

IMPLEMENTASI NAIVE BAYES UNTUK MEMPREDIKSI CALON MAHASISWA BARU YANG AKAN LANJUT MELAKUKAN PEMBAYARAN ATAU TIDAK (Studi Kasus : PMB UIN Bandung)

Raka Fajar Salingih¹, Undang Syaripudin², Yana Aditia Gerhana³

¹Teknik Informatika, Sains & Teknologi, Universitas Islam Negeri Sunan Gunung Djati Bandung, Bandung, 40614, Indonesia

²Teknik Informatika, Sains & Teknologi, Universitas Islam Negeri Sunan Gunung Djati Bandung, Bandung, 40614, Indonesia

³Matematika, Sains & Teknologi, Universitas Islam Negeri Sunan Gunung Djati Bandung, Bandung, 40614, Indonesia

1147050132@uinsgd.ac.id, undang_if@uinsgd.ac.id, yanagerhana@uinsgd.ac.id

Abstract. Sunan Gunung Djati State Islamic University Bandung has 5700 graduation quotas for new students in 2018. However, in each selection of prospective new students who resigned by not making payments beforehand, initially the graduation quota needed by the university was wasted. The Naïve Bayes algorithm is an algorithm that has an accuracy value that is quite easy to implement in the Bayes Theorem (or with Bayes rules) and has a strong (naïve) assumption of independence (incompatibility). The development method made to predict payment status is a prototype. The results of the verification obtained from the application of naïve Bayes algorithm. That is 91,811% of the data 6240 training data also test data.

Keywords : Naïve Bayes, Payment, Payment Status Prediction

Abstrak. Universitas Islam Negeri Sunan Gunung Djati Bandung memiliki 5700 kuota kelulusan untuk calon mahasiswa baru pada tahun 2018. Namun disetiap seleksi banyak calon mahasiswa baru yang mengundurkan diri dengan cara tidak melakukan pembayaran ukt semester awal sehingga terbuangnya kuota kelulusan yang sudah ditentukan oleh pihak universitas. Algoritma Naïve Bayes merupakan algoritma yang memiliki nilai akurasi yang cukup tinggi serta mudah diimplementasikan dikarenakan probabilitas sederhana yang memiliki dasar pada penerapan *Teorema Bayes* (atau dengan aturan *bayes*) dan memiliki asumsi independensi (ketidak ketergantungan) yang kuat (naif). Metode pengembangan yang dibuat untuk prediksi status pembayaran adalah prototype. Hasil akurasi yang didapat dari penerapan algoritma *naïve bayes*. yaitu 91,811% dari data 6240 data latih maupun data uji.

Kata Kunci : Naïve Bayes, Pembayaran, Prediksi Status Pembayaran

1. Pendahuluan

Seleksi Penerimaan Mahasiswa Baru adalah suatu aktivitas yang selalu menghadirkan calon mahasiswa baru yang akan berkuliah di universitas yang dilaksana pada setiap tahunnya. Dalam seleksi penerimaan mahasiswa baru, terdapat beberapa jalur pilihan masuk ke universitas, yaitu dengan mengikuti seleksi ujian tertulis atau seleksi jalur prestasi yang dimiliki calon mahasiswa baru. Seleksi ujian tertulis dibagi 3 kategori yaitu Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), Ujian Masuk Perguruan Tinggi Keagamaan Islam Negeri (UM-PTKIN) dan Ujian Mandiri (Seleksi yang dilakukan secara mandiri oleh masing-masing perguruan tinggi negeri).

Sedangkan seleksi jalur prestasi dibagi menjadi 2 yaitu Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Prestasi Akademik Nasional Perguruan Tinggi (SPAN-PTKIN) dan masih banyak seleksi-seleksi lainnya yang dilakukan oleh universitas. Dalam seleksi yang telah dilakukan oleh UIN Sunan Gunung Djati Bandung, UIN Bandung telah memberikan kuota disetiap seleksinya dengan jumlah kuota yang diberikan 5700 pada tahun 2018 [1]. Namun disetiap kuota yang telah diberikan, masih banyak calon mahasiswa baru yang telah lulus seleksi PMB tidak melakukan registrasi, yaitu berupa tidak membayar UKT semester awal, sehingga banyak terbuangnya kuota yang telah ditentukan oleh pihak UIN Sunan Gunung Djati Bandung.

Dalam hal ini Universitas Islam Negeri Sunan Gunung Djati Bandung ingin melakukan suatu prediksi untuk mengetahui calon mahasiswa baru yang melakukan pembayaran atau tidak, sehingga dapat mencari solusi untuk memperkecil jumlah calon mahasiswa yang tidak lanjut melakukan pembayaran. Maka dari itu diperlukannya suatu metode yang dapat memprediksi pengunduran diri yaitu dengan data mining. Data mining merupakan aktifitas yang melakukan pengumpulan, pemakaian data sebelumnya untuk mendapatkan pola dan hubungan dalam set data yang berukuran besar [2].

Salah satu metode data mining yaitu klasifikasi, Klasifikasi adalah teknik data mining yang memiliki kemampuan klasifikasi yang bertujuan mengambil suatu keputusan dengan memprediksi suatu kasus, berdasarkan hasil klasifikasi yang didapat [3]. Salah satu metode data mining yang dapat membantu memprediksi atas masalah calon mahasiswa yang tidak melakukan pembayaran adalah metode *Naive Bayes*. *Naive Bayes* adalah sebuah pengklasifikasian, berupa sebuah metode probabilitas dan statistik dengan memprediksi peluang di masa depan berdasarkan pengalaman sebelumnya [4] dan pengklasifikasian statistiknya dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [5].

Beberapa penelitian telah membahas tentang prediksi dengan menggunakan berbagai metode yang telah diterapkan.

Pada penelitian ini akan mengembangkan kembali penelitian sebelumnya dengan data training yang menjadi objek penelitian adalah data penerimaan mahasiswa baru tahun 2018. Maka akan dibangun aplikasi prediksi pembayaran atau tidak pada calon mahasiswa baru dengan judul **“Implementasi *Naive Bayes* Untuk Memprediksi Calon Mahasiswa Baru Yang Akan Lanjut Melakukan Pembayaran atau Tidak”**.

2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini terdiri dari dua tahap yaitu metode pengumpulan data dan metode pengembangan perangkat lunak.

2.1 Metode Pengumpulan Data

Metode yang digunakan dalam pengumpulan data pada penelitian ini terdiri dari 2 tahap yaitu :

1. Observasi

Teknik pengumpulan data dengan penelitian/peninjauan langsung terhadap permasalahan yang diambil untuk objek penelitian.

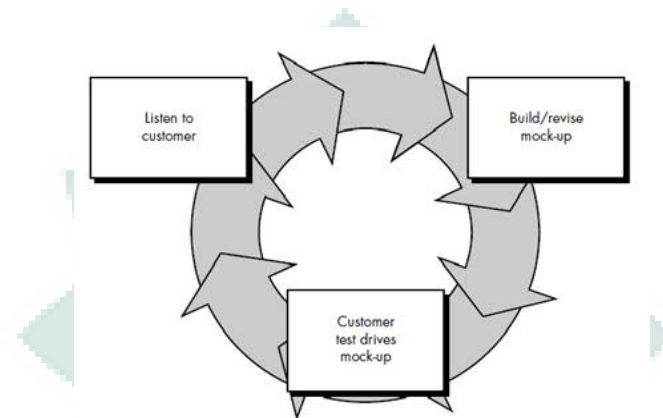
2. Studi Literatur

Pengumpulan data dengan cara mengumpulkan jurnal, literatur, paper, dan bacaan-bacaan yang ada kaitannya dengan judul penelitian.

2.2 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang akan digunakan untuk membangun aplikasi dalam penelitian ini menggunakan metode *Prototype*. *Prototype* merupakan proses pembuatan sistem perangkat lunak yang menggunakan pendekatan dalam membuat perangkat lunak, dengan proses pembuatannya dapat berinteraksi bagi pengembang dan pelanggan sehingga memudahkan proses pembuatan perangkat lunaknya [6].

Untuk proses pengulangan iteratif dari metode *Prototype* ini dapat dilihat pada gambar :



Gambar 1. Prototype Model . Listen to customer, Build/revise mockup, & Customer test driver mockup.

Prototype memiliki tiga tahap yang dapat dilakukan. Berikut penjelasan mengenai tahapan-tahapan yang terdapat pada metode *Prototype*:

1. Pengumpulan kebutuhan : pengemban dan pelanggan akan saling bertemu dan memberikan tujuan umum tentang pembuatan perangkat lunak akan bagaimana, gambaran kebutuhan yang akan datang berikutnya
2. Perancangan : perancangan ini biasanya menjadi awal pembuatan prototype, karena dilakukan secara cepat dengan memberikan rancangan yang mewakili aspek perangkat lunak yang ingin dibuat..
3. Evaluasi *Prototype* : pelanggan selalu mengevaluasi model prototype yang telah dibuat untuk digunakan dalam memperjelas kebutuhan perangkat lunak yang telah disepakati.

2.3 Pengertian Pembayaran

Menurut pasal Undang – Undang No 23 Tahun 1999 Pasal 1 Ayat 6 dengan bunyi : “Pembayaran adalah suatu sistem yang mencakup separangkat aturan, lembaga, dan mekanisme, yang digunakan untuk melaksanakan pemindahan dana guna memenuhi suatu kewajiban yang timbul dari suatu kegiatan ekonomi”. Pembayaran dapat melakukan mekanisme untuk memindahkan mata uang menjadi suatu barang, jasa atau informasi dari pembayar kepada penerima, baik secara langsung maupun melalui perbankan. Sebaliknya jika pembayar tidak melakukan pembayaran maka perpindahan mata uang menjadi satu barang, jasa atau informasi tidak akan mungkin terjadi.

2.4 Data Mining

Data Mining adalah data yang diproses menjadi informasi penting karena data tersebut sangat menarik, yang didalamnya menghasilkan informasi yang sangat bernilai [7]. Proses terjadinya *data mining* yaitu dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terhubung dari berbagai database yang jumlahnya sangat besar [8]. *Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugasnya yaitu [9] :

1. Klasifikasi

Proses membedakan kelas data dalam beberapa kategori.

2. Deskripsi
Teknik yang dipergunakan dalam menggambarkan pola dan kecenderungan yang dimiliki oleh data
3. Estimasi
Estimasi hampir sama dengan klasifikasi, namun variabel target estimasi lebih tertuju ke arah numerik dari pada ke arah kategori.
4. Prediksi
Prediksi memiliki kemiripan dengan klasifikasi dan estimasi, namun nilai dari hasil prediksi akan ada di masa yang akan datang.
5. Klaster
Klaster adalah pengelompokan sejumlah data yang memiliki kemiripan ke dalam kelompok-kelompok data.
6. Asosiasi
Teknik yang digunakan untuk mencari hubungan antar karakteristik tertentu dalam satu waktu.

Data mining juga merupakan sebuah sekumpulan data yang jumlahnya sangat besar sehingga membutuhkan proses pemilihan data yang akan memberikan informasi yang akurat.

Tahapan Proses *Data Mining* [10] :

1. *Data Selection* (Seleksi data), memilih data yang akan dijadikan proses data *mining* dari banyaknya data.
2. *Data Cleaning* (Membersihkan data) merupakan langkah pembersihan data yang memiliki atribut, nilai, dan konsiten data, dengan data yang diharapkan.
3. *Data Transformation* (Transformasi data), data yang ditransformasikan atau digabungkan untuk dijadikan data yang sesuai dengan proses data *mining*.
4. *Data Mining*, proses mencari pola atau informasi dari data yang telah dipilih dengan menggunakan teknik atau metode tertentu.
5. *Pattern Evaluation* (Pola evaluasi), pola atau informasi data yang dihasilkan dari proses *data mining* yang mencakup, apakah pola atau informasi yang telah diproses serta sesuai dengan hasil fakta atau hipotesis.
6. *Knowledge presentation* (Presentasi pengetahuan), tahapan akhir proses *data mining* yang akan ditampilkan dalam bentuk yang dapat dimengerti.

2.5 Metode Klasifikasi

Klasifikasi merupakan proses untuk mencari fungsi dan model yang dapat membedakan atau menjelaskan konsep atau kelas data dengan maksud memperkirakan kelas yang tidak diketahui dari suatu objek [11]. Klasifikasi merupakan bagian dari *data mining*, dimana *data mining* merupakan salah satu yang digunakan dalam menguraikan penemuan pengetahuan didalam *database* [12]. Contoh dari klasifikasi adalah menganalisa seseorang akan membeli *handphone* atau tidak.

2.6 Naïve Bayes

Naïve Bayes merupakan sebuah teknik prediksi yang berbassis pada probabilitas sederhana yang memiliki dasar pada penerapan *Teorema Bayes* (atau dengan aturan *bayes*) dan memiliki asumsi independensi (ketidak ketergantungan) yang kuat (naif) [13]. *Teorema Bayes* mempunyai kemampuan berupa klasifikasi serupa dengan *decision tree* dan *neural network* [12]. *Naïve Bayes* telah terbukti memiliki akurasi dan kecepatan dalam mengaplikasikan ke dalam database dengan data yang besar [12]. Secara umum, *teorema bayes* dinyatakan persamaan 1 :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Dimana :

1. *X* adalah data dengan *class* belum diketahui
2. *H* adalah hipotesis data *X* merupakan suatu *class* spesifik

3. $P(H/X)$ adalah probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)
4. $P(H)$ adalah probabilitas hipotesis H (probabilitas)
5. $P(X/H)$ adalah probabilitas hipotesis X berdasarkan kondisi H
6. $P(X)$ adalah probabilitas dari X

Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya memerlukan jumlah data latih (*data training*) yang kecil untuk memastikan estimasi parameter yang diperlukan dalam proses pengklasifikasian [4].

2.7 Confusion Matrix

Confusion matrix merupakan suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining [14] atau juga sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data yang salah diklasifikasikan [15]. Recall adalah berapa persen data kategori positif yang terklasifikasikan dengan benar oleh system, Presisi yaitu nilai yang menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif, dan Akurasi merupakan perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus [14]. Berikut rumus perhitungan recall, presisi, dan akurasi :

- a. Recall : $TP / (FN+TP) * 100\%$ (2)
- b. Presisi : $TP / (FP+TP) * 100\%$ (3)
- c. Akurasi : $(TN+TP) / (TP + TN + FP + FN) * 100\%$ (4)

Tabel 1 Confusion Matrix

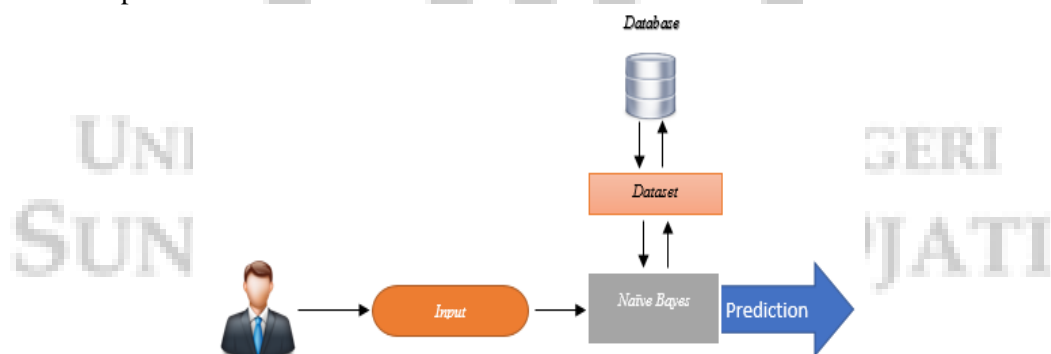
Classification	Predicted Class	
	Class = Ya	Class = Tidak
Class = Ya	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
Class = Tidak	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Dengan keterangan sebagai berikut :

1. TP = Hasil yang terklasifikasi oleh naïve bayes YA dengan class asli YA
2. FN = Hasil yang terklasifikasi oleh naïve bayes YA dengan class asli TIDAK
3. FP = Hasil yang terklasifikasi oleh naïve bayes TIDAK dengan class asli YA
4. TN = Hasil yang terklasifikasi oleh naïve bayes TIDAK dengan class asli TIDAK

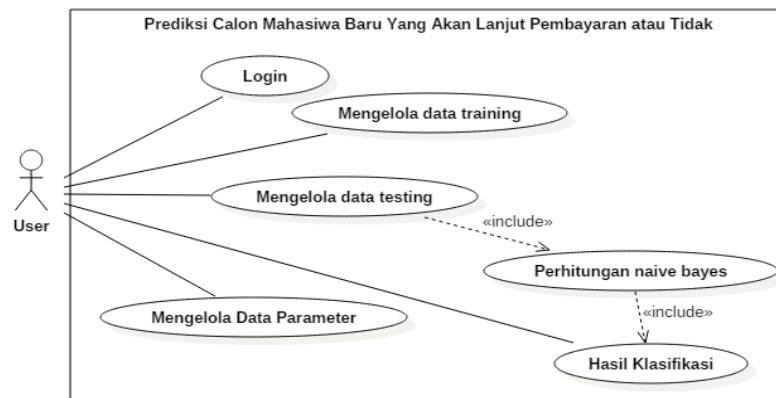
3. Hasil dan Pembahasan

3.1 Arsitektur Aplikasi



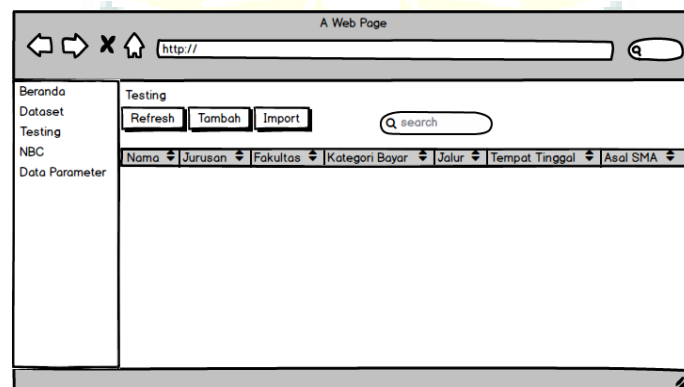
Gambar 2. Arsitektur Sistem . Arsitektur sistem menjelaskan gambaran secara umum dari sistem yang akan dibuat, dan dapat diimplementasikan sesuai kebutuhan dan tetap terarah sesuai dengan fungsionalitasnya karena sudah memiliki acuan sistem yang akan dibuat.

3.3 Use Case Diagram

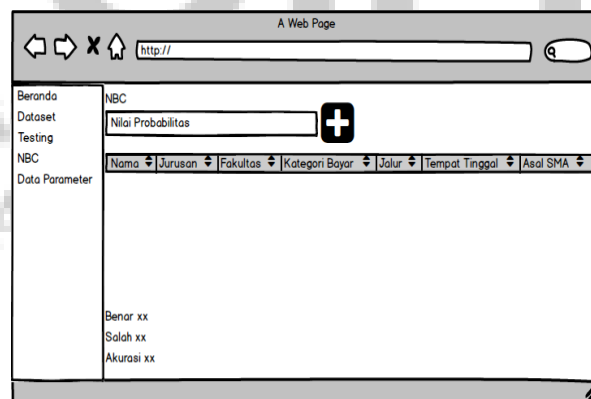


Gambar 3 . Use Case Diagram . Use case diagram merupakan diagram yang menggambarkan interaksi antara actor dengan aplikasi yang dibuat. Selain itu use case diagram mendeskripsikan fungsi yang terdapat dalam aplikasi. Dalam gambar tersebut menjelaskan fungsional

3.6 Perancangan Antarmuka



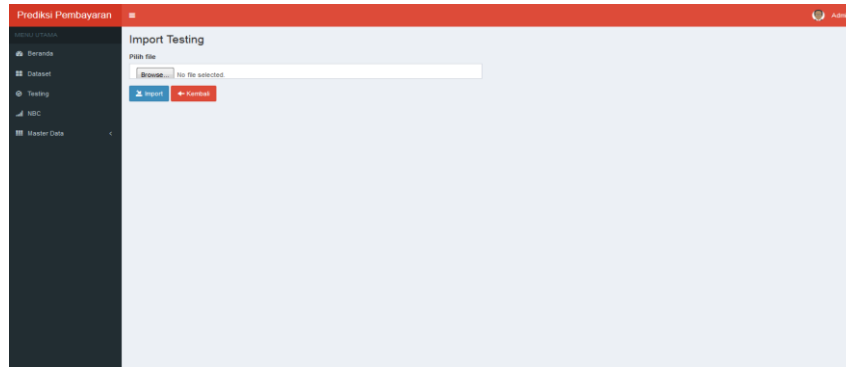
Gambar 4. Antarmuka Halaman Testing . Pada perancangan antarmuka ini pengguna dapat mengupload data testing.



Gambar 5. Antarmuka NBC. Pada antarmuka NBC, pengguna dapat melihat hasil prediksi *naive bayes*.

4. Implementasi dan Pengujian

4.1 Implementasi Antarmuka



Gambar 6. Halaman Testing . Pada perancangan antarmuka ini pengguna dapat mengupload data testing.

Atribut	Klasifikasi	Domain	Nilai
Jurusan	Bayar	Perbandingan Hadzad dan Hukum	0.0124
Jurusan	Bayar	Akuntansi Syariah	0.0171
Jurusan	Bayar	Pengembangan Masyarakat Islam	0.0288
Jurusan	Bayar	Manajemen Pendidikan Islam	0.0135
Jurusan	Bayar	Pendidikan Guru (PGSD)	0.0176
Jurusan	Bayar	Pendidikan Agama Islam	0.0476
Jurusan	Bayar	Pendidikan Matematika	0.0241
Jurusan	Bayar	Administrasi Publik	0.0435
Jurusan	Bayar	Sastra Inggris	0.0347
Jurusan	Bayar	Hukum Keluarga (Ahwal AL-Syakhshiyah)	0.0176
Jurusan	Bayar	Hukum Ekonomi Syariah (Muamalah)	0.0212
Jurusan	Bayar	Aqidah dan Fikih Islam	0.0105
Jurusan	Bayar	Hukum Tata Negara (Syasati)	0.0171
Jurusan	Bayar	Manajemen Keuangan Syariah	0.0471
Jurusan	Bayar	Pendidikan Bahasa Arab	0.0241
Jurusan	Bayar	Pendidikan Biologi	0.0147
Jurusan	Bayar	Pendidikan Bahasa Inggris	0.0239

Gambar 7. Halaman NBC. Pada antarmuka NBC, pengguna dapat melihat hasil prediksi *naïve bayes*.

4.2 Implementasi Hasil Prediksi

Implementasi hasil prediksi merupakan tahap pengujian aplikasi untuk memprediksi data apakah calon mahasiswa baru melakukan pembayaran atau tidak secara otomatis. Data training yang digunakan 6240 data dengan data status bayar 5739 dan tidak bayar 501. Adapun hasil ringkasan dapat dilihat di bawah ini :

Tabel 2 Data Uji

No	Atribut	Nilai
1	Nama	Firman
2	Jurusan	Studi Agama-Agama
3	Fakultas	USHULUDDIN
4	Kategori UKT	K2
5	Jenis Kelamin	Laki-Laki
6	Jalur	UMPTKIN
7	Tempat Tinggal	Bersama Orang Tua
8	Kendaraan Pribadi	Tidak Memiliki
9	Asal SMA	Administrasi Perkantoran
10	Pekerjaan Ibu	Tidak Bekerja
11	Gaji Ibu	0 S.D 400.000

12	Pekerjaan Ayah	Buruh (Tani/Pabrik/Bangunan)
13	Gaji Ayah	750.001 S.D 1.250.000
14	Status Bayar	?

1. Menghitung jumlah kelas/label pada kolom keterangan status bayar berdasarkan klasifikasi yang sudah dibentuk.

- 1) P1 (class keterangan status bayar ="Bayar") = jumlah total "bayar" pada kolom keterangan status bayar.

$$= \frac{5739}{6240} = 0,9197115384615385$$

- 2) P2 (class keterangan status bayar="Tidak Bayar") = jumlah total "tidak bayar" pada kolom keterangan status bayar.

$$= \frac{501}{6240} = 0,0802884615384615$$

$$= 0,080288461538462$$

2. Menghitung jumlah kasus yang sama pada setiap atribut dari kelas status bayar berdasarkan testing.

- 1) Menghitung probabilitas atribut jurusan dengan nilai studi agama-agama berdasarkan atribut status bayar.

- a. $P(\text{jalur}=\text{"studi agama-agama"} \mid \text{class status bayar}=\text{"bayar"})$

$$= 81/5739$$

$$= 0.014113957135389$$

- b. $P(\text{jalur}=\text{"studi agama-agama"} \mid \text{class status bayar}=\text{"tidak bayar"})$

$$= 12/501$$

$$= 0.023952095808383$$

- 2) Menghitung probabilitas atribut fakultas dengan nilai ushuluddin berdasarkan atribut status bayar.

- a. $P(\text{fakultas}=\text{"ushuluddin"} \mid \text{class status bayar}=\text{"bayar"})$

$$= 548/5739$$

$$= 0.095487018644363$$

- b. $P(\text{fakultas}=\text{"ushuluddin"} \mid \text{class status bayar}=\text{"tidak bayar"})$

$$= 65/501$$

$$= 0.12974051896208$$

- 3) Menghitung probabilitas berdasarkan atribut kategori ukt dengan nilai k2 berdasarkan atribut status bayar.

- a. $P(\text{kategori ukt}=\text{"k2"} \mid \text{class status bayar}=\text{"bayar"})$

$$= 697/5739$$

$$= 0.1214497299181$$

- b. $P(\text{kategori ukt}=\text{"k2"} \mid \text{class status bayar}=\text{"tidak bayar"})$

$$= 77/501$$

$$= 0.15369261477046$$

- 4) Menghitung probabilitas atribut jalur dengan nilai umptkin berdasarkan atribut status bayar.

- a. $P(\text{jalur}=\text{"umptkin"} \mid \text{class status bayar}=\text{"bayar"})$

$$= 878/5739$$

$$= 0.15298832549225$$

- b. $P(\text{jalur}=\text{"umptkin"} \mid \text{class status bayar}=\text{"tidak bayar"})$

$$= 126/501$$

$$= 0.25149700598802$$

- 5) Menghitung probabilitas berdasarkan atribut jenis kelamin dengan nilai berdasarkan atribut status bayar.

- a. $P(\text{jenis kelamin}=\text{"laki-laki"} \mid \text{class status bayar}=\text{"bayar"})$

$$= 2465/5739$$

- $$= 0.42951733751525$$
- $$\text{b. } P(\text{jenis kelamin} = \text{"laki-laki"} \mid \text{class status bayar} = \text{"tidak bayar"}) = 202/501$$
- $$= 0.40319361277445$$
- 6) Menghitung probabilitas berdasarkan atribut tempat tinggal dengan nilai berdasarkan atribut status bayar.
- $$\text{a. } P(\text{tempat tinggal} = \text{"bersama orang tua"} \mid \text{class status bayar} = \text{"bayar"})$$
- $$= 4841/5739$$
- $$= 0.84352674682$$
- $$\text{b. } P(\text{tempat tinggal} = \text{"bersama orang tua"} \mid \text{class status bayar} = \text{"tidak bayar"})$$
- $$= 422/501$$
- $$= 0.84231536926148$$
- 7) Menghitung probabilitas berdasarkan atribut kendaraan pribadi dengan nilai berdasarkan atribut status bayar.
- $$\text{a. } P(\text{kendaraan pribadi} = \text{"tidak memiliki"} \mid \text{class status bayar} = \text{"bayar"})$$
- $$= 1701/5739$$
- $$= 0.29639309984318$$
- $$\text{b. } P(\text{kendaraan pribadi} = \text{"tidak memiliki"} \mid \text{class status bayar} = \text{"tidak bayar"})$$
- $$= 144/501$$
- $$= 0.2874251497006$$
- 8) Menghitung probabilitas berdasarkan atribut asal sma dengan nilai berdasarkan atribut status bayar.
- $$\text{a. } P(\text{asal sma} = \text{"administrasi perkantoran"} \mid \text{class status bayar} = \text{"bayar"})$$
- $$= 28/5739$$
- $$= 0.0048788987628507$$
- $$\text{b. } P(\text{asal sma} = \text{"administrasi perkantoran"} \mid \text{class status bayar} = \text{"tidak bayar"})$$
- $$= 3/501$$
- $$= 0.0059880239520958$$
- 9) Menghitung probabilitas berdasarkan atribut pekerjaan ibu dengan nilai berdasarkan atribut status bayar.
- $$\text{a. } P(\text{pekerjaan ibu} = \text{"tidak bekerja"} \mid \text{class status bayar} = \text{"bayar"})$$
- $$= 3137/5739$$
- $$= 0.54661090782366$$
- $$\text{b. } P(\text{pekerjaan ibu} = \text{"tidak bekerja"} \mid \text{class status bayar} = \text{"tidak bayar"})$$
- $$= 257/501$$
- $$= 0.51297405189621$$
- 10) Menghitung probabilitas berdasarkan atribut gaji ibu dengan nilai berdasarkan atribut status bayar.
- $$\text{a. } P(\text{gaji ibu} = \text{"0 S.D 400.000"} \mid \text{class status bayar} = \text{"bayar"})$$
- $$= 3851/5739$$
- $$= 0.67102282627635$$
- $$\text{b. } P(\text{gaji ibu} = \text{"0 S.D 400.000"} \mid \text{class status bayar} = \text{"tidak bayar"})$$
- $$= 354/501$$
- $$= 0.70658682634731$$
- 11) Menghitung probabilitas berdasarkan atribut pekerjaan ayah dengan nilai berdasarkan atribut status bayar.
- $$\text{a. } P(\text{pekerjaan ayah} = \text{"Buruh (Tani/Pabrik/Bangunan)"} \mid \text{class status bayar} = \text{"bayar"})$$
- $$= 595/5739$$
- $$= 0.10367659871058$$
- $$\text{b. } P(\text{pekerjaan ayah} = \text{"Buruh (Tani/Pabrik/Bangunan)"} \mid \text{class status bayar} = \text{"tidak bayar"})$$
- $$= 56/501$$

$$= 0.11177644710579$$

12) Menghitung probabilitas berdasarkan atribut gaji ayah dengan nilai berdasarkan atribut status bayar.

$$a. P(\text{gaji ayah} = "750.001 \text{ S.D } 1.250.000" \mid \text{class status bayar} = \text{"bayar"})$$

$$= 645/5739$$

$$= 0.11238891792995$$

$$b. P(\text{gaji ayah} = "750.001 \text{ S.D } 1.250.000" \mid \text{class status bayar} = \text{"tidak bayar"})$$

$$= 47/501$$

$$= 0.093812375249501$$

3. Kalikan semua hasil variabel bayar atau tidak bayar.

1) Semua atribut class status bayar = "bayar"

$$P(X \mid \text{class status bayar} = \text{"bayar"})$$

$$= 0.014113957135389 * 0.095487018644363 * 0.1214497299181 * 0.42951733751525 * \\ 0.15298832549225 * 0.84352674682 * 0.29639309984318 * 0.0048788987628507 * \\ 0.54661090782366 * 0.67102282627635 * 0.10367659871058 * 0.11238891792995 \\ = 5.60709E-11$$

2) Semua atribut class status bayar = "tidak bayar"

$$P(X \mid \text{class status bayar} = \text{"tidak bayar"})$$

$$= 0.023952095808383 * 0.12974051896208 * 0.15369261477046 * 0.40319361277445 * \\ 0.25149700598802 * 0.84231536926148 * 0.2874251497006 * 0.0059880239520958 * \\ 0.51297405189621 * 0.70658682634731 * 0.11177644710579 * 0.093812375249501 \\ = 0.0000000002669$$

4. Kalikan prior probability dengan semua atribut class status = "bayar" atau "tidak bayar".

$$1) P(C_i) \mid \text{class status bayar} = \text{"bayar"} \times P(X \mid \text{class status bayar} = \text{"bayar"})$$

$$= 5.60709E-11 * 0.919711538461538$$

$$= 5.1569020007112E-11$$

$$2) P(C_i) \mid \text{class status bayar} = \text{"tidak bayar"} \times P(X \mid \text{class status bayar} = \text{"tidak bayar"})$$

$$= 0.0000000002669 * 0.0802884615384615$$

$$= 2.1425261746368E-11$$

5. Bandingkan hasil class Bayar dan Tidak Bayar karena hasil $P(C_i) \mid \text{Bayar}$ lebih besar dari $P(C_i) \mid \text{Tidak Bayar}$ maka keputusan adalah "Bayar".

5. Penutup

5.1 Kesimpulan

Kesimpulan yang didapat dari hasil analisis prediksi calon mahasiswa baru yang akan lanjut melakukan pembayaran atau tidak dengan menggunakan algoritma *naïve bayes* adalah sebagai berikut:

1. Algoritma Naïve Bayes yang diterapkan pada aplikasi ini memberikan hasil yang cukup akurat dalam menentukan prediksi calon mahasiswa baru yang akan lanjut melakukan pembayaran atau tidak.
2. Algoritma Naïve Bayes ini memberikan tingkat akurasi yang sangat besar yaitu 91,811% dengan data testing yang digunakan sebanyak 6240 data.

5.2 Saran

Agar kerja dari aplikasi prediksi ini berjalan lebih optimal, penulis memberikan beberapa saran, yaitu :

1. Perlu adanya penambahan kriteria atau atribut untuk memperbesar dari prediksi calon mahasiswa baru yang akan lanjut melakukan pembayaran atau tidak.
2. Perlu adanya *maintance* sistem secara berkala untuk menghindari ada bug pada aplikasi prediksi ini.
3. Perlu adanya pengujian data yang belum terklasifikasi sama sekali

Referensi

- [1] Fadhilah Rama, "Lima Jalur Pendaftaran, UIN Bandung Akan Terima 5700 Mahasiswa Baru," *SUAKAONLINE*, 2018. [Online]. Available: <http://suakaonline.com/12827/2018/04/26/lima-jalur-pendaftaran-uin-bandung-akan-terima-5700-mahasiswa-baru/>. [Accessed: 30-Mar-2019].
- [2] M. H. Rifqo, A. Wijaya, and J. Pseudocode, "Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit," vol. IV, no. September, pp. 120–128, 2017.
- [3] D. Wahyudi, A. H. Mirza, M. Kom, P. H. Merriayu, and M. Kom, "IMPLEMENTASI DATA MINING DENGAN NAIVE BAYES CLASSIFIER UNTUK Mendukung Strategi PROMOSI (Studi kasus Universitas Bina Darma Palembang)," pp. 1–10.
- [4] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec J.*, vol. 2, no. 3, pp. 207–217, 2015.
- [5] M. Efendi, "IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES DALAM MENENTUKAN PENGUNDURAN DIRI CALON MAHASISWA PADA UNIVERSITAS DIAN NUSWANTORO SEMARANG Muhammad Efendi," pp. 1–5, 2014.
- [6] P. M. Swamidass, "Prototype Model," pp. 595–595, 2000.
- [7] Nurdin and D. Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Pada Supermarket Sejahtera Lhokseumawe," vol. 6, no. 1, pp. 134–155, 2015.
- [8] D. M. Bates and D. G. Watts, "Data Mining," *Nonlinear Regres. Anal. Its Appl.*, vol. 32, no. 2, p. 90, 1988.
- [9] Adrian Satria Putra, "KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN NAIVE BAYES CLASSIFICATION (STUDI KASUS POSYANDU NGUDI LUHUR)," 2018.
- [10] Aditya, F. Marisa, and D. Purnomo, "Penerapan Algoritma Apriori Terhadap Data Penjualan di Toko Gudang BM," *Jointecs*, vol. 1, no. 1, pp. 1–5, 2016.
- [11] D. K. Ilmiah and P. T. Informatika, "IMPLEMENTASI METODE NAIVE BAYES CLASSIFICATION DALAM KLASIFIKASI KELAYAKAN CALON PENDONOR DARAH (STUDI KASUS PMI KAB . DEMAK)," pp. 1–5, 2016.
- [12] I. A. Dahlia, "PERBANDINGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK PREDIKSI PERCERAIAN (Studi Kasus : Pengadilan Agama Cimahi)," Universitas Islam Negeri Sunan Gunung Djati Bandung, 2017.
- [13] H. Wasiati and D. Wijayanti, "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes," *J. Ilm. NERO*, vol. 1, no. 4, pp. 1–7, 2014.
- [14] T. Rosandy, "Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree Untuk Menganalisa Kelancaran Pembiayaan," vol. 02, no. 01, pp. 52–62, 2016.
- [15] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier," *Semin. Nas. Apl. Teknol. Inf.*, vol. 20, no. ISSN: 1907-5022, pp. 5–10, 2014.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG