

Data Mining Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes dalam Prediksi Penerimaan Beasiswa

Ahmad Ilham Kushartanto¹, Rima Tamara Aldisa^{2,*}

¹ Fakultas Teknologi Komunikasi dan Informatika, Informatika, Universitas Nasional, Jakarta, Indonesia

² Fakultas Teknologi Komunikasi dan Informatika, Sistem Informasi, Universitas Nasional, Jakarta, Indonesia

Email: ¹ilhamahmadre@gmail.com, ^{2,*}rimatamaraa@gmail.com

Email Penulis Korespondensi: rimatamaraa@gmail.com

Submitted: 11/11/2023; Accepted: 30/11/2023; Published: 30/11/2023

Abstrak—Salah satu tujuan dari Negara Indonesia yang tertuang pada Undang – Undang Dasar Negara Republik Indonesia adalah mencerdaskan kehidupan bangsa. Pemerintah melalui Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbudristek) menjalankan program wajib belajar 12 tahun. Program Kartu Indonesia Pintar Kuliah mulai terlaksana pada tahun 2021 dimana pada awalnya program ini bernama bidikmisi. Program Kartu Indonesia Pintar Kuliah diperuntukan membantu bagi anak – anak yang mengalami kesulitan ekonomi atau terkendala terhadap biaya untuk melanjutkan pendidikan pada tingkat perguruan tinggi. Bagi penerima bantuan program Kartu Indonesia Pintar Kuliah nantinya akan mendapatkan biaya penuh untuk pelaksanaan perkuliahan dimulai dari awal masuk hingga sampai menyelesaikan perkuliahan dengan jangka waktu yang telah disepakati. Hal tersebut bertujuan agar kiranya bagi anak – anak yang telah terpilih menjadi mahasiswa tidak lagi harus pusing ataupun merasa takut akan keberlangsungan hidup ketika melaksanakan perkuliahan. Pemberian batasan terhadap jumlah penerima Kartu Indonesia Pintar Kuliah bagi Perguruan Tinggi Swasta (PTS) menjadi sebuah permasalahan yang harus diselesaikan dengan baik dan cermat. Proses seleksi bagi calon mahasiswa yang mendaftar pada perguruan tinggi yang dituju bertujuan untuk mendapatkan penerima yang memang layak untuk mendapatkan bantuan program Kartu Indonesia Pintar Kuliah tersebut. Proses seleksi bagi calon mahasiswa yang mendaftar pada perguruan tinggi yang dituju bertujuan untuk mendapatkan penerima yang memang layak untuk mendapatkan bantuan program Kartu Indonesia Pintar Kuliah tersebut. Hasil yang didapatkan dari penerapan penelitian bahwasannya dari perbandingan hasil algoritma K-NN dan Naive Bayes didapatkan hasil yang sama terhadap data pengujian yaitu Diterima

Kata Kunci: Data Mining; Prediksi; Penerimaan; KIP Kuliah; K-NN; Naïve Bayes

Abstract—One of the goals of the Indonesian State as stated in the Constitution of the Republic of Indonesia is to make the nation's life more intelligent. The government, through the Ministry of Education, Culture, Research and Technology (Kemendikbudristek), is implementing a 12-year compulsory education program. The Indonesia Smart College Card program will start to be implemented in 2021, where initially this program was called bidikmisi. The Indonesia Smart College Card program is intended to help children who experience economic difficulties or are constrained by the costs of continuing their education at tertiary level. Beneficiaries of the Indonesian Smart College Card program will receive full tuition fees from the start of admission to completion of the course within the agreed time period. This aims to ensure that children who have been selected to become students no longer have to worry or feel afraid about their survival when studying. Placing limits on the number of recipients of the Indonesia Smart Tuition Card for Private Universities (PTS) is a problem that must be resolved properly and carefully. The selection process for prospective students who register at the target university aims to find recipients who are worthy of assistance from the Indonesia Smart College Card program. The selection process for prospective students who register at the target university aims to find recipients who are worthy of assistance from the Indonesia Smart College Card program. The results obtained from the application of the research were that from comparing the results of the K-NN and Naive Bayes algorithms, the same results were obtained for the test data, namely Acceptable

Keywords: Data Mining; Prediction; Reception; KIP Lecture; K-NN; Naïve Bayes

1. PENDAHULUAN

Salah satu tujuan dari Negara Indonesia yang tertuang pada Undang – Undang Dasar Negara Republik Indonesia adalah mencerdaskan kehidupan bangsa. Berdasarkan hal tersebut sudah sepatutnya pemerintah memperhatikan lebih baik terhadap pencerdasan kehidupan di Negara Indonesia. Salah satu cara yang dilakukan untuk mencerdaskan kehidupan bangsa adalah memberikan pendidikan yang baik dan layak bagi anak – anak generasi penerus bangsa[1], [2]. Pemerintah melalui Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbudristek) menjalankan program wajib belajar 12 tahun dimana seluruh anak – anak diwajibkan untuk menyelesaikan pendidikan hingga jenjang Sekolah Menengah Atas (SMA) / Sekolah Menengah Kejuruan (SMK). Dimana untuk menjamin terlaksananya program tersebut pemerintah memberikan berbagai macam bantuan guna mendukung seluruh anak di Indonesia dapat menyelesaikan pendidikan hingga jenjang SMA / SMK[3]–[5].

Dalam peningkatan dan pelaksanaan mencerdaskan generasi penerus di Indonesia, pada saat ini pemerintah melalui Kemendikbudristek sudah memberikan bantuan pendidikan hingga tingkat perguruan tinggi. Program bantuan yang diberikan pemerintah pada tingkat jenjang perguruan tinggi adalah Kartu Indonesia Pintar Kuliah (KIP – Kuliah)[6], [7]. Program Kartu Indonesia Pintar Kuliah mulai terlaksana pada tahun 2021 dimana pada awalnya program ini bernama bidikmisi. Program Kartu Indonesia Pintar Kuliah merupakan bantuan beasiswa kepada para generasi penerus untuk melanjutkan pendidikan pada tingkat perguruan tinggi. Program Kartu Indonesia Pintar Kuliah diperuntukan membantu bagi anak – anak yang mengalami kesulitan ekonomi

atau terkendala terhadap biaya untuk melanjutkan pendidikan pada tingkat perguruan tinggi[8]–[10]. Pemberian beasiswa Kartu Indonesia Pintar Kuliah terhadap anak – anak di Indonesia bukan saja dilihat berdasarkan dengan kesulitan ekonomi tetapi program ini lebih terfokus terhadap bagi anak – anak yang memiliki prestasi secara akademik ataupun non akademik yang sekiranya masih layak untuk melanjutkan pendidikan pada tingkat perguruan tinggi[11], [12]. Bagi penerima bantuan program Kartu Indonesia Pintar Kuliah nantinya akan mendapatkan biaya penuh untuk pelaksanaan perkuliahan dimulai dari awal masuk hingga sampai menyelesaikan perkuliahan dengan jangka waktu yang telah disepakati. Selain bantuan biaya pendidikan seperti pembayaran uang kuliah, uang praktikum, uang ujian dan lainnya program Kartu Indonesia Pintar Kuliah juga memberikan bantuan biaya hidup bagi penerima program tersebut[13], [14].

Hal tersebut bertujuan agar kiranya bagi anak – anak yang telah terpilih menjadi mahasiswa tidak lagi harus pusing ataupun merasa takut akan keberlangsungan hidup ketika melaksanakan perkuliahan. Dengan memberikan bantuan biaya hidup tersebut agar kiranya yang mendapatkan program Kartu Indonesia Pintar Kuliah tersebut dapat lebih fokus untuk menyelesaikan perkuliahannya dimasa waktu yang telah ditentukan dan juga agar nilai dalam setiap semesternya mendapatkan hasil yang maksimal[15], [16]. Penyelenggaraan program Kartu Indonesia Pintar Kuliah bukan saja diberikan bagi anak – anak yang ingin masuk pada Perguruan Tinggi Negeri (PTN) saja tetapi juga pada Perguruan Tinggi Swasta (PTS). Pemberian pada beasiswa di PTN dan PTS diharapkan dapat memberikan ataupun pelaksanaan pembelajaran yang merata dan seimbang. Selain itu juga agar tidak terjadi perbedaan bagi PTN ataupun PTS. Pemberian jumlah penerima Kartu Indonesia Pintar Kuliah bagi PTN dan PTS diatur ataupun dibagikan oleh lembaga yang berwenang dan bertanggung jawab pada setiap wilayahnya. Pemberian batasan terhadap jumlah penerima Kartu Indonesia Pintar Kuliah bagi Perguruan Tinggi Swasta (PTS) menjadi sebuah permasalahan yang harus diselesaikan dengan baik dan cermat. Hal tersebut dikarenakan jumlah kuota yang diterima lebih sedikit dibandingkan dengan jumlah yang mendaftar pada program Kartu Indonesia Pintar Kuliah. Perbedaan jumlah yang diterima dan mendaftar tersebut mengharuskan bagi penyelenggara penerima bantuan Kartu Indonesia Pintar Kuliah harus melakukan seleksi[17]–[19]. Proses seleksi bagi calon mahasiswa yang mendaftar pada perguruan tinggi yang dituju bertujuan untuk mendapatkan penerima yang memang layak untuk mendapatkan bantuan program Kartu Indonesia Pintar Kuliah tersebut. Proses seleksi yang dilakukan secara asal – asalan dikhawatirkan calon mahasiswa tersebut tidak dapat untuk mengikuti perkuliahan dengan baik sehingga nilai yang didapatkan pada setiap semester menjadi buruk dan juga masa studi yang dijalani melewati dari kesepakatan penyelesaian waktu studi[20], [21].

Bagi PTS yang menyelenggarakan proses penyeleksian terhadap penerima bantuan Kartu Indonesia Pintar Kuliah menjadi sebuah polemik baru, dimana bagi tim penyeleksi diharuskan untuk menentukan siapa saja yang layak untuk terpilih mendapatkan bantuan program Kartu Indonesia Pintar Kuliah tersebut. Dimana proses yang harus dilakukan bukan saja mempertimbangkan terhadap kemampuan ekonomi saja tetapi juga mempertimbangkan terhadap faktor prestasi akademik dan non akademik sesuai dengan tujuan dari pada program ini dilaksanakan[22], [23]. Masalah yang dihadapi bukanlah sebuah masalah yang bisa dianggap sepele, jika masalah ini tidak diselesaikan dengan baik dan bijak akan memberikan dampak – dampak lainnya bagi perguruan tinggi ataupun juga bagi mahasiswa tersebut. Bagi perguruan tinggi jika proses pemilihan ataupun mahasiswa yang terpilih tidak dapat mengikuti proses pembelajaran dengan baik maka bagi perguruan tinggi tersebut tidak mendapatkan kuota kembali terhadap penerimaan mahasiswa melalui jalur Kartu Indonesia Pintar Kuliah. Sedangkan bagi mahasiswa sendiri dampak terburuk adalah diberhentikannya terhadap pemberian beasiswa Kartu Indonesia Pintar Kuliah atau bahkan sampai dengan pengembalian dana. Permasalahan dan juga dampak yang cukup besar tersebut maka kiranya harus diselesaikan dengan baik, benar dan cermat. Maka dari itu diperlukan sebuah cara untuk membantu proses penyelesaian permasalahan tersebut. Dimana cara yang dapat dipergunakan dengan cara melakukan proses pengolahan data terkhususnya dari data – data penerima bantuan program Kartu Indonesia Pintar Kuliah sebelumnya. Proses penyelesaian permasalahan tersebut dapat diselesaikan dengan menggunakan data mining.

Data mining merupakan sebuah teknik pengolahan data, dimana pada data mining melakukan proses pengolahan data berdasarkan dengan data – data dimasa lampau. Data dimasa lampau tersebut yang telah disimpan dilakukan proses pengolahan kembali hingga akhirnya ditemukan sebuah informasi – informasi baru ataupun sebuah pola baru yang dapat dimanfaatkan untuk proses pengambilan keputusan[24], [25]. Penerapan data mining pada penyelesaian permasalahan tersebut sangatlah tepat, hal tersebut dikarenakan sebelumnya sudah terdapat data mahasiswa yang telah diterima mendapatkan bantuan Kartu Indonesia Pintar Kuliah dan telah melakukan perkuliahan. Dari data tersebut dapat dilihat pola seperti apa bagi calon mahasiswa yang nantinya dapat mengikuti perkuliahan dengan baik tanpa harus perlu dilakukan pembinaan secara rutin oleh dosen. Pada data mining sendiri terdapat banyak bagian proses penyelesaiannya salah satunya adalah klasifikasi[26]–[28].

Klasifikasi pada data mining bertujuan untuk melakukan pengelompokan – pengelompokan berdasarkan dengan atribut – atribut tertentu. Pengelompokan yang dilakukan berdasarkan dengan tujuan kelas dari data dan juga nilai – nilai atribut yang dimiliki dari setiap data. Pada proses klasifikasi data diharuskan memiliki tujuan kelas agar kiranya dapat dengan mudah untuk dilakukan proses pengelompokan. Pada klasifikasi data mining terdapat berbagai macam algoritma seperti K-Nearest Neighbor (K-NN) dan juga Naïve Bayes[20]–[23].

Algoritma K-Nearest Neighbor termasuk dalam bagian klasifikasi data mining, hal tersebut dikarenakan untuk dapat menyelesaikan dengan algoritma tersebut harus kiranya diketahui terlebih dahulu untuk target kelasnya. Proses penyelesaian yang dilakukan dengan menggunakan Algoritma K-Nearest Neighbor berdasarkan nilai kedekatan dari objek baru dengan objek yang lama. Proses awal dilakukan dengan menentukan nilai K sebagai dasar pengambilan keputusan nantinya, setelah itu dilakukan proses perhitungan jarak terdekat dari objek baru dengan objek lama berdasarkan dengan euclidean distance[29]–[31]. Algoritma lainnya selain K-Nearest Neighbor yaitu Algoritma Naïve Bayes. Algoritma Naïve Bayes bagian dari pada klasifikasi data mining dikarenakan proses pada algoritma tersebut berdasarkan dengan adopsi proses pada teorema bayes. Dimana proses dilakukan dengan melakukan perhitungan nilai probabilitas terhadap kelas yang akan dituju dari objek baru[32]–[34].

Pelaksanaan penelitian tidak terlepas dari penelitian terdahulu beberapa penelitian terdahulu seperti yang dilakukan oleh Rino Bahtiar pada tahun 2023 dengan judul penelitian Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor dimana dari penelitian yang dilakukan didapatkan hasil bahwa algoritma K-NN dapat diperuntukan mengklasifikasi objek dengan hasil prediksi yang didapatkan sebesar 88,89%. Dari hasil prediksi yang didapatkan dapat dikategorikan sangat baik dan hasil dapat digunakan untuk pengambilan keputusan[31]. Di tahun yang sama yaitu 2023 juga dilakukan penelitian oleh Rahmadini, dkk dengan judul penelitian Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor dimana hasil penelitian yang didapatkan yaitu algoritma K-NN dapat melakukan prediksi berdasarkan dengan data dimasa lampau[30].

Penelitian lainnya yang dilakukan oleh Ahmad Muhadi dan Alvino Octaviano ditahun 2023 juga dengan judul penelitian yang dilakukan Penerapan Data Mining Untuk Prediksi Hasil Keuntungan Lelang Mesin X-Ray Tahun 2020 Dengan Metode K-Nearest Neighbor (Studi Kasus : PT.Ramadika Mandiri) hasil yang didapatkan dari proses penelitian yaitu Dari hasil prediksi yang dilakukan didapatkan hasil prediksi keuntungan penjualan mesin x-ray sebanyak 3 tipe yaitu: 8 tipe bandara dinyatakan untung, 1 tipe medis dinyatakan untung dan 2 tipe hotel dinyatakan untung dengan akurasi data sebesar 77,68%[29]. Selain algoritma K-NN juga terdapat penelitian terdahulu dari algoritma Naïve Bayes seperti penelitian yang dilakukan oleh Qurotul A'yuniyah, dkk pada tahun 2023 dimana judul penelitian yang dilakukan Analisa Algoritma Naïve Bayes Classifier (NBC) Untuk Prediksi Penjualan Alat Kesehatan dengan hasil penelitian yang didapatkan Hasil dari data penjualan alat Kesehatan dengan 2 kelas yaitu "Laris" dan "Tidak Laris". Nilai yang dihasilkan sangat baik dan dapat dijadikan sebagai landasan untuk mengklasifikasikan penjualan alat kesehatan dengan menganalisa stok barang pada[34].

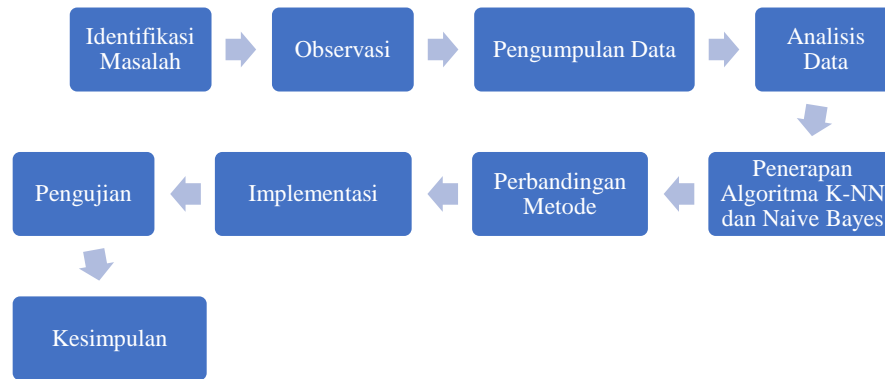
Penelitian lainnya yang telah dilakukan seperti penelitian yang dilakukan oleh Nurfazriah Attamam, dkk pada tahun 2023 dengan judul penelitian Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan dimana hasil penelitian yang didapatkan bahwa tingkat akurasi yang didapatkan dari algoritma Naïve Bayes sebesar 92,97% dimana hal ini menandakan bahwa algoritma Naïve Bayes dapat dipergunakan untuk melakukan proses prediksi[33]. Penelitian lainnya juga dilakukan pada tahun 2023 oleh Nur Syifa Fauzia dan Raditya Danar Dana dengan judul penelitian yang dilakukan Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari bahwasannya algoirtma Naïve Bayes dapat dipergunakan untuk melakukan proses prediksi, dimana hasil tingkat akurasi yang didapatkan sebesar 93,69%[32].

Pada penelitian ini akan dilakukan proses penyelesaian dengan menggunakan algoritma K-NN dan Naïve Bayes. Penelitian ini bertujuan untuk melakukan perbandingan hasil dari kedua algoritma tersebut. Perbandingan dilakukan untuk melihat hasil yang didapatkan apakah sudah benar layak termasuk dalam mahasiswa yang mendapatkan program beasiswa Kartu Indonesia Pintar Kuliah. Perbandingan dari kedua algoritma bertujuan untuk mendapatkan hasil yang lebih pasti dari proses prediksi yang dilakukan. Sering kali didapatkan hasil penyelesaian dengan menggunakan algoritma klasifikasi hasil yang didapatkan masih belum teruji dengan baik ataupun tingkat akurasi yang cukup rendah. Maka dari itu dengan dilakukan perbandingan dari kedua algoritma tersebut nantinya diharapkan untuk dapat memberikan rekomendasi yang paling bagus dari hasil yang didapatkan dari penerapan algoritma tersebut.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan gambaran terhadap proses – proses yang dilakukan pada penelitian, dengan membuat kerangka kerja penelitian maka proses yang dilakukan pada penelitian tidaklah lagi berulang. Selain itu kerangka kerja penelitian juga sebagai struktur dasar dari pelaksanaan penelitian dimulai dari tahapan awal hingga tahapan akhir penelitian. Adapun pada penelitian ini kerangka kerja penelitian dapat dilihat pada gambar alur proses di bawah ini:



Gambar 1. Kerangka Kerja Penelitian

2.2 Data Mining

Data mining merupakan bagian dari pengolahan data pada komputer. Data mining memiliki peran terhadap proses pengolahan data yang cukup besar. Data yang besar tersebut tersimpan pada sebuah wadah yang bernama datawarehouse. Dengan data yang tersimpan tersebut kemudian nantinya akan dilakukan proses pengolahan data hingga didapatkan kembali informasi – informasi baru yang tersimpan pada kumpulan data tersebut. Pencarian data tersebutlah merupakan peran dari data mining, data mining bertanggung jawab untuk menemukannya sebuah informasi atau pola yang tersembunyi dari kumpulan data. Informasi ataupun pola tersebut nantinya akan dimanfaatkan dan digunakan bagi pemilik data untuk dilakukan proses pengambilan keputusan[5], [8]. Pada sekarang ini data mining sudah banyak dimanfaatkan pada banyak kalangan dan bidang ilmu, bukan saja dimanfaatkan pada perusahaan dengan skala besar tetapi perusahaan berskala kecil ataupun bahkan perorangan juga menerapkan data mining untuk melakukan proses pengolahan data. Hasil data yang didapatkan dari data mining beragam bentuknya dan disesuaikan dengan kebutuhan dari pemilik ataupun pengguna data tersebut[7], [14], [17].

2.3 Algoritma K-NN

Algoritma K-Nearest Neighbor merupakan algoritma pada data mining yang termasuk dalam klasifikasi data mining. Proses pada algoritma K-Nearest Neighbor dimulai dengan penentuan nilai K terlebih dahulu. Setelah penentuan nilai K maka selanjutnya proses yang dilakukan menghitung kedekatan nilai antar objek. Pada K-Nearest Neighbor objek yang terpilih untuk memenuhi nilai K dihitung berdasarkan dengan euclidean distance. Euclidean distance merupakan proses perhitungan jarak terdekat antara objek baru dengan objek lama yang tersimpan pada data. Adapun rumus yang digunakan untuk menyelesaikan algoritma K-Nearest Neighbor yaitu[35]–[37]:

$$Dq = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_n - b_n)^2} \tag{1}$$

2.4 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan algoritma klasifikasi pada data mining. Algoritma Naïve Bayes melakukan proses klasifikasi berdasarkan dengan konsep statistika, hal tersebut didasari dengan algoritma Naïve Bayes mengadopsi dari Teorema Bayes. Setiap kelas pada algoritma Naïve Bayes memiliki nilai probabilitas tersendiri yang nantinya dapat digunakan untuk proses pengambilan keputusan. Adapun persamaan umum yang digunakan yaitu[38], [39]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{2}$$

3. HASIL DAN PEMBAHASAN

3.1 Analisis Masalah

Dalam menunjang program wajib belajar pemerintah memberlakukan program beasiswa yang bernama Kartu Indonesia Pinta Kuliah (KIP Kuliah). Program Kartu Indonesia Pintar Kuliah bertujuan untuk memberikan bantuan materi bagi anak – anak yang ingin melanjutkan pendidikan pada tingkat perguruan tinggi namun terkendala terhadap biaya perkuliahan. Selain itu juga, program Kartu Indonesia Pintar Kuliah juga memberikan bantuan tunjangan biaya hidup ketika menjalankan masa studi dengan tujuan agar lebih fokus untuk menjalankan perkuliahan hingga mendapatkan hasil yang maksimal pada setiap matakuliah dan juga dapat menyelesaikan perkuliahan dengan tepat waktu. Dalam pelaksanaannya Kartu Indonesia Pintar Kuliah diberlakukan secara merata bagi Perguruan Tinggi Negeri dan Perguruan Tinggi Swasta. Proses pelaksanaannya setiap perguruan tinggi mendapatkan kuota yang telah ditentukan oleh lembaga penanggung jawab dimasing – masing wilayah,

dimana setiap perguruan tinggi kemungkinan mendapatkan kuota yang berbeda berdasarkan dengan pertimbangan didalamnya. Pembatasan jumlah kuota tersebut menjadi sebuah polemik baru bagi perguruan tinggi, dimana jumlah calon mahasiswa yang mendaftar lebih banyak dibandingkan dengan jumlah kuota yang diterima. Jumlah pelamar yang melebihi jumlah kuota maka diperlukan proses penyeleksian, dimana proses penyeleksian harus dilakukan dengan sebaik mungkin, agar kiranya mahasiswa yang terpilih merupakan mahasiswa yang pantas untuk mendapatkan bantuan program Kartu Indonesia Pintar Kuliah. Pemberian bantuan Kartu Indonesia Pintar Kuliah bukan saja diberikan bagi calon mahasiswa yang kekurangan secara ekonomi tetapi lebih terfokus terhadap calon mahasiswa yang memiliki prestasi baik dibidang akademik ataupun non akademik yang kiranya masih memiliki potensi untuk melanjutkan pendidikan pada tingkat perguruan tinggi. Proses penyeleksian yang dilakukan diharuskan untuk mempertimbangkan atribut – atribut pendukung, hal tersebut dikarenakan kesalahan dalam proses penerimaan bukan saja menimbulkan keraguan salah satu pihak tetapi juga seluruh pihak.

Bagi perguruan tinggi jika mahasiswa bermasalah dalam proses pendidikannya maka tidak akan diberikan kembali kuota Kartu Indonesia Pintar Kuliah hingga terselesainya permasalahan tersebut dan bagi mahasiswa dampak yang terjadi hingga sampai pembatalan program beasiswa Kartu Indonesia Pintar Kuliah. Pentingnya peran atribut dalam proses penyeleksian maka kiranya harus diselesaikan dengan menggunakan teknik khusus ataupun teknik tersebut bernama data mining. Dalam proses penyelesaiannya data mining bertujuan untuk melakukan proses pengolahan data dengan meninjau kembali kumpulan data yang tersimpan pada datawarehouse. Kumpulan data diproses hingga didapatkan sebuah pola baru ataupun informasi– informasi baru yang dapat dipergunakan untuk proses penyelesaian permasalahan. Informasi baru yang didapatkan kembali pada proses data mining bisa bersifat prediksi, asosisasi, estimasi, klasifikasi ataupun klasterisasi. Dalam penelitian ini proses penyelesaian dengan menggunakan klasifikasi dan terdapat algoritma yang digunakan yaitu Algoritma K-NN dan Algoritma Naïve Bayes. Algoritma K-NN dan Algoritma Naïve Bayes dipergunakan pada data mining untuk membantu dalam proses pengklasifikasi data dan juga untuk melakukan prediksi data. Dimana kedua algoritma tersebut sudah banyak teruji dapat membantu proses penyelesaian permasalahan pada data mining. Pada penelitian ini akan dipergunakan kedua algoritma tersebut untuk membantu proses penyelesaian masalah. Dimana dengan kedua algoritma tersebut nantinya akan dilakukan proses perbandingan hasil yang didapatkan. Perbandingan dari kedua algoritma bertujuan untuk mendapatkan hasil yang lebih terpercaya dari hasil yang didapatkan.

3.1.1 Pengumpulan Data

Tahapan ataupun langkah awal sebelum dilakukan proses pengolahan data harus dilakukan terlebih dahulu pengumpulan data. Pengumpulan data bertujuan untuk mengetahui dan juga membatasi data – data mana saja yang dapat dipergunakan untuk proses pengolahan data hingga didapatkan informasi baru yang berharga bagi pemilik data. Adapun data yang digunakan pada penelitian dapat dilihat pada tabel 1 berikut:

Tabel 1. Data Penelitian

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondisi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	Rp. 2.500.000	3	Permanen	Milik Sendiri	1	Diterima
A2	Rp. 2.250.000	2	Tidak Permanen	Sewa	3	Diterima
A3	Rp. 2.100.000	2	Semi Permanen	Milik Sendiri	2	Diterima
A4	Rp. 1.850.000	2	Permanen	Sewa	3	Diterima
A5	Rp. 3.100.000	2	Semi Permanen	Milik Sendiri	1	Tidak Diterima
A6	Rp. 3.050.000	1	Permanen	Milik Sendiri	2	Tidak Diterima
A7	Rp. 1.500.000	2	Tidak Permanen	Sewa	3	Diterima
A8	Rp. 1.800.000	3	Permanen	Sewa	2	Diterima
A9	Rp. 2.200.000	2	Semi Permanen	Milik Sendiri	1	Tidak Diterima
A10	Rp. 2.500.000	1	Permanen	Sewa	1	Tidak Diterima

Pada tabel 1 dapat dilihat untuk data yang digunakan pada penelitian, dimana pada data tersebut digunakan data objek sebanyak 10 orang yang merupakan data pendaftar program Kartu Indonesia Pintar Kuliah sebelumnya. Pada tabel tersebut juga terlihat atribut dan nilai atribut yang digunakan pada penelitian. Selain data penelitian atau sampel data, dalam proses penyelesaian data mining harus dibutuhkan data pengujian atau data yang akan dilakukan proses pengambilan keputusan. Adapun data tersebut dapat dilihat pada tabel 2 berikut

Tabel 2. Data Pengujian / Data Baru

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondisi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	Rp. 2.100.000	2	Semi Permanen	Sewa	3	???

3.1.2 Preprocessing Data

Selain pengumpulan data dalam proses penyelesaian data mining juga diperlukan preprocessing data. Preprocessing data merupakan proses peninjauan kembali terhadap data apakah data terdapat kesalahan data ataupun juga data sudah sesuai dan siap pakai terhadap algoritma yang digunakan. Dalam penelitian ini terdapat 2 (dua) preprocessing data yaitu

Tabel 3. Hasil Preprocessing Data Untuk Algoritma K-NN

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondisi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	2	3	3	1	1	Diterima
A2	2	2	1	2	3	Diterima
A3	2	2	2	1	2	Diterima
A4	1	2	3	2	3	Diterima
A5	3	2	2	1	1	Tidak Diterima
A6	3	1	3	1	2	Tidak Diterima
A7	1	2	1	2	3	Diterima
A8	1	3	3	2	2	Diterima
A9	2	2	2	1	1	Tidak Diterima
A10	2	1	3	2	1	Tidak Diterima

Pada tabel 3 dapat dilihat preprocessing data untuk algoritma K-NN. Proses penyelesaian pada algoritma K-NN diselesaikan dengan menggunakan data numerical, sehingga data – data harus dirubah dalam bentuk numerical. Selanjutnya juga dilakukan preprocessing terhadap algoritma Naïve Bayes. Dimana hasil preprocessing data algoritma Naïve Bayes dapat dilihat pada tabel 4 berikut:

Tabel 4. Hasil Preprocessing Data Untuk Algoritma Naïve Bayes

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondisi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	Sedang	Sedang	Permanen	Milik Sendiri	Sedikit	Diterima
A2	Sedang	Sedikit	Tidak Permanen	Sewa	Sedang	Diterima
A3	Sedang	Sedikit	Semi Permanen	Milik Sendiri	Sedikit	Diterima
A4	Sedikit	Sedikit	Permanen	Sewa	Sedang	Diterima
A5	Banyak	Sedikit	Semi Permanen	Milik Sendiri	Sedikit	Tidak Diterima
A6	Banyak	Sedikit	Permanen	Milik Sendiri	Sedikit	Tidak Diterima
A7	Sedikit	Sedikit	Tidak Permanen	Sewa	Sedang	Diterima
A8	Sedikit	Sedang	Permanen	Sewa	Sedikit	Diterima
A9	Sedang	Sedikit	Semi Permanen	Milik Sendiri	Sedikit	Tidak Diterima
A10	Sedang	Sedikit	Permanen	Sewa	Sedikit	Tidak Diterima

Pada tabel 4 dapat dilihat hasil preprocessing data yang akan digunakan untuk melakukan proses penyelesaian dengan menggunakan algoritma Naïve Bayes. Berbeda halnya dengan K-NN, proses penyelesaian pada Naïve Bayes data yang digunakan merupakan data kategorikal / pengelompokan.

3.1.3 Penerapan Algoritma K-NN

Setelah dilakukan proses preprocessing data maka selanjutnya dapat diselesaikan untuk proses penerapan dari algoritma K-NN. Proses penyelesaian dengan algoritma K-NN nantinya akan dilakukan perhitungan nilai jarak terdekat dari data testing dengan nilai objek pada data sebelumnya. Adapun data yang digunakan pada proses algoritma K-NN dapat dilihat di tabel 5.

Tabel 5. Dataset Algoritma K-NN

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondisi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	2	3	3	1	1	Diterima
A2	2	2	1	2	3	Diterima
A3	2	2	2	1	2	Diterima
A4	1	2	3	2	3	Diterima
A5	3	2	2	1	1	Tidak Diterima
A6	3	1	3	1	2	Tidak Diterima
A7	1	2	1	2	3	Diterima
A8	1	3	3	2	2	Diterima
A9	2	2	2	1	1	Tidak Diterima

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondisi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A10	2	1	3	2	1	Tidak Diterima

Selanjutnya setelah diketahui data tersebut terlebih dahulu ditentukan nilai K sebagai dasar pengambilan keputusan. Pada penelitian ini nilai K yang akan digunakan adalah K=5. Setelah itu diharuskan terlebih dahulu untuk mengetahui data uji yang akan digunakan pada penyelesaian algoritma K-NN adapun data uji dapat dilihat pada tabel berikut

Tabel 6. Data Pengujian Algoritma K-NN

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondisi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	2	2	2	2	3	???

Setelah diketahui data uji yang akan digunakan pada proses penyelesaian dengan algoritma K-NN maka selanjutnya dapat dilakukan proses penyelesaian dengan melakukan perhitungan nilai jarak. Adapun proses perhitungan nilai jarak dapat dilihat pada pross berikut:

$$Dq = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_n - b_n)^2}$$

$$D1 = \sqrt{(2 - 2)^2 + (2 - 3)^2 + (2 - 3)^2 + (2 - 1)^2 + (3 - 1)^2}$$

$$= \sqrt{(0)^2 + (-1)^2 + (-1)^2 + (1)^2 + (2)^2}$$

$$= \sqrt{0 + 1 + 1 + 1 + 4}$$

$$= \sqrt{7}$$

$$= 2,65$$

$$D2 = \sqrt{(2 - 2)^2 + (2 - 2)^2 + (2 - 1)^2 + (2 - 2)^2 + (3 - 3)^2}$$

$$= \sqrt{(0)^2 + (0)^2 + (1)^2 + (0)^2 + (0)^2}$$

$$= \sqrt{0 + 0 + 1 + 0 + 0}$$

$$= \sqrt{1}$$

$$= 1$$

$$D3 = \sqrt{(2 - 2)^2 + (2 - 2)^2 + (2 - 2)^2 + (2 - 1)^2 + (3 - 2)^2}$$

$$= \sqrt{(0)^2 + (0)^2 + (0)^2 + (1)^2 + (1)^2}$$

$$= \sqrt{0 + 0 + 0 + 1 + 1}$$

$$= \sqrt{2}$$

$$= 1,41$$

$$D4 = \sqrt{(2 - 1)^2 + (2 - 2)^2 + (2 - 3)^2 + (2 - 2)^2 + (3 - 3)^2}$$

$$= \sqrt{(1)^2 + (0)^2 + (-1)^2 + (0)^2 + (0)^2}$$

$$= \sqrt{1 + 0 + 1 + 0 + 0}$$

$$= \sqrt{2}$$

$$= 1,41$$

$$D5 = \sqrt{(2 - 3)^2 + (2 - 2)^2 + (2 - 2)^2 + (2 - 1)^2 + (3 - 1)^2}$$

$$= \sqrt{(-1)^2 + (0)^2 + (0)^2 + (1)^2 + (2)^2}$$

$$= \sqrt{1 + 0 + 0 + 1 + 4}$$

$$= \sqrt{6}$$

$$= 2,45$$

$$D6 = \sqrt{(2 - 3)^2 + (2 - 1)^2 + (2 - 3)^2 + (2 - 1)^2 + (3 - 2)^2}$$

$$= \sqrt{(-1)^2 + (1)^2 + (-1)^2 + (1)^2 + (1)^2}$$

$$= \sqrt{1 + 1 + 1 + 1 + 1}$$

$$= \sqrt{5}$$

$$= 2,24$$

$$D7 = \sqrt{(2 - 1)^2 + (2 - 2)^2 + (2 - 1)^2 + (2 - 2)^2 + (3 - 3)^2}$$

$$= \sqrt{(1)^2 + (0)^2 + (1)^2 + (0)^2 + (0)^2}$$

$$= \sqrt{1 + 0 + 1 + 0 + 0}$$

$$= \sqrt{2}$$

$$= 1,41$$

$$D8 = \sqrt{(2 - 1)^2 + (2 - 3)^2 + (2 - 3)^2 + (2 - 2)^2 + (3 - 2)^2}$$

$$= \sqrt{(1)^2 + (-1)^2 + (-1)^2 + (0)^2 + (1)^2}$$

$$= \sqrt{1 + 1 + 1 + 0 + 1}$$

$$= \sqrt{4}$$

$$= 2$$

$$D9 = \sqrt{(2 - 2)^2 + (2 - 2)^2 + (2 - 2)^2 + (2 - 1)^2 + (3 - 1)^2}$$

$$= \sqrt{(0)^2 + (0)^2 + (0)^2 + (1)^2 + (2)^2}$$

$$= \sqrt{0 + 0 + 0 + 1 + 4}$$

$$= \sqrt{5}$$

$$= 2,24$$

$$D10 = \sqrt{(2 - 2)^2 + (2 - 1)^2 + (2 - 3)^2 + (2 - 2)^2 + (3 - 1)^2}$$

$$= \sqrt{(0)^2 + (1)^2 + (-1)^2 + (0)^2 + (2)^2}$$

$$= \sqrt{0 + 1 + 1 + 0 + 4}$$

$$= \sqrt{6}$$

$$= 2,45$$

Setelah dilakukan proses perhitungan jarak dari data testing dengan data sebelumnya, maka selanjutnya dilakukan penyusunan data. Dimana penyusunan data perhitungan jarak dapat dilihat pada tabel 7.

Tabel 7. Hasil Perhitungan Jarak

No	Objek	Hasil Perhitungan Jarak
1	A1	2,65
2	A2	1
3	A3	1,41
4	A4	1,41
5	A5	2,45
6	A6	2,24
7	A7	1,41
8	A8	2
9	A9	2,24
10	A10	2,45

Setelah disusun terhadap hasil perhitungan jarak maka selanjutnya dilakukan proses perankingan, dimana proses perankingan dimulai dari jarak terdekat hingga jarak terjauh. Tujuan dilakukan perankingan agar memudahkan dalam proses pemilihan nilai K. Adapun hasil perankingan dapat dilihat pada tabel 8.

Tabel 8. Hasil Perankingan Jarak

No	Objek	Hasil Perhitungan Jarak	Ranking
1	A2	1	1
2	A3	1,41	2
3	A4	1,41	3
4	A7	1,41	4
5	A8	2	5
6	A6	2,24	6
7	A9	2,24	7
8	A5	2,45	8
9	A10	2,45	9
10	A1	2,65	10

Setelah dilakukan proses perankingan data, selanjutnya memisahkan data berdasarkan dengan nilai K. Pada penelitian ini nilai K yang digunakan adalah K=5, maka dari itu nilai jarak terdekat dari ranking 1 dan 5 akan dipisahkan dengan data yang lainnya. Adapun hasil pengelompokan nilai K dapat dilihat berikut pada tabel 9.

Tabel 9. Hasil Data Nilai K=5

No	Objek	Hasil Perhitungan Jarak	Ranking	Hasil
1	A2	1	1	Diterima
2	A3	1,41	2	Diterima

No	Objek	Hasil Perhitungan Jarak	Ranking	Hasil
3	A4	1,41	3	Diterima
4	A7	1,41	4	Diterima
5	A8	2	5	Diterima

Setelah dilakukan proses pemilihan data berdasarkan dengan nilai K-5, maka selanjutnya dilakukan proses pengambilan keputusan berdasarkan dengan hasil terbanyak dari data apakah Diterima atau Tidak Diterima. Dari hasil nilai K=5 dapat dilihat bahwasannya seluruh data memiliki hasil Diterima oleh karena itu hasil prediksi yang dilakukan yaitu Diterima. Ataupun hasil dapat dilihat berikut

Tabel 10. Hasil Penerapan Algoritma K-NN

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondsi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	Rp. 2.100.000	2	Semi Permanen	Sewa	3	Diteirma

3.1.4 Penerapan Algoritma K-Naïve Bayes

Setelah dilakukan proses dengan menggunakan algoritma K-NN selanjutnya proses dilakukan dengan menerapkan algoritma Naïve Bayes. Proses pada Naïve Bayes berdasarkan dengan teorema bayes dan juga besarnya tingkat probabilitas dari data. Adapun data yang digunakan pada proses penyelesaian algoritma Naïve Bayes dapat dilihat pada tabel 11.

Tabel 11. Dataset Algoritma Naïve Bayes

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondsi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	Sedang	Sedang	Permanen	Milik Sendiri	Sedikit	Diterima
A2	Sedang	Sedikit	Tidak Permanen	Sewa	Sedang	Diterima
A3	Sedang	Sedikit	Semi Permanen	Milik Sendiri	Sedikit	Diterima
A4	Sedikit	Sedikit	Permanen	Sewa	Sedang	Diterima
A5	Banyak	Sedikit	Semi Permanen	Milik Sendiri	Sedikit	Tidak Diterima
A6	Banyak	Sedikit	Permanen	Milik Sendiri	Sedikit	Tidak Diterima
A7	Sedikit	Sedikit	Tidak Permanen	Sewa	Sedang	Diterima
A8	Sedikit	Sedang	Permanen	Sewa	Sedikit	Diterima
A9	Sedang	Sedikit	Semi Permanen	Milik Sendiri	Sedikit	Tidak Diterima
A10	Sedang	Sedikit	Permanen	Sewa	Sedikit	Tidak Diterima

Pada tabel 11 dapat dilihat data yang akan digunakan untuk melakukan proses penyelesaian dengan algoritma Naïve Bayes. Sama halnya dengan algoritma K-NN pada Naïve Bayes juga dibutuhkan data uji untuk dilakukan prediksi Diterima atau Tidak Diterima. Data uji dapat dilihat berikut pada tabel 12.

Tabel 12. Data Pengujian Algoritma Naïve Bayes

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondsi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	Sedang	Sedang	Semi Permanen	Sewa	Sedang	???

Setelah didapatkan data uji untuk dilakukan prediksi maka selanjutnya dilakukan proses penyelesaian dengan algoritma Naïve Bayes. Adapun tahapan proses dari algoritma Naïve Bayes dapat dilihat berikut

1. Menghitung Nilai $p(XK|H)$ tiap kelas
 - a. $P(\text{Atribut Penghasilan Orang Tua} = \text{"Sedang"} | \text{Hasil} = \text{"Diterima"})$
 $P(\text{Atribut Penghasilan Orang Tua} = \text{"Sedang"}) = 3/6 = 0,5$
 - b. $P(\text{Atribut Penghasilan Orang Tua} = \text{"Sedang"} | \text{Hasil} = \text{"Tidak Diterima"})$
 $P(\text{Atribut Penghasilan Orang Tua} = \text{"Sedang"}) = 2/4 = 0,5$
 - c. $P(\text{Atribut Jumlah Tanggungan} = \text{"Sedang"} | \text{Hasil} = \text{"Diterima"})$
 $P(\text{Atribut Jumlah Tanggungan} = \text{"Sedang"}) = 2/6 = 0,33$
 - d. $P(\text{Atribut Jumlah Tanggungan} = \text{"Sedang"} | \text{Hasil} = \text{"Tidak Diterima"})$
 $P(\text{Atribut Jumlah Tanggungan} = \text{"Sedang"}) = 0/4 = 0$
 - e. $P(\text{Atribut Kondisi Rumah} = \text{"Semi Permanen"} | \text{Hasil} = \text{"Diterima"})$
 $P(\text{Atribut Kondisi Rumah} = \text{"Semi Permanen"}) = 1/6 = 0,17$
 - f. $P(\text{Atribut Kondisi Rumah} = \text{"Semi Permanen"} | \text{Hasil} = \text{"Tidak Diterima"})$
 $P(\text{Atribut Kondisi Rumah} = \text{"Semi Permanen"}) = 2/4 = 0,5$
 - g. $P(\text{Atribut Kepemilikan Rumah} = \text{"Sewa"} | \text{Hasil} = \text{"Diterima"})$
 $P(\text{Atribut Kepemilikan Rumah} = \text{"Sewa"}) = 4/6 = 0,67$

- h. $P(\text{Atribut Kepemilikan Rumah} = \text{"Sewa"} \mid \text{Hasil} = \text{"Tidak Diterima"})$
 $P(\text{Atribut Kepemilikan Rumah} = \text{"Sewa"}) = 1/4 = 0,25$
 - i. $P(\text{Atribut Jumlah Prestasi} = \text{"Sedang"} \mid \text{Hasil} = \text{"Diterima"})$
 $P(\text{Atribut Jumlah Prestasi} = \text{"Sedang"}) = 3/6 = 0,5$
 - j. $P(\text{Atribut Jumlah Prestasi} = \text{"Sedang"} \mid \text{Hasil} = \text{"Tidak Diterima"})$
 $P(\text{Atribut Jumlah Prestasi} = \text{"Sedang"}) = 0/4 = 0$
2. Menghitung Nilai $p(X|H)$ untuk tiap kelas label
- a. $P(X|\text{Hasil} = \text{"Diterima"})$
 $= 0,5 * 0,33 * 0,17 * 0,67 * 0,5 = 0,00939675$
 - b. $P(X|\text{Hasil} = \text{"Tidak Diterima"})$
 $= 0,5 * 0 * 0,5 * 0,25 * 0 = 0$
3. Menghitung Nilai $p(X|H) * p(H)$
- a. $(P(X|\text{Hasil} = \text{"Diterima"}) * P(\text{Hasil} = \text{"Diterima"}))$
 $= 0,00939675 * 6/10$
 $= 0,00939675 * 0,6 = 0,00563805$
 - b. $(P(X|\text{Hasil} = \text{"Tidak Diterima"}) * P(\text{Hasil} = \text{"Tidak Diterima"}))$
 $= 0 * 4/10$
 $= 0 * 0,4 = 0$
4. Menentukan Kelas

Setelah dilakukan proses penyelesaian dengan algoritma Naïve Bayes nilai probabilitas unuk hasil Diterima sebesar **0,00563805**. Sedangkan untuk nilai probablitas hasil Tidak Diterima adalah **0**. Berdasarkan dengan hasil proses nilai probabilitas tersebut maka dapat disimpulkan bahwasannya data uji termasuk dalam kategori diterima. Adapun proses dapat dilihat pada tabel 13 berikut.

Tabel 13. Hasil Penerapan Algoritma K-NN

	Penghasilan Orang Tua	Jumlah Tanggungan	Kondsi Rumah	Kepemilikan Rumah	Jumlah Prestasi	Hasil
A1	Rp. 2.100.000	2	Semi Permanen	Sewa	3	Diteirma

3.2 Pembahasan

Setelah dilakukan proses pengujian dengan algoritma Naïve Bayes dan K-NN untuk seluruh data penerima Kartu Indonesia Pintar Kuliah, maka selanjutnya dilakukan pembahasan hasil. Adapun hasil dari kedua tersebut dapat dilihat pada tabel berikut

Tabel 14. Hasil Proses Penerapan

No	Algoritma	Hasil
1	K-NN	Diterima
2	Naïve Bayes	Diterima

Dari hasil pembahasan dan penerapan dari kedua algoritma tersebut dapat dikatakan bahwasannya hasil yang didapatkan dari kedua algoritma adalah sama yaitu Diterima. Dengan hasil ini maka untuk proses pengambilan keputusan tidaklah perlu terdapat kesalahan kembali.

4. KESIMPULAN

Langkah akhir yang dilakukan pada penelitian yaitu penarikan sebuah kesimpulan. Adapun hasil kesimpulan yang didapatkan dari proses penelitian bahwa algoirtma data mining merupakan teknik yang cukup mudah digunakan dalam proses prediksi pada data uji. Selain itu proses yang dilakukan dapat dilihat bahwa hasil yang didapatkan dari penerapan kedua algoritma adalah sama yaitu "Diterima"

REFERENCES

- [1] I. Arfyanti, M. Fahmi, and P. Adytia, "Penerapan Algoritma Decision Tree Untuk Penentuan Pola Penerima Beasiswa KIP Kuliah," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1196–1201, 2022, doi: 10.47065/bits.v4i3.2275.
- [2] R. N. H. Hutasuhut, H. Okprana, and B. E. Damanik, "Penerapan Data Mining Untuk Menentukan

- Penerima Program Bidikmisi Menggunakan Algoritma K-Medoids,” *TIN Terap. Inform. Nusant.*, vol. 2, no. 11, pp. 667–672, 2022.
- [3] S. Aminah and T. Susanti, “Implementasi Algoritma K-Means Clustering Penerima Bantuan Beasiswa UKT Pada Institut Teknologi Pagar Alam,” *J. Ilm. Teknosains*, vol. 9, no. 1, 2023, [Online]. Available: <https://journal.upgris.ac.id/index.php/JITEK/article/view/15889%0Ahttps://journal.upgris.ac.id/index.php/JITEK/article/viewFile/15889/6936>.
- [4] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, “Data Mining Algorithm Testing For SAND Metaverse Forecasting,” *J. Appl. Intell. Syst.*, vol. 7, no. 3, pp. 259–267, 2022, doi: 10.33633/jais.v7i3.7155.
- [5] M. Martiano, Y. Sary, and F. Akbar, “Analysis and Optimization of the K-Means Algorithm in Determining Course Scheduling,” *J. Inf. Syst. Res.*, vol. 5, no. 1, 2023, doi: 10.47065/josh.v5i1.4343.
- [6] R. A. Andari, N. P. Lestari, and A. Rosadi, “PENERIMA BEASISWA BIDIKMISI DAN NON BEASISWA BIDIKMISI,” *JIPM*, vol. 4, no. 1, pp. 47–53, 2022.
- [7] M. W. D. Wirawan, S. K. Gusti, Jasril, and Pizain, “Penerapan Algoritma C4 . 5 Mengklarifikasi Penerimaan Bantuan,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 1, pp. 69–78, 2023, doi: 10.30865/json.v5i1.6653.
- [8] D. K. Utami, N. Irawati, and Sumantri, “Analisis Metode k-Means pada Clustering Penerimaan Bantuan PKH Desa Pulau Rakyat Tua,” *Sist. J. Sist. Inf.*, vol. 12, no. 3, pp. 953–961, 2023.
- [9] E. Budiarto, R. Rino, S. Hariyanto, and D. Susilawati, “Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan Algoritma C4.5,” *J. ALGOR*, vol. 3, no. 2, pp. 23–34, 2022, doi: 10.31253/alogr.v3i2.1019.
- [10] Asriyanik and A. Pambudi, “Machine Learning-Based Classification for Scholarship Selection,” *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 11, no. 2, pp. 447–460, 2023.
- [11] F. Purwani, R. T. Wahyudi, and I. D. Jaya, “Penerapan Algoritma K-Nearest Neighbor dengan Euclidean Distance untuk Menentukan Kelompok Uang Kuliah Tunggal Mahasiswa,” *Edumatic J. Pendidik. Inform.*, vol. 6, no. 2, pp. 344–353, 2022, doi: 10.29408/edumatic.v6i2.6547.
- [12] A. Mulyani, D. Kurniadi, M. R. Nashrulloh, I. T. Julianto, and M. Regita, “the Prediction of Ppa and Kip-Kuliah Scholarship Recipients Using Naive Bayes Algorithm,” *J. Tek. Inform.*, vol. 3, no. 4, pp. 821–827, 2022, doi: 10.20884/1.jutif.2022.3.4.297.
- [13] A. Hadi and I. Ali, “Menentukan Kelayakan Penerima Kip Menggunakan Klasifikasi Dengan Metode Algoritma Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 366–372, 2023, doi: 10.36040/jati.v7i1.6311.
- [14] A. Triayudi, “Penerapan Algoritma C5.0 Data Mining Untuk Mengetahui Pola Kepuasan Mahasiswa Terhadap Pelayanan Akademik,” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2361, 2022, doi: 10.30865/mib.v6i4.4961.
- [15] D. B. Siswanto, D. Normawati, F. T. Industri, U. Ahmad, D. Yogyakarta, and K. Bantul, “Sistem Klasifikasi Monitoring dan Evaluasi Kelayakan Penerima Beasiswa UAD Menggunakan Algoritma Naïve Bayes Indonesia terkait Sumber Daya Manusia,” *J. Saintekom Sains, Teknol. Komput. dan Manaj.*, vol. 13, no. 2, pp. 2–6, 2023.
- [16] E. Sabrina, A. Simorangkir, N. Yakub, and A. Manalu, “Jurnal Mantik The implementation of machine learning for classifying eligible students for a scholarship at Budidarma University,” *J. Mantik*, vol. 6, no. 3, 2022.
- [17] M. H. Hasibuan and R. Rosnelly, “Optimizing the Selection of KIP Scholarship Recipients at STKIP Al Maksu Langkat using K-Means Clustering Method,” *Proceeding Medan Int. Conf. Econ. Bus.*, vol. 1, no. January, pp. 2697–2704, 2023.
- [18] D. T. Yuliana, M. I. A. Fathoni, and N. Kurniawati, “Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering,” *J. Focus Action Res. Math. (Factor M)*, vol. 5, no. 1, pp. 127–141, 2022, doi: 10.30762/f_m.v5i1.570.
- [19] A. Iskandar, “Sistem Pendukung Keputusan Kelayakan Penerima Bantuan Dana KIP Kuliah Menggunakan Metode ROC-EDAS,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 2, pp. 856–864, 2022, doi: 10.47065/bits.v4i2.2265.
- [20] A. Bengnga and R. Ishak, “Penerapan XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah,” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 2, pp. 192–196, 2023, doi: 10.37905/jjee.v5i2.20253.
- [21] R. Al Ghani, A. F. Winanda, and M. L. Hamzah, “Sistem Pendukung Keputusan Pemberian Beasiswa KIP-K (Studi Kasus: ‘UIN SUSKA RIAU’),” *Pros. Semin. Nas. ...*, pp. 236–239, 2022, [Online]. Available: <http://ojs.udb.ac.id/index.php/Senatib/article/download/1892/1482>.
- [22] S. Aprilia, E. Elmayati, and A. Sobri, “Data Mining Penerapan Metode K-Means Clustering Untuk Mengelompokkan Penerima Bantuan Kartu Indonesia Pintar (Studi Kasus ...,” *Escaf*, pp. 1095–1100, 2023, [Online]. Available: <https://semnas.univbinainsan.ac.id/index.php/escaf/article/view/487%0Ahttps://semnas.univbinainsan.ac.id/index.php/escaf/article/download/487/320>.

- [23] H. Di Kesuma and ul Hamidani, "Penerapan data mining menggunakan algoritma k-means clustering dalam pengelompokan penerima beasiswa KIP kuliah," *J. Ilm. Bin. STMIK Bina Nusant. Jaya*, vol. 05, no. 01, pp. 2657–2117, 2023.
- [24] K. Anam, B. Nurhakim, and C. Juliane, "Komparasi Algoritma Klasifikasi Data Mining Menggunakan Optimize Selection untuk Peminatan Program Studi," *Build. Informatics, Technol. Sci.*, vol. 4, no. 2, pp. 606–613, 2022, doi: 10.47065/bits.v4i2.2160.
- [25] Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, and Ricky Perdana Kusuma, "Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) Kuliah Menggunakan Naïve Bayes Classifier," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 193–199, 2022, doi: 10.30656/jsii.v9i2.4376.
- [26] M. Sompaa and R. Ishak, "Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar (KIP) Kuliah Metode K-Means," *J. Ilm. Ilmu Komput. Banthayo Lo Komput.*, vol. 1, no. 2, pp. 65–71, 2022, doi: 10.37195/balok.v1i2.175.
- [27] U. S, "Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3411.
- [28] F. Nuraeni, D. Kurniadi, and G. Fauzian Dermawan, "Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma K-Means++," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 437–443, 2023, doi: 10.32736/sisfokom.v11i3.1439.
- [29] A. Muhadi and A. Octaviano, "Penerapan Data Mining Untuk Prediksi Hasil Keuntungan Lelang Mesin X-Ray Tahun 2020 Dengan Metode K-Nearest Neighbor (Studi Kasus : PT.Ramadika Mandiri)," *J. Inform. MULTI*, vol. 1, no. 2, pp. 2985–8860, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim126>.
- [30] R. Rahmadini, Enjel Erika LorencisLubis, Aji Priansyah, Yolanda R.W.N, and Tuti Meutia, "Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor," *J. Mhs. Akunt. Samudra*, vol. 4, no. 4, pp. 223–235, 2023, doi: 10.33059/jmas.v4i4.7074.
- [31] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *J. Inform. MULTI*, vol. 1, no. 3, pp. 200–214, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim203>.
- [32] N. S. Fauziah and R. D. Dana, "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.
- [33] N. Attamami, A. Triayudi, and R. T. Aldisa, "Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 2, pp. 262–269, 2023, doi: 10.35870/jtik.v7i2.756.
- [34] Q. A'yuniyah, W. Elvira, N. Nazira, I. Ambarani, S. F. Intan, and D. Ramadhani, "Analisa Algoritma Naïve Bayes Classifier (NBC) Untuk Prediksi Penjualan Alat Kesehatan," *IJIRSE Indones. J. Inform. Res. Softw. Eng. J.*, vol. 3, no. 2, pp. 119–126, 2023.
- [35] A. E. Gumanti, Taslim, S. Handayani, and D. Toresa, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Topik Skripsi Mahasiswa di Fakultas Ilmu Komputer," *JITaCS J. Inf. Technol. Comput. Sci.*, vol. 1, no. 2, pp. 44–50, 2022.
- [36] S. Z. H. Rukmana, A. Aziz, and W. Harianto, "Optimasi Algoritma K-Nearest Neighbor (Knn) Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 439–445, 2022.
- [37] H. W. Azizah, O. Nurdiawan, G. Dwilestari, and E. Tohidi, "Klasifikasi Pemberian Bantuan Umkm Cirebon Dengan Menggunakan Algoritma K-Nearest Neighbor," vol. 3, no. 3, pp. 110–115, 2022, doi: 10.47065/josyc.v3i3.1392.
- [38] A. Y. Simanjuntak, I. S. E. S. Simatupang, and A. Anita, "IMPLEMENTASI DATA MINING MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER UNTUK DATA KENAIKAN PANGKAT DINAS KETENAGAKERJAAN KOTA MEDAN," *J. Sci. Soc. Res.*, vol. 5, no. 1, p. 85, 2022, doi: 10.54314/jssr.v5i1.804.
- [39] E. Y. Kodratillah, Daririn, and C. Naya, "PENERAPAN DATA MINING UNTUK PREDIKSI KELULUSAN SISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES PADA SMK GARUDA," *J. Teknol. Pelita Bangsa*, vol. 12, no. 4, pp. 33–40, 2021.