

IMPLEMENTASI TEORI NAÏVE BAYES DALAM KLASIFIKASI CALON MAHASISWA BARU STMIK KHARISMA MAKASSAR

Irayori Loelianto¹, Moh. Sofyan S. Thayf², Husni Angriani³

^{1,2,3}STMIK KHARISMA Makassar

Jl. Baji Ateka No.20, Baji Mappakasunggu, Kec. Mamajang, Kota Makassar, Sulawesi Selatan 90121, Indonesia

e-mail: irayoriloelianto_17@kharisma.ac.id¹, sofyan.thayf@kharisma.ac.id², husniangriani@kharisma.ac.id³

Received : August, 2020

Accepted : September, 2020

Published : October, 2020

Abstract

STMIK KHARISMA Makassar has graduated thousands of alumni since it was founded. However, the number of students registering is uncertain every year, although from 2016 to 2019 there has been an increase in the number of registrations. The problem is the percentage of the number of prospective students registering has actually decreased significantly. The purpose of this research is to implement the Naive Bayes theory in classification of STMIK KHARISMA Makassar prospective students. This research basically uses the Naive Bayes theory as a classifier, and is made using the Python programming language. At the classifier design stage, there were a total of 499 data collected from 2016 to 2019. The data was divided by a ratio of 80:20 for training data and test data. The result from the research indicate the level of accuracy of the classifier reaches 73%.

Keywords: Prospective Students, Naive Bayes, Python

Abstrak

STMIK KHARISMA Makassar telah meluluskan ribuan alumni sejak didirikan. Namun, jumlah calon mahasiswa baru yang mendaftar tidak menentu setiap tahunnya, meskipun sejak tahun 2016 hingga tahun 2019 jumlah pendaftar mengalami peningkatan. Tetapi persentase dari calon mahasiswa yang registrasi justru mengalami penurunan yang signifikan. Tujuan dari penelitian ini adalah mengimplementasikan teori Naive Bayes dalam pengklasifikasian calon mahasiswa baru STMIK KHARISMA Makassar. Pada penelitian ini menggunakan metode Naive Bayes untuk membuat classifier dan dibuat menggunakan bahasa pemrograman Python. Pada tahap perancangan classifier, 499 data dikumpulkan yaitu data dari angkatan 2016 sampai 2019. Data tersebut dibagi dengan rasio 80:20 untuk data latih dan data uji. Hasil pengujian yang dilakukan menunjukkan tingkat akurasi classifier mencapai 73%.

Kata Kunci: Calon Mahasiswa Baru, Naive Bayes, Python

1. PENDAHULUAN

STMIK KHARISMA Makassar adalah salah satu perguruan tinggi swasta, telah berdiri selama 20 tahun sejak tahun 2000. STMIK KHARISMA Makassar bergerak di bidang informatika dan komputer. Dalam proses penerimaan mahasiswa baru, STMIK KHARISMA Makassar

telah menerapkan sistem pendaftaran secara *online*. Sejak berdirinya STMIK KHARISMA Makassar telah menghasilkan ribuan alumni. Namun, jumlah mahasiswa yang mendaftar di STMIK KHARISMA Makassar semakin tidak menentu dari tahun ke tahun. Bahkan tidak jarang terdapat calon mahasiswa yang hanya

melakukan pendaftaran tetapi tidak melakukan registrasi lanjut.

Tabel 1 menunjukkan data calon mahasiswa baru tahun 2016 sampai 2019. Dari tahun 2016 sampai 2019 jumlah calon mahasiswa pendaftar STMIK KHARISMA Makassar mengalami peningkatan. Akan tetapi persentase dari calon mahasiswa yang melakukan registrasi justru menurun secara signifikan dari tahun ke tahun. Dengan begini, bisa dikatakan jumlah pendaftar dan jumlah registrasi semakin tidak seimbang dari tahun ke tahun.

Dengan menganalisis dan mengetahui karakteristik dari calon mahasiswa baru, maka dapat membantu *marketing* dan komunikasi STMIK KHARISMA Makassar dalam menentukan arah strategi promosi ke depannya, sehingga diharapkan dapat meningkatkan jumlah calon mahasiswa registrasi setiap tahunnya. Untuk itu, STMIK KHARISMA Makassar membutuhkan pengetahuan baru tentang calon mahasiswa baru. Teknik penambangan data atau *data mining* bisa menjadi salah satu alternatif untuk mengatasi masalah ini, yaitu dengan mengklasifikasikan calon mahasiswa baru.

Menurut Larose, *data mining* sendiri adalah suatu proses dalam menggali pola, kecenderungan, ataupun memahami hubungan dari berbagai variabel melalui serangkaian pemeriksaan terhadap sekumpulan data yang dimiliki dengan menggunakan teknik statistik ataupun matematika [1]. Terdapat beberapa metode yang sering dibahas dalam literatur *data mining*, diantaranya adalah *classification* (pengklasifikasian), *clustering* (pengelompokan), *association* (hubungan), *estimation* (estimasi), dan *prediction* (prediksi).

Beberapa penelitian terkait pemanfaatan teknik *data mining* telah dilakukan. Penelitian [2] menggunakan metode *Clustering* untuk meningkatkan potensi penerimaan mahasiswa baru. Ada juga penelitian [3] menerapkan metode *Naive Bayes Classifier* untuk perankingan kegiatan Fakultas. Penelitian [4] juga menggunakan metode *Naive Bayes Classifier* dalam mendeteksi minat beli pelanggan terhadap produk paket internet. Penelitian yang dilakukan oleh Handayani dan Pribadi mengimplementasikan algoritma *Naive Bayes* dalam pengklasifikasian teks otomatis [5]. Penelitian lain [6] menggunakan teknik *FP-Growth* untuk mengidentifikasi relasi (asosiasi) dari masing-masing atribut atau variabel.

Pada penelitian ini dilakukan pengklasifikasian calon mahasiswa baru di STMIK KHARISMA Makassar dengan menggunakan salah satu algoritma pengklasifikasian. Ada beberapa algoritma pengklasifikasian dalam teknik *data mining* yang populer, diantaranya yaitu *Naive Bayes*, *Decision Tree*, *K-Nearest Neighbour*, dan *Neural Network*. Teknik klasifikasi yang digunakan yaitu *Naive Bayes*.

Algoritma *Naive Bayes* dapat bekerja dengan baik terlepas dari jumlah set data yang banyak ataupun sedikit. Han dan Kamber menjelaskan bahwa dalam penerapannya terhadap *database* yang besar pun, *Teorema Bayes* tetap mampu bekerja maksimal dan mempunyai keakurasian yang tinggi [7]. Selain itu, Xhemali, Hinde, dan Stone dalam penelitiannya juga berhasil membuktikan bahwa apabila algoritma *Naive Bayes* dibandingkan algoritma *classifier* lainnya, algoritma klasifikasi *Naive Bayes* masih bekerja lebih baik [8]. Klasifikasi yang akan dilakukan adalah klasifikasi calon mahasiswa baru yang membagi data menjadi dua kelas/label, yaitu Potensial dan Tidak Potensial.

Tabel 1: Data Calon Mahasiswa Baru Tahun 2016 Sampai 2019
[Sumber: Hasil Wawancara dengan *Marketing* dan Komunikasi STMIK KHARISMA Makassar]

| Tahun PMB | Pendaftar | Registrasi | Tidak Registrasi | Persentase Registrasi | Penurunan Persentase Registrasi |
|-------------|-----------|------------|------------------|-----------------------|---------------------------------|
| 2016 - 2017 | 70 orang | 48 orang | 22 orang | 68,57 % | |
| 2017 - 2018 | 108 orang | 64 orang | 44 orang | 59,25 % | 9,32 % |
| 2018 - 2019 | 145 orang | 69 orang | 76 orang | 47,58 % | 11,67 % |
| 2019 - 2020 | 186 orang | 72 orang | 114 orang | 38,70 % | 8,88 % |

2. METODE PENELITIAN

Tahapan penelitian yang dilakukan meliputi tahapan *data mining* dalam menerapkan metode klasifikasi *Naive Bayes*. Istilah "*data mining*" ini digunakan cukup luas di industri Teknologi Informasi (TI). Ini sering diterapkan pada berbagai kegiatan pemrosesan data skala besar seperti pengumpulan, ekstraksi, pergudangan, dan analisis data.

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [9]. *Naive Bayes* menggunakan model fitur independen dengan asumsi independensi yang kuat (naif) [10]. Persamaan dari *Teorema Bayes* adalah:

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

dimana X menunjukkan *class* yang belum diketahui, H yaitu hipotesis data, $P(H|X)$ yaitu probabilitas hipotesis H berdasarkan X, $P(H)$ yaitu probabilitas hipotesis H, $P(X|H)$ yaitu probabilitas X berdasarkan hipotesis H, dan $P(X)$ yaitu probabilitas X [11].

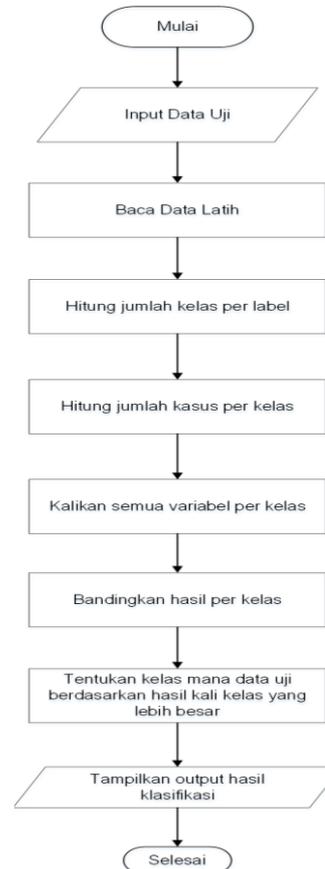
Persamaan di atas adalah persamaan *Teorema Bayes* yang digunakan untuk melakukan klasifikasi. Proses klasifikasi algoritma *Naive Bayes* digambarkan dalam bentuk *flowchart* diagram. *Flowchart* merupakan cara untuk menggambarkan proses kerja algoritma atau langkah-langkah pemecahan suatu masalah yang direpresentasikan menggunakan simbol-simbol tertentu.

Flowchart dari algoritma *Naive Bayes* ditunjukkan pada Gambar 1. Dengan demikian, dapat disimpulkan bahwa langkah kerja dalam proses klasifikasi *Naive Bayes*, yaitu:

- 1) Hitung jumlah kelas per label
- 2) Hitung jumlah kasus per kelas
- 3) Kalikan semua variabel per kelas
- 4) Bandingkan hasil per kelas

Keuntungan dari penggunaan *Naive Bayes* menurut Pattekari dan Parveen di antaranya adalah dapat mengestimasi parameter yang dibutuhkan dalam melakukan klasifikasi walaupun jumlah data latihnya relatif kecil. Dalam banyak kasus kompleks, *Naive Bayes* cenderung bekerja jauh lebih baik [12].

Penelitian ini menggunakan bahasa pemrograman Python dengan memanfaatkan *library-library* yang telah disediakan. Bahasa pemrograman Python merupakan bahasa pemrograman yang memiliki banyak fungsi interaktif berorientasi objek dan merupakan bahasa pemrograman tingkat tinggi [13].



Gambar 1. *Flowchart* Algoritma *Naive Bayes*

Tahapan penelitian ditunjukkan pada Gambar 2. Penelitian diawali dengan seleksi data, dilanjutkan dengan proses *preprocessing* (*data cleaning*), *transformation*, merancang dan melatih *classifier*, serta evaluasi *classifier*.

Berikut penjabaran dari setiap tahapan dalam penelitian ini.

1. Seleksi Data

Data yang ada tidak semuanya dipakai dalam penelitian ini. Untuk itu, perlu dilakukan seleksi data. Seleksi data yaitu mengambil data-data yang diperlukan saja dalam penelitian. Dalam hal ini berarti data dari atribut/variabel yang digunakan dalam penelitian, diantaranya adalah asal sekolah, jalur masuk, program studi

pilihan, jenis kelamin, penghasilan orang tua, serta label.

Untuk atribut Penghasilan Orang Tua, penulis terlebih dahulu mengkategorikannya sebagai berikut:

- a) <Rp 2.500.000 dikategorikan sebagai *Low*
- b) Rp 2.500.000-Rp 5000.000 dikategorikan sebagai *Medium*, dan
- c) >Rp 5.000.000 dikategorikan sebagai *High*

2. *Preprocessing (Data Cleaning)*

Preprocessing yaitu proses pembersihan data yang akan dianalisis. Data yang diperoleh masih mengandung data yang bernilai *null* dan data duplikasi. Oleh karena itu, sebelum menganalisis data, harus dipastikan data yang dianalisis adalah data bersih, lengkap (tidak mengandung nilai *null*), dan tidak memiliki data duplikasi karena dapat memberi hasil yang tidak akurat. Proses ini meliputi penghapusan dan pemodifikasian data yang tidak benar, tidak relevan, maupun yang duplikasi.

Pembersihan data dilakukan secara manual dengan menghapus data yang nilainya sama (data duplikasi). Hasil pembersihan data duplikasi ada sebanyak 10 data, sehingga data yang awalnya sebanyak 509 data menjadi 499 data. Dengan demikian, data yang diolah hanya sebanyak 499 data. Setelah bersih dari data duplikasi, selanjutnya adalah melakukan modifikasi untuk data yang kosong. Pemodifikasian dilakukan dengan memodifikasi/mengisi data yang kosong dengan data dominan atau data terbanyak.

3. *Data Transformation*

Metode-metode dalam *data mining* sering kali membutuhkan format atau struktur data khusus sebelum bisa diimplementasikan. Proses transformasi yaitu mengubah data yang ada dari satu format atau struktur ke format atau struktur lain yang siap diolah. Melalui proses transformasi memungkinkan *data mining* yang diperoleh dapat lebih efisien. Tidak hanya itu, pola yang ditemukan juga lebih mudah untuk kita pahami. Pada kasus ini, transformasi data dilakukan dari format data

string ke format data numerik yang siap diolah oleh algoritma pengklasifikasian *Naïve Bayes*.

4. Merancang dan Melatih *Classifier*

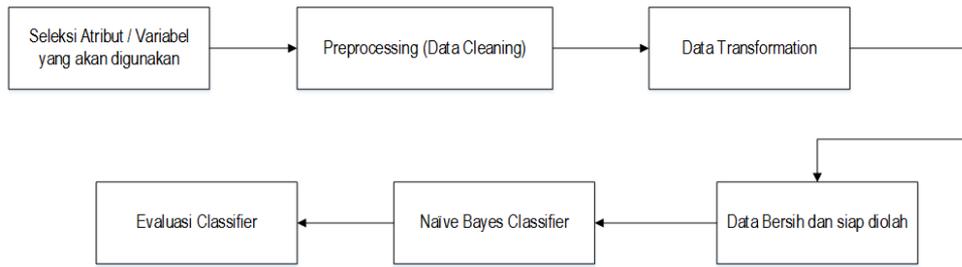
Pada tahap ini classifier mulai dirancang. *Naïve Bayes* melakukan klasifikasi dengan membagi data menjadi dua yaitu *data training* dan *data testing*. Untuk itu, data bersih yang siap diolah dibagi terlebih dahulu menjadi dua set data. Satu adalah data latih (*data training*) yang digunakan untuk membangun atau melatih model/*classifier*. Sedangkan yang lainnya adalah data uji (*data testing*), yang digunakan untuk menguji *classifier* yang telah dibangun untuk melihat seberapa akurat hasil klasifikasinya.

Pembagian data dilakukan dengan menggunakan prinsip *Pareto* yaitu prinsip 80:20. Sebanyak 80% dari data adalah data latih dan 20% sisanya adalah data uji. Pada kasus ini, jumlah data yang akan diolah ada sebanyak 499 data, sehingga 80% diantaranya yaitu sebanyak 399 data sebagai data latih, dan 20% sisanya yaitu sebanyak 100 data sebagai data uji.

5. Evaluasi *Classifier*

Evaluasi yaitu proses menyajikan pola-pola yang diperoleh dari hasil *data mining*. Selain itu, pada tahap ini juga akan dilakukan pemeriksaan terhadap pola atau informasi yang ditemukan dari hasil data mining apakah memang sesuai dengan fakta atau justru berlawanan dengan fakta atau hipotesis sebelumnya.

Pada tahap ini, *classifier* yang telah berhasil dirancang akan diukur tingkat ketepatannya. Pengujian sistem dilakukan dengan menggunakan metode *confusion matrix*. *Confusion matrix* dikenal sebagai metode yang paling umum digunakan dalam mengukur akurasi dari model *data mining* yang dibangun. *Confusion matrix* melakukan perhitungan yang dapat menghasilkan tiga nilai keluaran, yaitu: *recall*, *precision*, dan *accuracy*. Ini diperoleh dengan membandingkan hasil klasifikasi dari *classifier* dengan label data yang sesungguhnya.



Gambar 2. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Mengklasifikasikan Data Baru

Sistem dapat mengklasifikasikan data baru. Untuk memudahkan pengguna dalam mengklasifikasikan data baru, penulis menyediakan file template Excel bernama new dataset yang terdiri atas kolom nama mahasiswa, asal sekolah, jalur, program studi, jenis kelamin, penghasilan orang tua, dan label. *User* cukup menginput data kelima atribut dari calon mahasiswa baru yang ingin diklasifikasikan pada *file* Excel yang telah disediakan seperti Gambar 3. Untuk kolom label dibiarkan saja kosong. Kemudian menjalankan aplikasi cukup dengan *double click* pada aplikasi. Sistem akan mengklasifikasikan data tersebut secara otomatis. Setelah proses selesai, akan tampil kotak dialog pesan yang berisi "Proses Berhasil!". Selanjutnya, buka file Excel yang telah di-*input* tadi. Data pada kolom label telah terisi dengan data hasil klasifikasi *classifier* seperti Gambar 4.

3.2 Features Importance

Dalam banyak kasus atau bisnis, sama pentingnya untuk tidak hanya memiliki model yang akurat, tetapi juga dapat ditafsirkan. Seringkali, selain ingin mengetahui apa hasil klasifikasi dari sistem, kita juga bertanya-tanya mengapa hasilnya seperti itu, fitur atau atribut mana yang paling penting dan besar pengaruhnya dalam menentukan hasil klasifikasi. Secara umum, kepentingan fitur didefinisikan sebagai kontribusi relatifnya terhadap pengambilan keputusan algoritma.

Dengan mengidentifikasi besarnya pengaruh yang diberikan oleh masing-masing atribut, dapat membantu organisasi atau badan usaha dalam deteksi dini dan bahkan mungkin dapat meningkatkan strategi, produk, atau layanannya dengan lebih berfokus pada atribut penting tersebut.

Semakin akurat suatu model yang dibangun, semakin kita dapat mempercayai fitur-fitur penting dan interpretasi lainnya. Gambar 5 menunjukkan *features importance* dari setiap atribut yang digunakan dalam penelitian ini, dengan nilai *features importance* sebagai berikut:

- Asal Sekolah = 0.03398
- Jalur = 0.00645
- Program Studi = 0.00223
- Jenis Kelamin = 0.00301
- Penghasilan Orang Tua = 0.95430

Dengan demikian, fitur/atribut yang paling mempengaruhi hasil pengambilan keputusan algoritma (hasil klasifikasi) adalah atribut penghasilan orang tua.

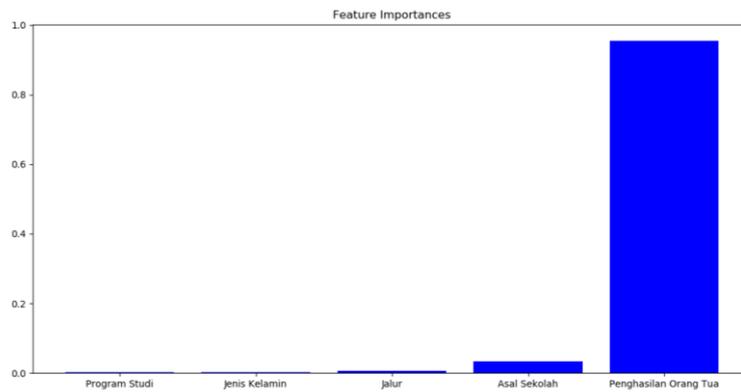
Berdasarkan Gambar 6, dapat dilihat bahwa calon Maba yang penghasilan orang tuanya tergolong *high* kemungkinan besar akan melakukan registrasi. Calon maba yang penghasilan orang tuanya tergolong *low* cenderung melakukan registrasi. Sementara itu, calon maba yang penghasilan orang tuanya tergolong *medium* justru cenderung tidak melakukan registrasi.

| | A | B | C | D | E | F | G |
|----|-----------------------------------|----------------------|----------|---------------|---------------|-----------------------|-------|
| 1 | Nama Mahasiswa | Asal Sekolah | Jalur | Program Studi | Jenis Kelamin | Penghasilan Orang Tua | Label |
| 2 | Nur Adrey Natasyia Admunady | SMA Negeri 7 | PMDK | SI | P | Low | |
| 3 | Edy Kurniawan Hardi | SMK Negeri 1 | PMDK | TI | L | Medium | |
| 4 | Raffi Gozallanto | SMA YHS | PMDK | TI | L | High | |
| 5 | Ika Rustin Aprilliani | SMK Tri Tunggal "45" | PMDK | TI | P | Medium | |
| 6 | Michael Owen Ouwyong | SMA Katolik Rajawali | PMDK | TI | L | High | |
| 7 | Imelda The | SMA Negeri 4 Bau-Bau | PMDK | SI | P | Medium | |
| 8 | Naldi | SMA Katolik Rajawali | PMDK | TI | L | Medium | |
| 9 | Musafir | SMA Negeri 1 Bajeng | PMDK | TI | L | Medium | |
| 10 | Muhammad Zaqiyamani Eka Indrajaya | SMA Perguruan Islam | PMDK | SI | L | Low | |
| 11 | Muh. Yusuf Anwar | SMK Negeri 2 | KONVERSI | TI | L | Low | |

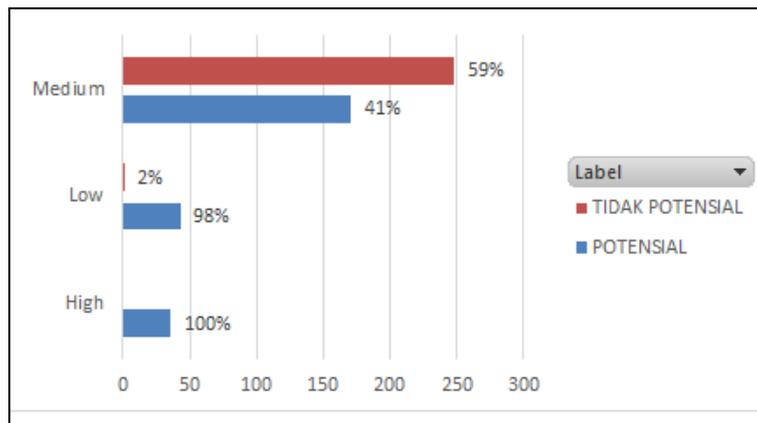
Gambar 3. Template File New Dataset

| 1 | | NamaMahasiswa | AsalSekolah | Jalur | ProgramStudi | JenisKelamin | PenghasilanOrangTua | Label |
|----|---|-----------------------------------|----------------------|----------|--------------|--------------|---------------------|-----------------|
| 2 | 0 | Nur Adrey Natasyia Admunady | SMA Negeri 7 | PMDK | SI | P | Low | POTENSIAL |
| 3 | 1 | Edy Kurniawan Hardi | SMK Negeri 1 | PMDK | TI | L | Medium | TIDAK POTENSIAL |
| 4 | 2 | Rafli Gozalianto | SMA YHS | PMDK | TI | L | High | POTENSIAL |
| 5 | 3 | Ika Rustin Aprilliani | SMK Tri Tunggal "45" | PMDK | TI | P | Medium | TIDAK POTENSIAL |
| 6 | 4 | Michael Owen Ouwyong | SMA Katolik Rajawali | PMDK | TI | L | High | POTENSIAL |
| 7 | 5 | Imelda The | SMA Negeri 4 Bau-Bau | PMDK | SI | P | Medium | TIDAK POTENSIAL |
| 8 | 6 | Naldi | SMA Katolik Rajawali | PMDK | TI | L | Medium | TIDAK POTENSIAL |
| 9 | 7 | Musafir | SMA Negeri 1 Bajeng | PMDK | TI | L | Medium | TIDAK POTENSIAL |
| 10 | 8 | Muhammad Zaqiyamani Eka Indrajaya | SMA Perguruan Islam | PMDK | SI | L | Low | POTENSIAL |
| 11 | 9 | Muh. Yusuf Anwar | SMK Negeri 2 | KONVERSI | TI | L | Low | POTENSIAL |

Gambar 4. Hasil Klasifikasi Classifier



Gambar 5. Grafik Features Importance



Gambar 6. Grafik Atribut Penghasilan Orang Tua

3.3 Pengujian

Setelah semua proses *data mining* telah dilakukan, proses terakhir adalah melakukan pengujian terhadap sistem yang telah dibuat. Dalam bidang *data mining*, pengujian merupakan proses menilai seberapa baik kinerja model pertambangan terhadap data nyata. Pengujian dilakukan dengan mengklasifikasikan label dalam data uji untuk mengukur tingkat ketepatan dari sistem yang dibangun. Berdasarkan hasil klasifikasi sistem, diketahui nilai *True Positive*, *True Negative*, *False Positive*, dan *False Negative* sebagai berikut:

| n = 100 | Predicted 0 | Predicted 1 |
|----------|-------------|-------------|
| | Actual 0 | 23 |
| Actual 1 | 1 | 50 |

Dengan demikian, diperoleh nilai akurasi, *sensitivity/recall*, dan *precision* sistem adalah:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = \frac{23 + 50}{100} = 0.73 = 73\%$$

$$Sensitivity = \frac{TP}{TP + FN} = \frac{23}{23 + 26} = 0.4693 = 46,93\%$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{23}{23 + 1} = 0.9583 = 95,83\%$$

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan yaitu mengimplementasikan teori *Naïve Bayes* dalam pengklasifikasian calon mahasiswa baru STMIK KHARISMA Makassar, dapat disimpulkan bahwa:

1. Penulis telah berhasil mengimplementasikan teori *Naïve Bayes* dalam pengklasifikasian calon mahasiswa baru STMIK KHARISMA Makassar.
2. Sistem dapat melakukan klasifikasi calon mahasiswa baru dengan terlebih dahulu melalui proses *preprocessing*.
3. Nilai *features importance* tertinggi adalah penghasilan orang tua, yang berarti atribut ini yang paling mempengaruhi hasil klasifikasi dari classifier yang dibangun.
4. Dari hasil pengujian sistem, diperoleh tingkat akurasi *Classifier* yang dibuat mencapai 73%, dengan nilai *sensitivity* dan *precision* sebesar 46,93% dan 95,83%. Hal ini menunjukkan kinerja *Classifier* yang cukup baik dalam mengklasifikasikan calon mahasiswa baru STMIK KHARISMA Makassar. Nilai *precision* yang tinggi menunjukkan tingkat ketepatan *Classifier* yang tinggi dalam mengklasifikasikan kejadian positif. Akan tetapi, nilai *sensitivity* dari *Classifier* yang dibuat masih agak kurang, yang berarti sistem masih kurang sensitif dalam mendeteksi kejadian positif.

Adapun saran untuk pengembangan penelitian selanjutnya adalah:

1. STMIK KHARISMA Makassar membutuhkan lebih banyak lagi data terkait calon mahasiswa baru, sehingga memungkinkan untuk menggunakan atribut atau variabel yang lebih bervariasi. Misalnya dapat menambahkan jarak rumah, hobi, ataupun sumber informasi terkait STMIK KHARISMA Makassar sebagai atribut.
2. Mencoba mencari atribut-atribut penting (*features importance*) lainnya.
3. Sebaiknya menggunakan data dalam jumlah yang lebih besar, sehingga pelatihan terhadap sekumpulan data dapat lebih maksimal. Dengan demikian, juga dapat meningkatkan nilai akurasi, *sensitivity*, dan *precision*.
4. Dapat mencoba mengimplementasikan metode atau algoritma pengklasifikasian

lainnya untuk dibandingkan dengan algoritma pengklasifikasian *Naïve Bayes*.

DAFTAR PUSTAKA

- [1] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. Canada: John Wiley & Sons, Inc, 2005.
- [2] Suharyadi and E. Sedyono, "Metode Microsoft Clustering untuk Meningkatkan Potensi Penerimaan Mahasiswa Baru," *J. Sist. Komput.*, vol. 7, no. 1, pp. 37–48, 2017.
- [3] A. N. Putri, "Penerapan Naive Bayesian untuk Perankingan Kegiatan di Fakultas TIK Universitas Semarang," *J. SIMETRIS*, vol. 8, no. 2, pp. 603–610, 2017.
- [4] A. S. Mubarak, D. R. Prehanto, and M. Ali, "Deteksi Minat Beli Pelanggan Terhadap Produk Paket Internet Menggunakan Algoritma Naïve Bayes," *INOVATE*, vol. 03, no. 02, pp. 58–63, 2019.
- [5] F. Handayani and F. S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *J. Tek. Elektro*, vol. 7, no. 1, pp. 19–24, 2015.
- [6] A. Ikhwan *et al.*, "A Novelty of Data Mining for Promoting Education Based on FP-Growth Algorithm," *Int. J. Civ. Eng. Technol.*, vol. 9, no. 7, pp. 1660–1669, 2018.
- [7] J. Han and M. Kamber, *Data Mining : Concepts and Techniques*, Second. San Fransisco: Elsevier, 2006.
- [8] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2009.
- [9] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec J.*, vol. 2, no. 3, pp. 207–217, 2015.
- [10] B. Harijanto, Y. Ariyanto, and L. Miftahurroifa, "Penerapan Algoritma Naive Bayes untuk Klasifikasi Retensi Arsip," *J. Inform. Polinema*, vol. 4, no. 2, pp. 155–160, 2018.
- [11] R. A. Anggraini, G. Widagdo, A. S. Budi, and M. Qomaruddin, "Penerapan Data

- Mining Classification untuk Data Blogger Menggunakan Metode Naïve Bayes," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 1, pp. 47–51, 2019.
- [12] S. A. Pattekari and A. Parveen, "Prediction System for Heart Disease using Naive Bayes," *Int. J. Adv. Comput. Math. Sci.*, vol. 3, no. 3, pp. 290–294, 2012.
- [13] A. N. Harrington, *Hands-On Python : A Tutorial Introduction for Beginners Python 3.1 Version*. Chicago: Computer Science Department, 2009.