

Klasifikasi Kelayakan Penerima Kartu Jakarta Pintar (KJP) Dengan Seleksi Fitur *Backward Elimination* Menggunakan Algoritma Klasifikasi Naïve Bayes

Jamalul Ikhsan¹, Catur Nugrahaeni., M.Kom², Desta Sandy Prasvita, S.Komp., M.Kom.³
Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450
jamaluli@upnvj.ac.id¹, catur,nugrahaeni@upnvj.ac.id², desta.sandya@upnvj.ac.id³

Abstrak. Kartu Jakarta Pintar atau biasa disebut sebagai KJP merupakan program yang diciptakan oleh pemerintah provinsi DKI Jakarta untuk memberikan bantuan pendidikan di wilayah DKI Jakarta, bantuan program pendidikan ini diharapkan dapat mengurangi tingkat putus sekolah yang terjadi dikalangan masyarakat kelas menengah kebawah. Tentunya setiap program memiliki sistem seleksi seperti halnya pada program Kartu Jakarta Pintar ini, pemerintah melakukan proses seleksi untuk memastikan masyarakat yang pantas untuk mendapatkan bantuan dan yang tidak. Dalam proses seleksi penerimaan terkadang terdapat suatu kelalaian yang dilakukan oleh pihak operator dalam melakukan input data sehingga yang seharusnya masyarakat bisa mendapatkan bantuan tetapi karena adanya kelalaian tersebut menjadi tidak dapat bantuan, selain itu adanya proses seleksi yang bersifat subyektif sehingga bantuan yang diberikan oleh pemerintah ini sering tidak tepat sasaran. Maka dari itu, penelitian ini bertujuan untuk membuat suatu sistem yang dapat melakukan langsung menentukan kelayakan penerima kartu jakarta pintar dengan tepat dan akurat menggunakan metode algoritma klasifikasi naïve bayes dan seleksi fitur backward elimination untuk menentukan fitur yang berperan penting dalam penentuan kelayakan penerima kartu jakarta pintar. Dataset didapatkan dari pihak sekolah dengan jumlah sebesar 158 data calon peserta penerima. Berdasarkan hasil penelitian ini, bahwa akurasi yang didapatkan oleh model dan setelah dilakukan perhitungan akurasi menggunakan confusion matrix terhadap model diperoleh dengan akurasi sebesar 90.625%, namun dengan bantuan seleksi fitur backward elimination, akurasi dapat meningkat cukup signifikan menjadi 96.875%.

Kata Kunci: Klasifikasi, Kartu Jakarta Pintar, Naïve Bayes, Machine Learning, Data Mining.

1 Pendahuluan

1.1 Latar Belakang.

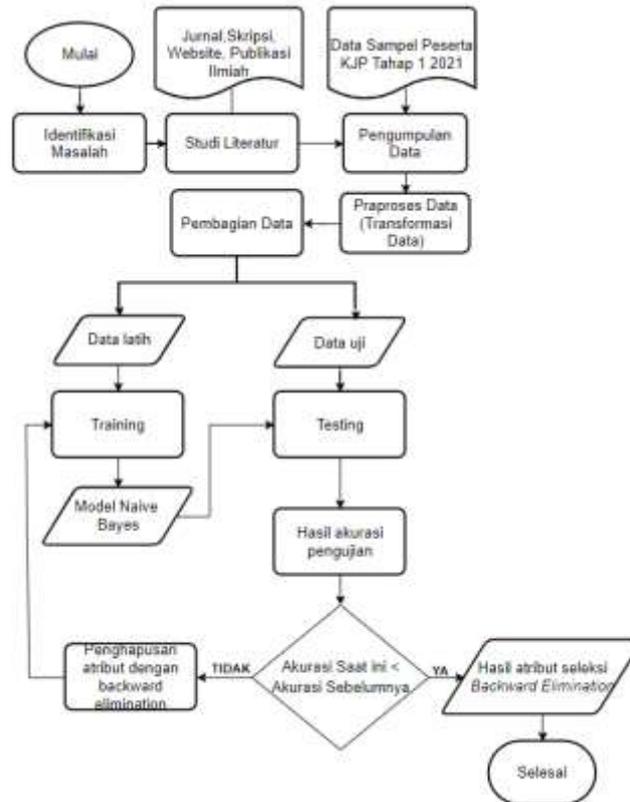
Pendidikan begitu penting dalam kehidupan manusia, hal tersebut diperjelas lagi dalam penelitian yang dilakukan oleh Sofya mengatakan bahwa pendidikan merupakan suatu kebutuhan pokok pada manusia yang dibutuhkan untuk bisa mengembangkan potensi dan bakat dalam diri yang dimilikinya serta juga memiliki kemampuan yang diperlukan untuk memperbaiki dari kualitas kehidupannya. Pendidikan juga merupakan jalan yang dapat memutus rantai dari kemiskinan, dengan sumber daya yang baik dan berkualitas melalui pendidikan maka tingkat produktivitas pastinya akan meningkat dan menjadi jalur untuk memiliki pendapatan. Dengan adanya dari penghasilan tersebut maka seseorang dapat memungkinkan untuk memiliki tabungan dan investasi lainnya [1].

Saat ini pendidikan di Indonesia belum tersebar secara merata di seluruh wilayah Indonesia diungkapkan [2] dalam penelitian lain permasalahan tersebut disebabkan karena faktor kemiskinan, kurangnya kualitas sumber daya manusia, rendahnya dari mutu pendidikan di Indonesia, rendahnya pelayanan pendidikan, rendahnya kemampuan literasi dari anak-anak Indonesia. Berdasarkan data yang dikumpulkan [3] bahwa terdapat anak putus sekolah di Indonesia sepanjang pada tahun 2019 dengan angka sebesar 4,3 juta jiwa siswa Indonesia yang putus sekolah di berbagai jenjang sekolah. Angka tersebut sebesar sekitar 6 persen dari keseluruhan usia anak sekolah yaitu sebesar 53 juta, didapatkan bahwa sebesar 54 persen anak yang putus sekolah karena masalah ekonomi sehingga mereka tidak memiliki biaya untuk melanjutkan sekolah ke tingkat yang lebih tinggi lagi.

Dalam analisis penelitian lainnya [4], mengatakan bahwa walaupun Provinsi DKI Jakarta merupakan kota

yang metropolitan, akan tetapi pada kenyataannya jika dilihat dari segi pendidikannya masih belum merata. Pendidikan di wilayah DKI Jakarta ini masih jauh dari harapan yang sebenarnya karena masih banyak anak putus sekolah karena bermasalah pada keterbatasannya kemampuan orang tua dalam memenuhi biaya pendidikan. Dan untuk mewujudkan proses pemerataan pendidikan di wilayah DKI Jakarta dengan program wajib belajar 12 tahun, pemerintah DKI Jakarta membuat suatu kebijakan dengan mengeluarkan suatu program yaitu Kartu Jakarta Pintar (KJP). Selain itu juga dalam penelitian lain [5], mengatakan bahwa dalam proses penerimaan kartu Jakarta Pintar ini masih bersifat subyektif. Hal tersebut tentu saja tidak efektif, oleh karena itu dalam penelitian ini dilakukan proses klasifikasi data untuk mengatasi permasalahan di atas dengan menggunakan algoritma klasifikasi naïve bayes.

2 Metodologi Penelitian



Gambar 1. Flowchart Metodologi Penelitian

2.1 Identifikasi Masalah

Pada bagian identifikasi masalah ini dilakukan pengamatan dan mencari fenomena permasalahan berdasarkan dari lingkungan sekitar atau permasalahan yang sedang terjadi pada masa saat ini. Setelah dilakukan suatu pengamatan maka didapatkan bahwa saat ini pemerintah provinsi DKI Jakarta memberikan bantuan pendidikan berupa program KJP, akan tetapi KJP ini masih terdapat beberapa permasalahan yang terjadi pada saat proses seleksi untuk melakukan penyaringan antara masyarakat yang miskin atau kurang mampu dengan masyarakat yang memang mampu secara finansial. Dalam proses seleksi masih kurang tepat sasaran karena proses yang bersifat subyektif [5]. Selain itu adanya kelalaian dari pihak manusia dalam kesalahan melakukan proses input data sehingga proses seleksi dalam penentuan kelayakan penerima KJP ini menjadi terhambat.

2.2 Studi Literatur

Pada kegiatan penelitian ini studi literatur yang dilakukan adalah mengumpulkan beberapa literatur paper dan informasi penelitian yang terkait dengan tema pada penelitian ini, sumber-sumber yang digunakan dalam mengumpulkan studi literatur melalui internet, seperti google scholar, website publikasi ilmiah, dan sumber lainnya yang terdapat publikasi ilmiah yang terkait dengan penelitian ini.

2.3 Pengumpulan Data

Pada bagian pengumpulan data ini, dilakukan proses pengumpulan data dengan melalui perantara petugas tata usaha SMAN 88 Jakarta yang mengurus mengenai data administrasi siswa disekolah yang mengurus segala hal mengenai administrasi siswa seperti pengumpulan data beasiswa, terutama data pendaftar Kartu Jakarta Pintar atau KJP, data yang didapatkan dari petugas tata usaha data pada periode pendaftaran KJP tahap 1 tahun 2021 yang terjadi pada tanggal 15 sampai 22 Maret 2021. Data ini diambil dengan menggunakan google form karena siswa tidak diperbolehkan untuk mengunjungi sekolah karena tidak adanya kegiatan belajar dan mengajar di sekolah dan demi menjaga protokol kesehatan sehingga siswa mengisi berkas dan mengirimkan kelengkapan data melalui google form yang sudah disediakan oleh pihak sekolah. Dataset yang didapatkan dalam penelitian ini berupa dokumen dengan format excel. Dalam dataset tersebut terdapat 158 baris data dan 37 kolom data. Dengan fitur variable independent/bebas sejumlah 36 fitur dan 1 fitur variable dependen/terikat.

2.4 Praproses Data

Tahapan Pre-Processing dalam penelitian ini yaitu dilakukan proses transformasi data, dengan melakukan perubahan data terhadap data kategorikal menjadi bentuk data numeric untuk mempermudah pengolahan data pada saat proses pembangunan pemodelan menggunakan algoritma naïve bayes.

2.5 Pembagian Data

Pada penelitian ini data dilakukan pembagian menjadi dua bagian yang terdiri dari data pelatihan dan data pengujian. Dalam penelitian ini data dibagi secara proporsional sebesar 80% untuk data pelatihan dan sisanya yaitu 20% untuk data pengujian. Proses pembagian data pada penelitian ini menggunakan bantuan library train test split dari sklearn. Setelah dilakukan pembagian data berikutnya data akan masuk kedalam proses pelatihan untuk dibangun sebuah pemodelan terhadap data tersebut.

2.6 Data Latih

Pada proses pelatihan dilakukan dengan penggunaan data pelatihan sebesar 80% data dari keseluruhan data yang akan diproses. Dalam proses pelatihan ini, data yang dilatih kemudian akan dilakukan juga proses secara bersamaan dengan pemilihan fitur menggunakan fitur seleksi backward elimination untuk mendapatkan atribut terbaik dan menghasilkan hasil nilai akurasi yang terbaik juga, setelah dilakukan pemilihan fitur data pelatihan ini masuk kedalam proses pelatihan model dengan Algoritma Naïve Bayes, dan akan menghasilkan model akhir.

2.7 Training

Proses training data latih setelah pembagian data dilakukan, data latih akan dimasukkan kedalam proses pelatihan dengan sebesar 80% data dari jumlah keseluruhan data sebesar 126 data. Dalam proses pelatihan ini dilakukan untuk membangun pemodelan naïve bayes menggunakan data latih tersebut, yang nantinya pemodelan akan dilakukan pencocokan terhadap data uji pada proses testing berikutnya.

2.8 Model Naïve Bayes

Untuk pembangunan model naïve bayes dihasilkan dari persamaan berikut ini;

$$P(\mathbf{X} | C_i) = \prod_{k=1}^n P(x_k | C_i) = P(x_1 | C_i) \times P(x_2 | C_i) \times \dots \times P(x_n | C_i)$$

Keterangan dari persamaan diatas sebagai berikut :

$P(X|C_i)$ = Suatu probabilitas kelas label dari data X.

$P(X_k|C_i)$ = Suatu probabilitas kelas label dari masing – masing data X1 , X2, dst.

Persamaan diatas untuk asumsi bahwa variabel tidak saling terikat (Conditionally Independent). Bila $P(\mathbf{X}|C_i)$ dapat diketahui melalui perhitungan di atas, maka kelas label dari data sampel X adalah kelas label yang memiliki nilai $P(\mathbf{X}|C_i) * P(C_i)$ maksimum.

Pemodelan naïve bayes telah dibangun dan akan dilakukan proses pengujian atau testing terhadap data uji untuk mengetahui tingkat hasil performa pada pemodelan yang telah ada.

2.9 Data Uji

Pada proses pengujian ini dilakukan dengan menggunakan data sebesar 20% dari keseluruhan data yang digunakan. Secara bersamaan didalam proses pengujian ini dilakukan dengan menguji secara langsung terhadap model yang dihasilkan pada proses pelatihan sebelumnya dengan menggunakan pemodelan Algoritma Naïve Bayes dan fitur seleksi berupa backward elimination. Pengujian dilakukan untuk memastikan hasil prediksi yang didapatkan antara data yang sudah dilatih dengan data uji kemudian didapatkan hasil nilai akurasi dari proses pengujian tersebut.

2.10 Testing

Proses testing data uji setelah pembagian data dilakukan, data uji akan dimasukkan kedalam pemodelan yang sudah dibangun pada saat proses pelatihan data sebelumnya dengan data uji sebesar 20% dari jumlah keseluruhan data sebesar 32 data. Dalam proses pengujian ini dilakukan pencocokan hasil antara data aktual dengan data prediksi sehingga akan menghasilkan akurasi dari pemodelan naïve bayes tersebut.

2.11 Seleksi Fitur *Backward Elimination*

Pada tahapan ini seleksi fitur dilakukan untuk memilih beberapa fitur pada dataset yang memiliki pengaruh penting dalam meningkatkan hasil performa pada pemodelan yang dibangun atau meminimalisir adanya error terhadap akurasi dari pemodelan. Seleksi fitur dalam tahapan ini terdiri dari beberapa proses didalamnya, yaitu semua fitur dalam dataset dipisahkan terlebih dahulu antara variabel bebas dan variabel terikat, hanya variabel bebas saja yang digunakan dalam proses seleksi fitur ini. Fitur dengan variabel bebas tersebut dimasukan kedalam pengolahan pemodelan naïve bayes, pemodelan yang terbentuk dengan beberapa fitur yang digunakan akan dilakukan prediksi untuk mengetahui hasil akurasi yang didapatkan oleh model tersebut, apabila didapatkan hasil akurasi yang tidak lebih baik dari sebelumnya maka beberapa fitur tersebut akan dihilangkan sehingga hanya tersisa fitur yang dapat menghasilkan tingkat performa akurasi pemodelan yang baik saja.

2.12 Hasil Akurasi Pengujian

Pada tahapan hasil akurasi pengujian pada model ini dilakukan perhitungan akurasi dari data pengujian terhadap hasil yang didapatkan dari proses pemodelan yang dibangun pada saat proses pelatihan sebelumnya. Pada penelitian ini untuk mengetahui akurasi pemodelan atau melihat hasil performa pada model menggunakan Confusion Matrix dan dilihat berdasarkan dari nilai akurasi, presisi, dan recall yang dihasilkan oleh model. Akurasi (1), Presisi (2), Recall(3).

Tabel 1 Tabel Confusion Matrix

| Kelas Aktual | Kelas Prediksi | |
|--------------|-----------------|-----------------|
| | Layak | Tidak Layak |
| Layak | (Benar Positif) | (Salah Negatif) |
| Tidak Layak | (Salah Positif) | (Benar Negatif) |

Berdasarkan dari tabel diatas maka didapatkan hasil akurasi, presisi, dan recall dengan rumus matematis berikut ini;

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots (1)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \dots (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \dots (3)$$

3 Hasil Penelitian

3.1 Data

Dalam tahapan ini data yang akan dilakukan proses pengolahan data yang didapatkan melalui pihak sekolah , data tersebut memiliki tipe format excel atau xlsx, dari data tersebut terdiri dari sebanyak 158 baris data dan 36 kolom data variabel bebas serta 1 kolom variable target atau kelas, dan masing-masing kolom data variabel memiliki nilai atribut yang berbeda-beda tentunya

Tabel 2 Tabel Variabel, Tipe, dan Atribut Data

| No | Variable | Tipe Data | Fitur Alias | Nilai Atribut Data |
|----|---|-----------|-------------|---|
| 1 | Kepala rumah tangga memiliki pendidikan SMA kebawah | Kategorik | v1 | (Ya, Tidak) |
| 2 | Pendidikan anggota rumah tangga yang sudah tidak bersekolah | Kategorik | v2 | (Tidak punya ijazah, SD/ sederajat, SMP/ sederajat, SMA/ sederajat, S2/S3, D1/D2/D3, D4/S1) |
| 3 | Apakah ada ibu hamil dalam anggota rumah tangga | Kategorik | v3 | (Ya, Tidak) |
| 4 | Apakah ada anak usia sekolah (6-18 tahun) dalam anggota rumah tangga | Kategorik | v4 | (Ya, Tidak) |
| 5 | Apakah ada anggota lanjut usia (usia lebih dari 65 tahun) dalam rumah tangga | Kategorik | v5 | (Ya, Tidak) |
| 6 | Apakah ada anggota disabilitas/penyandang cacat dalam rumah tangga | Kategorik | v6 | (Ya, Tidak) |
| 7 | Apakah ada anggota yang bekerja sebagai PNS/TNI/POLRI/DPR dalam rumah tangga | Kategorik | v7 | (Ya, Tidak) |
| 8 | Kepala Rumah Tangga mempunyai pekerjaan tidak tetap/PHK | Kategorik | v8 | (Ya, Tidak) |
| 9 | Kepala rumah tangga memiliki penyakit kronis yang menyulitkan untuk bekerja | Kategorik | v9 | (Ya, Tidak) |
| 10 | Salah Satu/Kedua Orang Tua Kandung Meninggal Dunia | Kategorik | v10 | (Ya, Tidak) |
| 11 | Ada anak umur 7-18 tahun yang tidak sekolah karena alasan ekonomi | Kategorik | v11 | (Ya, Tidak) |
| 12 | Ayah/ibu pegawai/karyawan swasta/wiraswatawan berpenghasilan cukup | Kategorik | v12 | (Ya, Tidak) |
| 13 | Ayah/ibu pegawai/karyawan swasta/wiraswatawan berpenghasilan cukup (kategori penghasilan) | Kategorik | v13 | (a, b, c, d) |
| 14 | Jumlah kepala keluarga dalam rumah tangga | Numerik | v14 | Bilangan Bulat |
| 15 | Status kepemilikan Tempat Tinggal | Kategorik | v15 | (Milik Sendiri, Kontrak/Sewa Bebas Sewa, Dinas, Lainnya) |
| 16 | Luas lantai bangunan tempat tinggal kurang dari 8 m2 | Kategorik | v16 | (Ya, Tidak) |
| 17 | Bahan bangunan utama atap rumah terluas | Kategorik | v17 | (Beton/Genteng Beton, Genteng Metal, Asbes, Sirap, Jerami/ijuk/daun, daunan/rumbia Genteng keramik, Genteng tanah liat, Seng, Bambu, Lainnya) |
| 18 | Bahan bangunan utama dinding terluas | Kategorik | v18 | (Tembok, Plesteran Anyaman, Bambu/Kawat, Batang Kayu Anyaman bambu, Bambu, Kayu, Lainnya) |
| 19 | Bahan bangunan utama lantai terluas | Kategorik | v19 | (Marmer/Granit, Keramik, Parket/Vinil/Permadani, Ubin/Tegel/Teras, Kayu/Papan Kualitas Tinggi, Semen/Bata Merah, Kayu/Papan Kualitas |

| | | | | |
|----|---|-----------|-----|--|
| | | | | Rendah, Bambu, Tanah, Lainnya) |
| 20 | Penggunaan fasilitas tempat buang air besar | Kategorik | v20 | (Sendiri, Bersama, Umum, Tidak Ada) |
| 21 | Sumber penerangan utama rumah tangga | Kategorik | v21 | (Bukan listrik, Listrik non PLN, Listrik PLN) |
| 22 | Sumber penerangan utama rumah tangga (kategori daya listrik rumah) | Kategorik | v22 | (a, b, c, d, e) |
| 23 | ID Pelanggan PLN | Numerik | -- | 11 Digit Bil Bulat |
| 24 | Sumber air minum berasal dari sumur | Kategorik | v23 | (Ya, Tidak) |
| 25 | Bahan bakar untuk memasak | Kategorik | v24 | (Listrik, Gas >3Kg, Gas 3Kg, Gas Kota/Biogas, Minyak Tanah, Briket, Arang, Kayu Bakar, Tidak Memasak di rumah) |
| 26 | Apakah terdapat perubahan kondisi sosial ekonomi rumah tangga dari tahun sebelumnya ? | Kategorik | v25 | (Ya, Tidak) |
| 27 | Apakah rumah tangga memiliki mobil | Kategorik | v26 | (Ya, Tidak) |
| 28 | Memiliki kapal motor | Kategorik | v27 | (Ya, Tidak) |
| 29 | Memiliki perahu motor lebih dari atau sama dengan dua | Kategorik | v28 | (Ya, Tidak) |
| 30 | Memiliki sepeda motor lebih dari dua | Kategorik | v29 | (Ya, Tidak) |
| 31 | Memiliki telepon rumah (bukan hp) | Kategorik | v30 | (Ya, Tidak) |
| 32 | Memiliki AC | Kategorik | v31 | (Ya, Tidak) |
| 33 | Memiliki tabung gas lebih dari 5,5 kg | Kategorik | v32 | (Ya, Tidak) |
| 34 | Anak memiliki HP berharga lebih Rp1.000.000 | Kategorik | v33 | (Ya, Tidak) |
| 35 | Rumah mewah | Kategorik | v34 | (Ya, Tidak) |
| 36 | Usaha rumah kontrakan atau kos | Kategorik | v35 | (Ya, Tidak) |
| 37 | Status | Kategorik | v36 | (Layak, Tidak Layak) |

3.2 Praproses Data

3.2.1 Cleaning Data

Proses pembersihan data pada tahapan ini yaitu hanya melakukan penghapusan kolom pada data yang tidak diperlukan, dan seperti yang sudah dipaparkan pada proses analisis data eksplorasi pada proses sebelumnya bahwa pada datasets tersebut tidak ditemukan adanya nilai kosong/missing value. Namun pada data tersebut terdapat satu kolom yang memiliki tipe data int64 yaitu pada kolom "ID Pelanggan PLN". Pada kolom ini dilakukan penghapusan karena kolom ini hanya berupa data identitas nomor id pelanggan pln saja dan kolom ini tidak memiliki makna sama sekali, sehingga dilakukan proses penghapusan pada kolom tersebut. Setelah dilakukan proses penghapusan maka datasets tersebut menjadi 36 kolom data dari semula 37 kolom data.

3.2.2 Transformasi Data

Dalam proses konversi pelabelan ini menggunakan metode label encoding secara manual tanpa library dengan bantuan fungsi mapping dan lambda pada bahasa python yang terdiri dari kunci (key) dan nilai (value).

Penerapan pelabelan tersebut untuk mengubah data kategorik menjadi numerik untuk sejumlah 34 data kategorik pada variabel independent.

Penerapan pelabelan tersebut untuk mengubah data kategorik menjadi numerik untuk sejumlah 34 data kategorik pada variabel independent.

3.3 Pembagian Data

Pembagian data dibagi secara proporsional sebesar 80% untuk data pelatihan yaitu sejumlah 126 data dan

sisanya 20% untuk data pengujian yaitu sejumlah 32 data. Proses pembagian data pada penelitian ini menggunakan library train test split dari sklearn. Setelah dilakukan pembagian data berikutnya data akan masuk kedalam proses pelatihan untuk dibangun sebuah pemodelan terhadap data tersebut.

3.4 Pemodelan

Pemodelan pada tahap ini akan dibangun setelah data melewati beberapa proses sebelumnya yaitu praproses data yang meliputi dari melakukan analisis terhadap data, dan karakteristik data, kemudian pembersihan data, dan transformasi data untuk membuat data menjadi bentuk yang seragam agar mudah diolah kedalam algoritma machine learning. Data yang diolah akan dibangun menggunakan algoritma klasifikasi naïve bayes dengan jenis multinomial naïve bayes dikarenakan pada data hampir seluruh nya memiliki jenis data yang kategorikal atau data teks.

3.5 Evaluasi Hasil

Pada bagian ini dilakukan evaluasi hasil menggunakan confusion matrix. Hasil confusion matrix dari keseluruhan data dalam dataset yang berjumlah total 158 data, yang terdiri dari 87 kelas layak dan 71 kelas tidak layak. Pembagian data training sebesar 80% atau 126 data dan data testing 20% atau 32 data menghasilkan performa model berdasarkan Tabel confusion matrix pada Tabel 4.79;

Tabel 3 Tabel Confusion Matrix

| Kelas Aktual | Kelas Prediksi | |
|-----------------|--------------------|-------------------|
| | Layak | Tidak Layak |
| Layak | (Benar Positif) 20 | (Salah Negatif) 0 |
| Tidak Layak | (Salah Positif) 3 | (Benar Negatif) 9 |

Berdasarkan dari tabel diatas maka didapatkan hasil akurasi, presisi, dan recall dengan rumus matematis berikut ini;

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% = \frac{29}{32} \times 100\% = 0.90625 \times 100\% = 90.625\% \dots (1)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% = \frac{20}{20+3} \times 100\% = 0.8697 \times 100\% = 86.97\% \dots (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% = \frac{20}{20+0} \times 100\% = 1 \times 100\% = 100\% \dots (3)$$

3.6 Seleksi Fitur Backward Elimination

Hasil seleksi fitur yang dilakukan oleh backward elimination didapatkan bahwa untuk tingkat hasil akurasi tertinggi sebesar 0.96875 atau 96.875%, hasil iterasi dengan tingkat akurasi yang konstan di angka 96.875% terakhir diperoleh pada iterasi ke-26 dengan 9 fitur yang didapatkan yaitu sebagai berikut ini ;

1. Kepala Rumah Tangga mempunyai pekerjaan tidak tetap/PHK
2. Ayah/ibu pegawai/karyawan swasta/wiraswatawan berpenghasilan cukup (kategori penghasilan)
3. Status kepemilikan Tempat Tinggal
4. Bahan bangunan utama atap rumah terlalu
5. Sumber penerangan utama rumah tangga (kategori daya listrik rumah)
6. Memiliki AC
7. Memiliki tabung gas lebih dari 5,5 kg
8. Anak memiliki HP berharga lebih Rp1.000.000
9. Rumah mewah

Fitur yang diperoleh pada tabel dibawah ini merupakan fitur yang telah diseleksi oleh backward elimination sehingga fitur tersebut yang memiliki peran cukup penting dalam memengaruhi hasil akurasi dan fitur

tersebut dapat digunakan sebagai parameter dalam menentukan proses penerimaan kelayakan dari calon pengguna kartu jakarta pintar.

Tabel 4 Tabel Seleksi Fitur Backward Elimination

| Iterasi Ke- | Fitur | Fitur Ter-eliminasi | Akurasi |
|-------------|---|--|---------|
| 1 | V1,V2,V3,V4,V5,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V2,V23,V24,V25,V26,V27,V28,V29,V30,V31,V33,V34,V35 | V32 | 93.75% |
| 2 | V1,V2,V3,V4,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V29,V30,V31,V32,V33,V34 | V5, V35 | 93.75% |
| 3 | V1,V3,V5,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2,V4, V29,V35 | 96.87% |
| 4 | V1,V3,V5,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V4, V29, V35 | 96.87% |
| 5 | V1,V5,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2,V3, V4, V29, V35 | 96.87% |
| 6 | V1,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4,V5, V29, V35 | 96.87% |
| 7 | V1,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5,V6, V29, V35 | 96.87% |
| 8 | V1,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6,V7, V29, V35 | 96.87% |
| 9 | V1,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7,V9, V29, V35 | 96.87% |
| 10 | V1,V8,V10,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9,V11, V29, V35 | 96.87% |
| 11 | V1,V8,,V10,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9,V11,V12, V29, V35 | 96.87% |
| 12 | V1,V7,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V8, V9,V10,V11, V12, V29, V35 | 96.87% |
| 13 | V1,V8,V13,V15,V16,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12,V14, V29, V35 | 96.87% |
| 14 | V1,V8, V13,V15,V17,V18,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14,V16, V29, V35 | 96.87% |
| 15 | V1,V8, V13,V15,V17,V19,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16,V18, V29, V35 | 96.87% |
| 16 | V1,V8,V13,V15,V17,V20,V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18,V19, V29, V35 | 96.87% |
| 17 | V1,V8, V13,V15,V17, V21,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19,V20, V29, V35 | 96.87% |
| 18 | V1,V8, V13,V15,V22,V23,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19, V20,V21, V29, V35 | 96.87% |
| 19 | V1,V8,V13,V15,V17,V22,V24,V25,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19, V20, V21,V23, V29, V35 | 96.87% |
| 20 | V1,V8, V13,V15,V17, V22,V24,V26,V27,V28,V30,V31,V32,V33,V34 | V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19, V20, V21, V23,V25, V29, V35 | 96.87% |
| 21 | V8,V9,V13,V15,V22,V24,V26,V27,V28,V30,V31,V32,V33,V34 | V1, V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19, V20, V21, V23, V25, V29, V35 | 96.87% |
| 22 | V8, V13,V15,V17,V22,V24,V27,V28,V30,V31,V32,V33,V34 | V1, V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19, V20, V21, V23, V25,V26, V29, V35 | 96.87% |

| | | | |
|----|--|---|--------|
| 23 | V8,V13 ,V15 ,V22, V27,V28 ,V30,V31,V32,V33,V34 | V1, V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19, V20, V21, V23,V24, V25, V26, V29, V35 | 96.87% |
| 24 | V8,V15,V17,V22,V28,V30,V31,V32,V33,V34 | V1, V2, V3, V4, V5, V6, V7, V9, V10 ,V11, V12, V14, V16, V18,V19, V20, V21, V23, V24, V25, V26,V27, V29, V35 | 96.87% |
| 25 | V8 ,V13 ,V15 ,V17,V22, V30,V31,V32,V33,V34 | V1, V2, V3, V4, V5, V6, V7, V9, V10,V11, V12, V14, V16, V18, V19, V20, V21, V23, V24, V25, V26, V27,V28, V29, V35 | 96.87% |
| 26 | V8,V13 ,V15 ,V22,V31,V32,V33,V34 | V1, V2, V3, V4, V5, V6, V7, V9, V10 ,V11, V12, V14, V16, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29,V30, V35 | 96.87% |
| 27 | V1,V13 ,V15 ,V17, V22 ,V31,V32 ,V34 | V2, V3, V4, V5, V6, V7,V8, V9, V10 ,V11, V12, V14, V16, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29, V30,V33, V35 | 93.75% |
| 28 | V13 ,V15 ,V17, V22, V31,V32 ,V34 | V1, V2, V3, V4, V5, V6, V7,V8, V9, V10 ,V11, V12, V14, V16, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29, V30, V33, V35 | 93.75% |
| 29 | V13 ,V15 ,V17 ,V22 ,V31,V32 | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10 ,V11, V12, V14, V16, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29, V30, V33, V34, V35 | 93.75% |
| 30 | V13 ,V15 ,V17 , V22, V31 | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10,V11, V12, V14, V16, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29, V30,V32, V33, V34, V35 | 90.62% |
| 31 | V15 , V17,V22, V32 | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12 V13, V14, V16, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29, V30, V31, V33, V34, V35 | 84.37% |
| 32 | V15,V22, V32, | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12,V13, V14, V16,V17, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29, V30, V31, V33, V34, V35 | 84.37% |
| 33 | V22,V32 | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12, V13, V14,V15, V16, V17, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28, V29, V30, V31, V33, V34, V35 | 81.25% |
| 34 | V22,V32 | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12, V13, V14,V15, V16, V17, V18, V19, V20, V21, V23, V24, V25, V26, V27, V28,V29, V30,V31, V33,V34,V35 | 62.50% |

4 Kesimpulan & Saran

Berdasarkan dari hasil penelitian diatas, dapat disimpulkan bahwa didapatkan hasil dari kelayakan penerima kartu jakarta pintar dengan menggunakan algoritma naïve bayes dan seleksi fitur backward elimination. Penentuan kelayakan peserta kartu jakarta pintar dengan algoritma naïve bayes memerlukan beberapa tahapan diantaranya, proses pengumpulan data, pengolahan data dengan melakukan pemeriksaan terhadap data atau praproses data, pembagian data untuk digunakan dalam proses pembangunan pemodelan naïve bayes, dan dilakukan pengujian data terhadap pemodelan yang sudah dibangun untuk mendapatkan hasil performa pada pemodelan sehingga didapatkan akurasi oleh model. Kemudian seleksi fitur dilakukan apabila didapatkan hasil akurasi yang tidak lebih baik dari sebelumnya maka beberapa fitur tersebut akan dihilangkan sehingga hanya tersisa fitur yang dapat menghasilkan tingkat performa akurasi pemodelan yang baik saja. Setelah dilakukan proses perhitungan akurasi oleh confusion matrix terhadap model didapatkan dengan akurasi sebesar 90.625%, namun dengan bantuan seleksi fitur akurasi dapat meningkat cukup signifikan menjadi 96.875%. Untuk fitur yang berperang penting dalam menentukan kelayakan dari peserta kartu jakarta pintar ini ada 9 fitur yang terdiri antara lain ; Kepala Rumah Tangga mempunyai pekerjaan tidak tetap/PHK, Ayah/ibu pegawai/karyawan swasta/wiraswatawan berpenghasilan cukup (kategori penghasilan), Status kepemilikan Tempat Tinggal, Bahan bangunan utama atap rumah terluas, Sumber penerangan utama rumah tangga (kategori daya listrik rumah), Memiliki AC, Memiliki tabung gas lebih dari 5,5 kg, Anak memiliki HP berharga lebih Rp1.000.000, Rumah mewah.

Saran untuk penelitian berikutnya untuk dapat mengolah data tersebut dengan menggunakan algoritma klasifikasi yang berbeda untuk mendapatkan hasil yang efektif dan tingkat akurasi yang akurat, dapat menggunakan algoritma seperti Decision Tree, Random Forest, dan algoritma lainnya yang mendukung dalam melakukan pemrosesan data dengan jenis data teks atau kategorikal.

5 Daftar Pustaka

- [1] Sofya, R., Siwi, M. K., & Oknaryana, O. (2018). *Kondisi Sosial Ekonomi Siswa Putus Sekolah*.

- Jurnal Inovasi Pendidikan Ekonomi*, 8(2), 90. <https://doi.org/10.24036/011024540>
- [2] Handoyo, A. D. dan Z. (2019). *Faktor-faktor Penyebab Pendidikan Tidak Merata di Indonesia. Prosiding Seminar Nasional*, 1(1), 21–24. <https://bimawa.uad.ac.id/wp-content/uploads/Paper-Seminar-Nasional-2.pdf>
- [3] Kementerian Perencanaan Pembangunan Nasional atau Badan Perencanaan Pembangunan Nasional (PPN/Bappenas), (2021). *Statistik pendidikan anak seluruh wilayah Indonesia*.
- [4] Muluk, A. P., Irbah, A., Bagastama, R. M., & Belakang, L. (2019). *Analisis Implementasi Kebijakan Kartu Jakarta Pintar (KJP) Plus di DKI Jakarta. Ilmu Pemerintahan Unbraw*.
- [5] Saryoko, A. (2017). *Penentuan Penerima Kartu Jakarta Pintar Dengan Metode Analytical Hierarchy Process*. XV(1), 7–16.