisis.

b. Preparasi Data
Preparasi data adalah membersihkan dan menyiapkan data agar siap digunakan dalam proses analisis atau pemodelan, dengan menghilangkan data yang tidak relevan atau rusak serta mengubah format data supaya konsisten dan mudah diproses. Adapun preparasi data diproses setelah diperoleh data secara lengkap.

c. Metode *Naïve Bayes*
Pada tahap ini, algoritma *Naive Bayes Classifier* digunakan untuk mengelompokkan calon mahasiswa penerima KIP Kuliah. Proses pengujian dilakukan dengan menggunakan teknik *Cross Validation* agar hasilnya lebih akurat dan dapat dipertanggungjawabkan.

d. Hasil
Tahap terakhir meliputi evaluasi terhadap model yang sudah dibuat, yang kemudian menghasilkan klasifikasi data calon mahasiswa penerima KIP Kuliah. Dari hasil klasifikasi tersebut, kesimpulan akan diambil.

3.3. Analisis

Analisis dilakukan dengan mengamati dan menginterpretasikan hasil klasifikasi yang diperoleh dari proses pemodelan menggunakan *RapidMiner*. Proses analisis meliputi penghitungan nilai probabilitas untuk setiap kelas ("Layak" dan "Tidak Layak") serta evaluasi performa model melalui pengukuran akurasi. Data diuji menggunakan teknik validasi seperti *cross validation* untuk memastikan kestabilan model. Hasil probabilitas membantu memahami pola data yang menentukan kelayakan beasiswa, sedangkan nilai akurasi digunakan untuk menilai keandalan prediksi algoritma *Naïve Bayes*.

4. HASIL DAN PEMBAHASAN

4.1. Dataset

Penelitian ini memanfaatkan dataset yang dari calon mahasiswa penerima KIP Kuliah pada tahun 2020, 2021 dan 2024 yang didapat dari Kemahasiswaan Politeknik TEDC Bandung. Dataset tersebut akan diproses terlebih dahulu agar siap digunakan dalam analisis. Data tersebut akan di preparasi dan di transformasi sebelum dilakukan pengujian dengan model klasifikasi *Naïve Bayes* di *RapidMiner*.

4.2. Preparasi dan Transformasi Data

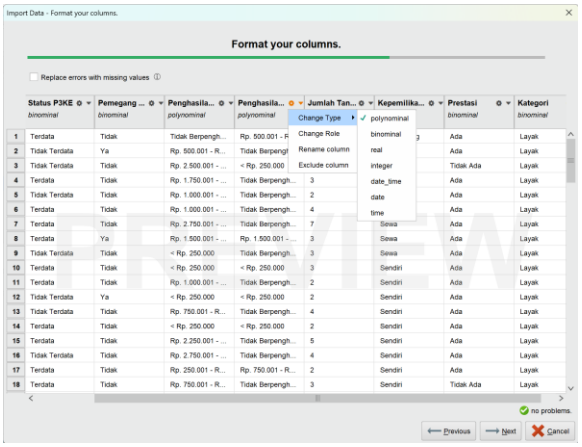
Atribut yang digunakan dalam klasifikasi ini awalnya berjumlah 42 Atribut. Setelah di preparasi menjadi 9 Atribut yakni Status DTKS, Status P3KE, Pemegang KIP, Penghasilan Ayah, Penghasilan Ibu, Jumlah Tanggungan, Kepemilikan Rumah, Prestasi dan juga Kategori yang dimanfaatkan sebagai label. Setelah data di preparasi maka data akan ditransformasi agar sesuai dengan analisis *data mining* yang diharapkan dalam penelitian ini.

	A	B	C	D	E	F	G	H	I
1	Status DTKS	Status P3KE	Pemegang KIP	Penghasilan Ayah	Penghasilan Ibu	Jumlah Tanggungan	Kepemilikan Rumah	Prestasi	Kategori
2	Tidak Terdata	Terdata	Tidak	Tidak Berpenghasilan	Rp. 500.001 - Rp. 750.000	3	Menumpang	Ada	Layak
3	Terdata	Tidak Terdata	Ya	Rp. 500.001 - Rp. 750.000	Tidak Berpenghasilan	2	Sendiri	Ada	Layak
4	Tidak Terdata	Tidak Terdata	Tidak	Rp. 2.500.001 - Rp. 2.750.000	< Rp. 250.000	2	Sewa	Tidak Ada	Layak
5	Tidak Terdata	Terdata	Tidak	Rp. 1.750.001 - Rp. 2.000.000	Tidak Berpenghasilan	3	Sendiri	Ada	Layak
6	Tidak Terdata	Tidak Terdata	Tidak	Rp. 1.000.001 - Rp. 1.250.000	Tidak Berpenghasilan	2	Sendiri	Ada	Layak
7	Tidak Terdata	Terdata	Tidak	Rp. 1.000.001 - Rp. 1.250.000	Tidak Berpenghasilan	4	Sendiri	Ada	Layak
8	Tidak Terdata	Terdata	Tidak	Rp. 2.750.001 - Rp. 3.000.000	Tidak Berpenghasilan	7	Sewa	Ada	Layak
9	Terdata	Terdata	Ya	Rp. 1.500.001 - Rp. 1.750.000	Rp. 1.500.001 - Rp. 1.750.000	3	Sewa	Ada	Layak
10	Tidak Terdata	Tidak Terdata	Tidak	< Rp. 250.000	Tidak Berpenghasilan	3	Sewa	Ada	Layak
11	Terdata	Terdata	Tidak	< Rp. 250.000	< Rp. 250.000	3	Sendiri	Ada	Layak
12	Terdata	Terdata	Tidak	Rp. 1.000.001 - Rp. 1.250.000	Tidak Berpenghasilan	2	Sendiri	Ada	Layak
13	Tidak Terdata	Tidak Terdata	Ya	< Rp. 250.000	< Rp. 250.000	2	Sendiri	Ada	Layak
14	Tidak Terdata	Tidak Terdata	Tidak	Rp. 750.001 - Rp. 1.000.000	Tidak Berpenghasilan	4	Sendiri	Ada	Layak
15	Tidak Terdata	Terdata	Tidak	< Rp. 250.000	< Rp. 250.000	2	Sendiri	Ada	Layak
16	Tidak Terdata	Terdata	Tidak	Rp. 2.250.001 - Rp. 2.500.000	Tidak Berpenghasilan	5	Sendiri	Ada	Layak
17	Tidak Terdata	Tidak Terdata	Tidak	Rp. 2.750.001 - Rp. 3.000.000	Tidak Berpenghasilan	4	Sendiri	Ada	Layak
18	Tidak Terdata	Terdata	Tidak	Rp. 250.001 - Rp. 500.000	Rp. 750.001 - Rp. 1.000.000	2	Sendiri	Ada	Layak
19	Terdata	Terdata	Tidak	Rp. 750.001 - Rp. 1.000.000	Tidak Berpenghasilan	3	Sendiri	Tidak Ada	Layak
20	Terdata	Terdata	Tidak	Rp. 2.000.001 - Rp. 2.250.000	Tidak Berpenghasilan	1	Sendiri	Ada	Layak
21	Tidak Terdata	Tidak Terdata	Tidak	Rp. 750.001 - Rp. 1.000.000	Tidak Berpenghasilan	2	Sewa	Ada	Layak
22	Tidak Terdata	Terdata	Tidak	Rp. 750.001 - Rp. 1.000.000	Tidak Berpenghasilan	1	Sendiri	Ada	Layak
23	Terdata	Tidak Terdata	Tidak	Tidak Berpenghasilan	Tidak Berpenghasilan	2	Sendiri	Ada	Layak
24	Tidak Terdata	Tidak Terdata	Tidak	Rp. 1.500.001 - Rp. 1.750.000	Tidak Berpenghasilan	3	Menumpang	Ada	Layak
25	Tidak Terdata	Terdata	Tidak	Rp. 500.001 - Rp. 750.000	Tidak Berpenghasilan	2	Sendiri	Ada	Layak
26	Tidak Terdata	Tidak Terdata	Tidak	Rp. 1.250.001 - Rp. 1.500.000	Rp. 2.250.001 - Rp. 2.500.000	3	Menumpang	Ada	Layak
27	Tidak Terdata	Terdata	Tidak	Rp. 3.750.001 - Rp. 4.000.000	Tidak Berpenghasilan	3	Sewa	Ada	Layak
28	Tidak Terdata	Terdata	Ya	Rp. 2.000.001 - Rp. 2.250.000	Tidak Berpenghasilan	6	Menumpang	Ada	Layak
29	Tidak Terdata	Terdata	Tidak	Rp. 2.250.001 - Rp. 2.500.000	Tidak Berpenghasilan	2	Sendiri	Ada	Layak
30	Tidak Terdata	Terdata	Tidak	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan	2	Menumpang	Ada	Layak
31	Tidak Terdata	Tidak Terdata	Tidak	Rp. 3.000.001 - Rp. 3.250.000	Rp. 1.250.001 - Rp. 1.500.000	1	Menumpang	Ada	Layak
32	Tidak Terdata	Tidak Terdata	Tidak	Rp. 3.250.001 - Rp. 3.500.000	Tidak Berpenghasilan	1	Sendiri	Ada	Layak
33	Tidak Terdata	Terdata	Tidak	Rp. 2.750.001 - Rp. 3.000.000	Tidak Berpenghasilan	3	Sewa	Ada	Layak

Gambar 4. Hasil Preparasi dan Transformasi data

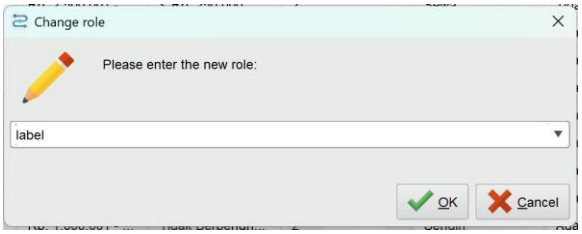
4.3. Proses Pengujian *Naïve Bayes* di *RapidMiner*

Pada tahap ini merupakan proses awal pengujian data pada *RapidMiner* yang mana dimulai dengan melakukan *import* data yang akan diuji dan mengubah type menjadi binomial terhadap kolom yang berisi 2 karakter, *integer* untuk kolom yang berisi nomor saja. Seperti pada gambar berikut ini.



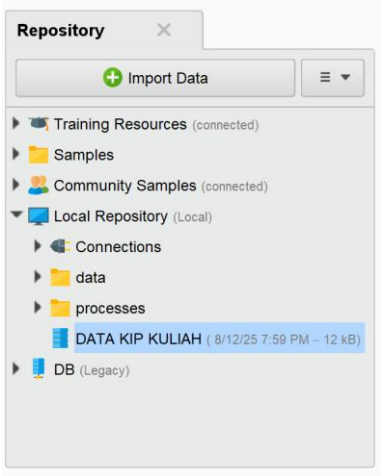
Gambar 5. Proses *Import* data dan *Change Type*

Selanjutnya mengubah role pada kolom Kategori menjadi label. Atribut ini berperan sebagai variable target atau atribut target untuk operator pembelajaran.



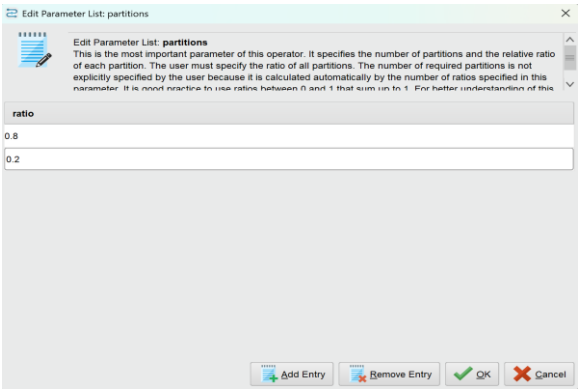
Gambar 6. Mengubah *Role* label

Setelah data berhasil ditambah maka akan tampil seperti gambar dibawah ini. Dalam hal ini data yang saya masukkan adalah “DATA KIP KULIAH”.



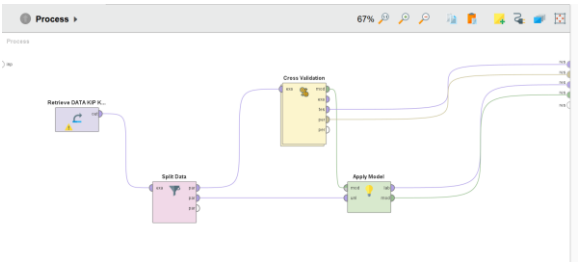
Gambar 7. Berhasil *Import* data

Adapun setelah berhasil *Import* data maka dapat menambah operator *Retrive* lalu menambahkan operator *split* data.



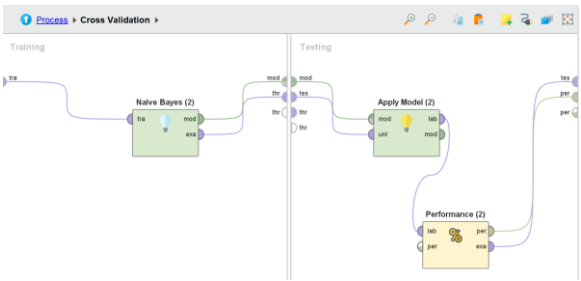
Gambar 8. *Split* data

Gambar diatas merupakan *Split* data untuk membagi data perbandingan yang dalam hal ini saya membagi 80% untuk *training* dan 20% untuk *testing.*



Gambar 9. Proses Pengujian

Gambar 9 menunjukkan proses pengujian awal yang melibatkan penggunaan operator *Cross Validation* pada data pelatihan untuk melatih dan mengevaluasi model, serta operator *Apply Model* untuk menerapkan model akhir yang sudah dilatih ke data pengujian yang terpisah. Dalam *Cross Validation*, data dibagi secara acak menjadi 10 bagian atau *fold*. Model dilatih menggunakan 9 *fold*, sementara *fold* ke-10 digunakan sebagai data uji. Proses ini diulang sebanyak 10 kali sehingga setiap *fold* bergantian menjadi data uji, sedangkan sisanya digunakan untuk pelatihan. Setelah itu, tingkat kesalahan (*error rate*) dihitung pada data uji tersebut. Rata-rata *error rate* dari 10 pengujian ini memberikan gambaran mengenai kemampuan model secara keseluruhan dalam melakukan prediksi. Metode *Cross Validation* ini bertujuan untuk memastikan bahwa model tidak mengalami *overfitting*, sehingga performanya tetap baik tidak hanya pada data pelatihan tetapi juga pada data baru yang belum pernah dilihat sebelumnya.



Gambar 10. Pengujian *Naïve Bayes*

Pada gambar 10 tampak penggunaan metode *Cross Validation* dalam pelatihan model yang terdiri dari tiga komponen utama. Pertama adalah operator *Naive Bayes* yang bertugas untuk melatih model menggunakan data pelatihan. Kedua, operator *Apply Model* digunakan untuk menerapkan model hasil pelatihan tersebut ke data pengujian yang terpisah. Terakhir, operator *Performance* berfungsi menampilkan *Confusion Matrix* yang menyajikan evaluasi hasil model, termasuk metrik seperti *Accuracy, Precision, Recall* untuk menilai kinerja prediksi model secara keseluruhan.

4.4. Hasil Pengujian

Pada pengujian dengan 322 data *training*, terdapat 242 data yang diprediksi "Layak" dan memang benar "Layak" (*True Positive*). Ada 64 data yang diprediksi "Layak" namun sebenarnya "Tidak Layak" (*False Positive*). Adapun untuk prediksi Tidak Layak ada Sebanyak 12 data diprediksi "Tidak Layak" tetapi sebenarnya "Layak" (*False Negative*). Sedangkan 4 data diprediksi "Tidak Layak" sesuai kenyataan (*True Negative*).

	true Layak	true Tidak Layak	class precision
pred. Layak	242	64	79.08%
pred. Tidak Layak	12	4	25.00%
class recall	95.28%	5.88%	

Gambar 11. *Confusion Matrix*

Adapun hasil *Perfomance Vector* pada *RapidMiner* dari data tersebut adalah sebagai berikut dibawah ini

4.5. *Accuracy*

accuracy: 76.39% +/- 3.38% (micro average: 76.40%)

	true Layak	true Tidak Layak	class precision
pred. Layak	242	64	79.08%
pred. Tidak Layak	12	4	25.00%
class recall	95.28%	5.88%	

Gambar 12. Hasil *Accuracy*

Sebagaimana hasil *Accuracy* pada gambar diatas menunjukkan bahwa akurasi algoritama *Naïve Bayes* adalah sebesar 76,39%

4.6. *Precision*

precision: 25.00% (positive class: Tidak Layak)

	true Layak	true Tidak Layak	class precision
pred. Layak	242	64	79.08%
pred. Tidak Layak	12	4	25.00%
class recall	95.28%	5.88%	

Gambar 13. Hasil *Precision*

Pada gambar diatas menunjukkan hasil *class precision* adalah 79,08% untuk *class* "Layak" dan 25% untuk *class* "Tidak Layak

4.7. *Recall*

Pada gambar 14 menunjukkan hasil *class Recall* adalah 95,28% untuk *class* "Layak" dan 5,88% untuk *class* "Tidak Layak.

recall: 5.54% +/- 7.17% (micro average: 5.88%) (positive class: Tidak Layak)

	true Layak	true Tidak Layak	class precision
pred. Layak	242	64	79.08%
pred. Tidak Layak	12	4	25.00%
class recall	95.28%	5.88%	

Gambar 14. Hasil *Recall*

Dari hasil pengujian yang dilakukan dengan menggunakan 322 Data Latih dan 80 data Uji mendapatkan hasil kesalahan prediksi sebanyak 64 untuk kelas "Layak" yang mana saat di prediksi dengan *RapidMiner* menjadi Tidak Layak dan 4 data pada kelas "Tidak Layak" saat diprediksi menggunakan *RapidMiner* hasilnya Layak. Secara keseluruhan dari total 322 data yang diuji, sebanyak 254 data berhasil diprediksi dengan benar, sementara 68 data diprediksi secara keliru.

5. KESIMPULAN DAN SARAN

Hasil penelitian dan pengujian Algoritma *Naïve Bayes* dalam memprediksi kelayakan calon mahasiswa beasiswa KIP Kuliah, mendapatkan hasil berupa tingkat akurasi sebesar 76,39%, *precision* 25,00%, *recall* 5,54%, Dari total 322 data yang diuji, sebanyak 254 data berhasil diprediksi dengan benar, sementara 68 data diprediksi secara keliru. Peneliti dapat menyimpulkan bahwa tingkat akurasi algoritna *Naïve Bayes* dipengaruhi oleh dataset dikarenakan pada penelitian sebelumnya yang dilakukan oleh Angga Pebdika [5], tingkat akurasi dengan menggunakan algoritma dan topik yang sama menghasilkan akurasi yang lebih dari 90%.

Penelitian selanjutnya disarankan dapat dilakukan perbandingan dengan algoritma lain seperti *Support Vector Machine, Random Forest, atau Neural Network* untuk memperoleh hasil yang lebih akurat dan optimal serta dapat dikembangkan sebagai sistem aplikasi pengambil Keputusan prediksi kelayakan penerima beasiswa KIP Kuliah di Politeknik TEDC Bandung dan institusi lain.

DAFTAR PUSTAKA

[1] R. M. I. Saputra, O. K. Rahmadhani, N. S. Hikmayanti, F. Amelia, A. D. Rosyada, and J. T. Nugraha, "Kontribusi Program KIP-K terhadap Peningkatan Motivasi dan Sikap Belajar Mahasiswa di Universitas Tidar," *Jurnal ISO: Jurnal Ilmu Sosial, Politik dan Humaniora*, vol. 5, no. 1, pp. 1–16, Jun. 2025, doi: 10.53697/iso.v5i1.2637.

[2] C. Nas, "*Data Mining* Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 11, no. 2, pp. 131–145, Oct. 2021, doi: 10.34010/jamika.v11i2.5506.

[3] H. Hartono, A. Hajjah, and Y. N. Marlim, "Penerapan Metode *Naïve Bayes Classifier* Untuk Klasifikasi Judul Berita," *Jurnal SimanteC*, vol. 12, no. 1, Feb. 2023.

[4] C. F. N. Halizah and M. Kartikasari, "Implementasi Algoritma *Naïve Bayes* Dalam Penentuan Pemberian Beasiswa KIP Kuliah (Studi Kasus STIKI Malang)," *Jurnal SimanteC*, vol. 12, no. 2, Jun. 2024.

[5] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode *Naive Bayes* Untuk Menentukan Calon Penerima PIP," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, 2023.

[6] E. Yuniarsih, R. Tiarani, R. Fariyanda, E. Y. A. Raki, and F. Damayanti, "Pengaruh Gaya Hidup Dan Mental *Accounting* Terhadap Pengelolaan Keuangan Mahasiswa Penerima KIP Kuliah (Studi Kasus: Mahasiswa FEB UNTAN)," *Jurnal Audit dan Akuntansi Fakultas Ekonomi Universitas Tanjunpura*, vol. 13, no. 1, pp. 111–137, Jun. 2024, doi: 10.26418/jaakfe.v13i1.81912.

[7] I. Ilham, I. G. Suwijana, and N. Nurdin, "Sistem Pendukung Keputusan Penerimaan Beasiswa Pada SMK 2 Sojol Menggunakan Metode AHP," *Jurnal Elektronik Sistem Informasi dan Komputer*, vol. 4, no. 2, Dec. 2020.

[8] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data *Mining* Data Netflix Menggunakan Aplikasi *Rapid Miner*," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.

[9] J. Eska, "Penerapan *Data Mining* untuk Prediksi Penjualan Walpaper Menggunakan Algoritma C4.5," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 2, no. 2, pp. 9–13, Mar. 2020.

[10] Mustika *et al.*, *Data Mining dan Aplikasinya*. Bandung: Widina Bhakti Persada Bandung, 2021. [*Online*]. Available: www.penerbitwidina.com

[11] E. Martantoh and N. Yanih, "Implementasi Metode *Naïve Bayes* Untuk Klasifikasi Karakteristik Kepribadiaan Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP MySQL," *JTSI*, vol. 3, no. 2, pp. 166–175, Sep. 2022.

[12] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis *Data Mining* Data Netflix menggunakan Aplikasi *RapidMiner*," *Journal of Business and Audit Information Syatems*, vol. 4, no. 1, 2021.

[13] R. R. R. Arisandi, B. Warsito, A. R. Hakim, " Aplikasi *Naïve Bayes Classifier* (NBC) Pada Klasifikasi Status Gizi Balita *Stunting* Dengan Pengujian *K-Fold Cross Validation*," *Jurnal Gaussian*, vol. 11, no 1, 2022.

[14] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*, Tonni Limbong. Medan: Yayasan Kita Menulis, 2020.

[15] I. Priyanto, E. M. Dewanti, Tundo, M. Nurdin, and R. Kasiono, "Penerapan Algoritma Metode *Naïve Bayes* Untuk Penentuan Penerimaan Bantuan Program Indonesia Pintar (PIP) (Studi Kasus SMP PGRI 1 Cilacap)," *Jurnal Manajamen Informatika Jayakarta*, vol. 4, no. 2, p. 162, Apr. 2024, doi: 10.52362/jmijayakarta.v4i2.1355.

[16] I. P. Oktavia and N. L. Anggreini, "Implementasi Algoritma *Naive Bayes* dengan Metode Klasifikasi dalam Menentukan Siswa Penerima Bantuan Program Indonesia Pintar (Studi Kasus : SMPN 3 Cihampelas)," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 6, Dec. 2024.

[17] Nurhidayati, Yahya, Fathurrahman, L. M. Samsu, and W. Amnia, "Implementasi Algoritma *Naive Bayes* Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi)," *Infotek: Jurnal Informatika dan Teknologi*, vol. 6, no. 1, pp. 177–188, Jan. 2023, doi: 10.29408/jit.v6i1.7529