

Penerapan Metode *Naïve Bayes* Dalam Prediksi Penentuan Jurusan Mahasiswa TI

Implementation of the Naïve Bayes Method in Predicting the Determination of IT Student Majors

Kharis Luqman Firdaus^{1,*}

¹*Fakultas Teknologi Informasi, Universitas Islam Balitar
Jl. Majapahit No.2- 4, Sananwetan, Kec. Sananwetan, Kota Blitar, Jawa Timur 66137 Indonesia*

^{1,*}Penulis Korespondensi: kharisluqman5@gmail.com

Received on 03-08-2020, accepted on 31-01-2021, published on 31-01-2021

Abstract

Penjurusan mahasiswa adalah bagian dari suatu fakultas atau sekolah tinggi untuk mengelola suatu bidang studi sebagai pemberi arah tujuan mahasiswa. Hal ini yang mempengaruhi penentuan jurusan seorang mahasiswa. Saat ini penjurusan mahasiswa TI di UNISBA Blitar yang dilakukan oleh prodi TI masih dilakukan secara manual. Penelitian ini bertujuan membantu penjurusan mahasiswa TI dengan menggunakan aplikasi sehingga dapat mengoptimalkan waktu saat penjurusan dan dapat sebagai bahan pertimbangan mahasiswa dengan prediksi yang telah dilakukan di aplikasi tersebut. Metode *Naïve Bayes* adalah teknik prediksi berbasis probabilistik sederhana yang berdasarkan pada penerapan teorema *Bayes* (atau aturan *Bayes*) dengan asumsi independensi yang kuat. Hasil yang didapat menggunakan algoritma *naïve bayes* dalam proses prediksi menggunakan pengujian *confusion matrix* didapat nilai akurasi sebesar 57% sedangkan nilai presisi mendapatkan hasil 71% dan untuk hasil recall didapat nilai sebesar 46% dari 44 data set yang telah diuji, terdapat hasil 25 prediksi benar dan 19 prediksi salah.

Kata Kunci: *Confusion Matrix, Naïve Bayes, Penjurusan mahasiswa, Prediksi*

Abstract

Student majors are part of a faculty or high school to manage a field of study as a direction for student goals. This affects the determination of a student's major. Currently, the IT student majors at UNISBA Blitar conducted by the IT study program are still done manually. This study aims to help IT student majors by using the application so that it can streamline the time during majors and can be taken into consideration by students with predictions that have been made in the application. The Naïve Bayes method is a simple probabilistic-based prediction technique based on the application of Bayes' theorem (or Bayes' rule) with the assumption of strong independence. The results obtained using the naive Bayes algorithm in the prediction process using the confusion matrix test obtained an accuracy value of 57% while the precision value obtained 71% results and for the recall results obtained a value of 46% from 44 data sets that have been tested, there are 25 correct predictions and 19 wrong predictions

Keywords: Confusion Matrix, Majoring student, Naïve Bayes, Prediction

I. PENDAHULUAN

Penjurusan mahasiswa adalah bagian dari suatu fakultas atau sekolah tinggi untuk mengelola suatu bidang studi sebagai pemberi arah tujuan mahasiswa[1]. Hal ini yang mempengaruhi penentuan jurusan. Bagi mahasiswa TI UNISBA Blitar untuk memfokuskan mahasiswa dalam bidang konsentrasi tertentu. Di dalam penjurusan mahasiswa TI UNISBA Blitar memiliki dua penjurusan yaitu penjurusan SC (Sistem Cerdas) dan RPL (Rekayasa Perangkat Lunak).

Saat ini penjurusan mahasiswa TI yang dilakukan oleh prodi TI masih dilakukan secara manual. Penelitian ini bertujuan membantu penjurusan mahasiswa TI dengan menggunakan aplikasi sehingga dapat mengoptimalkan waktu saat penjurusan dan dapat sebagai bahan pertimbangan mahasiswa dengan prediksi yang telah dilakukan di aplikasi tersebut.

Dalam penelitian ini akan dikembangkan aplikasi dengan menggunakan metode *naïve bayes*. Pendekatan ini merupakan pendekatan mengacu pada teorema *bayes* yang merupakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Prinsip ini kemudian digunakan untuk memecahkan masalah klasifikasi[2]. Penggunaan algoritma ini dinilai sesuai karena *naïve bayesian classifier* merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi. Aplikasi yang dikembangkan dalam penelitian ini akan melakukan pengklasifikasian data KHS semester sebelumnya yang memiliki nilai-nilai sesuai pengklarifikasian penentuan jurusan.

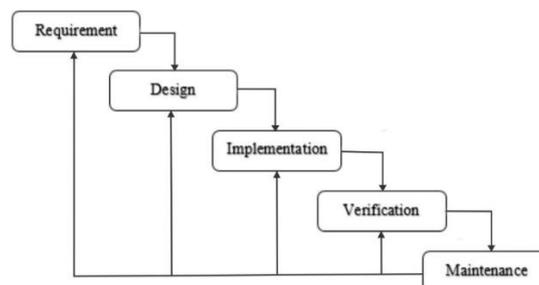
II. KAJIAN PUSTAKA

Penelitian Hartatik pada tahun 2017 yang berjudul “Klasifikasi Konsentrasi Penjurusan Mahasiswa Universitas Amikom Yogyakarta”. Penelitian tersebut bertujuan untuk melakukan klasifikasi penentuan konsentrasi penjurusan yang tepat bagi mahasiswa Universitas AMIKOM Yogyakarta dengan berdasar pada nilai matakuliah di semester 1 sampai dengan semester 4. Hasil dari penelitian tersebut pertama didapatkan hasil berupa perancangan kebutuhan fungsional DFD dan ERD dengan menggunakan metode *Learning Vector Quantization (LVQ)* dan perancangan berbasis destop. Kedua hasil dari penelitian tersebut menggunakan 150 data latih dan 50 data uji didapat nilai akurasi sebesar 76% [3].

Penelitian Heryono dan Acun Kardianawati pada tahun 2018 yang berjudul “Implementasi Metode *Naive Bayes* Untuk Klasifikasi Kredit Motor “. Penelitian ini bertujuan untuk digunakan untuk memprediksi kelayakan kredit untuk menghindarkan terjadinya kredit macet, dan mengevaluasi performance model *Naive Bayes*. dikarenakan OTO Kredit Motor masih sangat kesulitan di dalam menentukan calon debitur yang nantinya akan layak untuk mendapatkan kredit tersebut. Oleh karena itu data mining digunakan untuk dapat memprediksi resiko terjadinya kredit macet dengan melakukan pengklasifikasian terhadap calon debitur kendaraan nantinya. Dari hasil penelitian tersebut didapatkan hasil penelitian untuk perhitungan terhadap pengujian nilai akurasi algoritma klasifikasi *naïve bayes classifier* sebesar 65 %. Kedua penelitian ini mampu mengidentifikasi tingkat kelancaran kredit pada bulan januari hingga maret hal ini di tunjukan dari besar nilai lancer yang meningkat[4].

III. METODELOGI PENELITIAN

Pada bab ini membahas mengenai proses perancangan aplikasi web yang dapat mengelompokkan tingkat pemahaman mahasiswa terhadap mata kuliah yang ditempuh dengan k-means clustering. metode pengembangan Research and Development (R&D). Peneliti mengembangkan produk berupa web aplikasi dengan menggunakan proses pengembangan yang mengacu pada tahapan model pengembangan perangkat lunak waterfall.



Gambar 1. Ilustrasi model *waterfall*

A. SIMULASI PENELITIAN

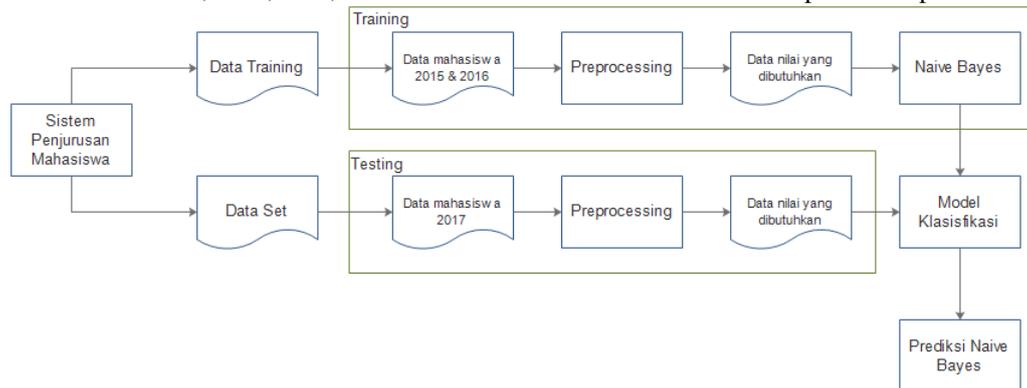
Proses penelitian menggunakan dataset yang didapat dari **pengumpulan data akademi mahasiswa pada tahun 2016 dan 2017**. Data yang dibutuhkan adalah nilai pada mahasiswa yang berpengaruh pada

penjurusan SC dan RPL pada UNISBA Blitar. Isi data yang dibutuhkan mencakup nilai RPL, nilai SBD, nilai SPK, nilai APB nilai PBD, dan Jurusan mahasiswa yang diperoleh mahasiswa. Data training yang akan digunakan adalah tabel sebagai berikut.

Tabel 1. Data Training.

| No | TAHUN | RPL | SBD | SPK | APS | PBD | JURUSAN |
|----|-------|-----|-----|-----|-----|-----|---------|
| 1 | 2016 | A | A | A | A | A | SC |
| 2 | 2016 | A | C | A | B | B | RPL |
| 3 | 2016 | A | A | A | A | A | SC |
| 4 | 2016 | A | A | A | A | A | SC |
| 5 | 2016 | A | A | B | A | A | RPL |
| 6 | 2016 | A | A | B | A | A | RPL |
| 7 | 2016 | A | C | A | A | B | RPL |
| 8 | 2016 | A | B | A | A | C | RPL |
| 9 | 2016 | A | A | A | A | A | SC |
| 10 | 2016 | B | B | A | A | B | RPL |
| 11 | 2016 | A | A | A | A | A | SC |
| 12 | 2016 | A | A | A | A | A | SC |
| 13 | 2016 | A | A | B | A | A | RPL |
| 14 | 2016 | A | C | B | A | B | RPL |
| 15 | 2016 | A | B | A | A | A | SC |
| 16 | 2016 | A | A | B | A | A | RPL |
| 17 | 2016 | A | B | A | A | A | SC |
| 18 | 2016 | A | A | B | B | B | SC |
| 19 | 2016 | B | A | A | B | B | SC |
| 20 | 2016 | A | B | B | B | B | RPL |
| 21 | 2016 | A | B | A | A | B | RPL |
| 22 | 2016 | A | B | B | A | C | RPL |
| 23 | 2016 | A | B | B | A | A | RPL |
| 24 | 2016 | A | B | B | B | B | RPL |
| 25 | 2016 | B | B | B | A | B | RPL |
| 26 | 2016 | B | B | A | A | B | SC |
| 27 | 2016 | B | B | A | A | B | SC |
| 28 | 2016 | A | B | A | B | A | SC |
| 29 | 2016 | A | A | A | A | A | SC |
| 30 | 2016 | A | B | B | A | B | RPL |
| 31 | 2016 | A | A | A | B | A | SC |
| 32 | 2016 | A | A | A | A | A | SC |
| 33 | 2016 | A | B | B | A | A | RPL |
| 34 | 2015 | C | B | B | A | C | SC |
| 35 | 2015 | B | A | B | B | B | SC |
| 36 | 2015 | B | A | B | A | C | SC |
| 37 | 2015 | B | A | B | A | C | SC |
| 38 | 2015 | B | A | B | A | C | SC |
| 39 | 2015 | B | A | B | A | B | SC |
| 40 | 2015 | B | A | B | A | C | SC |
| 41 | 2015 | A | A | B | A | B | SC |
| 42 | 2015 | B | A | B | B | B | SC |
| 43 | 2015 | B | A | B | A | C | SC |
| 44 | 2015 | A | A | B | B | B | RPL |
| 45 | 2015 | A | B | B | B | B | RPL |
| 46 | 2015 | B | A | B | B | B | RPL |
| 47 | 2015 | A | A | A | A | A | RPL |
| 48 | 2015 | A | A | B | A | B | RPL |
| 49 | 2015 | A | A | B | A | B | RPL |
| 50 | 2015 | A | A | B | A | A | RPL |
| 51 | 2015 | B | A | B | B | B | RPL |
| 52 | 2015 | B | A | A | B | B | RPL |
| 53 | 2015 | B | A | B | B | B | RPL |
| 54 | 2015 | B | A | A | A | B | RPL |
| 55 | 2015 | B | A | B | C | B | RPL |

Pada penelitian ini data yang dimiliki selanjutnya dimasukkan dalam format xlsx untuk diunggah ke aplikasi web. Aplikasi ini telah dirancang untuk dapat melakukan prediksi penjurusan mahasiswa. Data yang diolah akan di ambil nilai yang berpengaruh pada saat penjurusan mahasiswa SC atau RPL yaitu nilai pada mata kuliah RPL, SBD, SPK, APS dan PBD. Gambaran umum sistem dapat dilihat pada Gambar 2.



Gambar 2. Sistem Penjurusan Mahasiswa

B. SKENARIO PENGUJIAN

Pada penelitian ini penelitian dilakukan dengan cara memasukan data set nilai mahasiswa yang ada pada kolom web pada aplikasi, dimana data training sebagai data klasifikasi model dengan data set dan dapat mengelompokan jurusan mahasiswa sesuai dengan nilai yang dimasukan.

C. PARAMETER ANALISI DATA

Pada penelitian ini ini digunakan 2 paramater yaitu menggunakan naïve bayes sebagai untuk prediksi penjurusan mahasiswa[5], dan confusion matrix[6] untuk mengukur nilai akurasi sitem dengan membandingkan data set dengan data fakta hasil sebenarnya.

1. Naïve Bayes

Prediksi *Bayes* didasarkan pada teorema *Bayes* dengan formula umum yang dituliskan di persamaan (1):

$$P(H | E) = \frac{P(E|H) \times P(H)}{P(E)} \tag{1}$$

Penjelasan dari formula tersebut adalah sebagai berikut :

$P(H|E)$: Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi.

$P(E|H)$: Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.

$P(H)$: Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

$P(E)$: Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

2. Confusion Matrix

Contoh gambar *Confusion matrix* ditunjukkan pada gambar 3

| | | Predicted class | | Total |
|--------------|-----|-----------------|----|-------|
| | | yes | no | |
| Actual class | yes | TP | FN | P |
| | no | FP | TN | N |
| Total | | P' | N' | P + N |

Gambar 3 Confusion Matrix

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3)$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100 \quad (4)$$

$$\text{Recall} = \frac{TP}{FN+TP} \times 100 \quad (5)$$

Dimana :

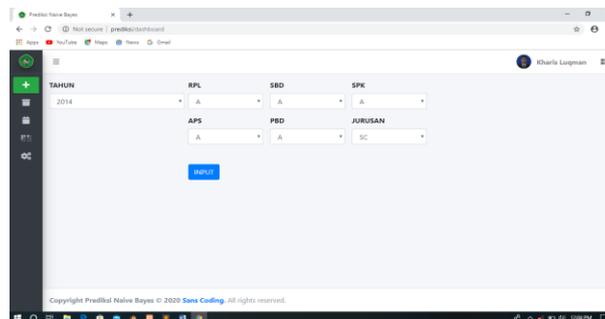
- TP adalah *True Positive*, yaitu jumlah data positif yang terklarifikasi dengan benar oleh sistem.
- TN adalah *True Negative*, yaitu jumlah data negative yang terklarifikasi dengan benar oleh sistem.
- FN adalah *False Negative*, yaitu jumlah data negative namun terklarifikasi salah oleh sistem.
- FP adalah *False Positive*, yaitu jumlah data positif namun terklarifikasi salah oleh sistem .

IV. HASIL DAN ANALISA

Setelah semua rancangan selesai disusun dan sistem juga telah dibangun, maka tahap berikutnya adalah menggunakan sistem tersebut untuk menghasilkan prediksi dari data training terhadap data set, dan menguji nilai akurasi model prediksi yang didapat dari aplikasi dengan data set serta membandingkan nilai confusion matrix berdasarkan hasil prediksi dengan hasil sebenarnya.

A. Penambahan data training

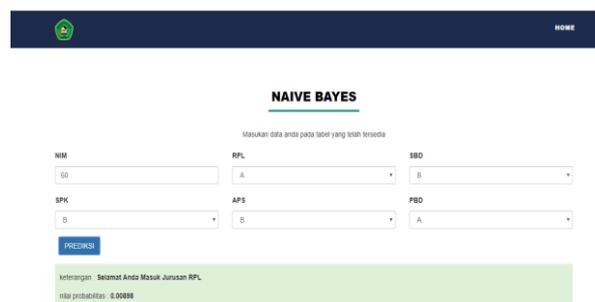
Penambahan data training dilakukan dengan mengisi nilai yang tertera pada kolom seperti gambar untuk menambahkan data training sebagai acuan pengklasifikasian naïve bayes untuk memprediksi penjurusan mahasiswa.



Gambar 4. Halaman Data Trainng

B. Pengujian data set

Pengujian data set dilakukan untuk mendapatkan hasil prediksi penjurusan mahasiswa dengan memasukan nilai sesuai tabel, dan sistem akan mengklasifikasi nilai data set dengan data training dan dipatkan hasil prediksi seperti gambar



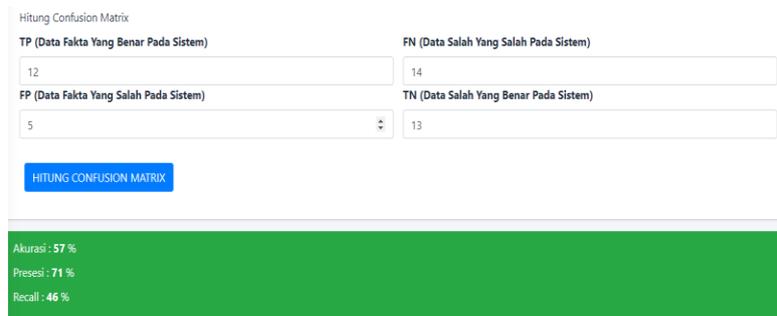
Gambar 5. Halaman Data Set

C. Pengujian nilai akurasi

Pengujian ini dilakukan dengan cara menggunakan metode confusion matrix dimana melakukan terlebih dahulu pengelompokan data prediksi yang dilakukan oleh sistem dan data fakta yang didapat untuk mengetahui atribut pada confusion matrix, hasil yang didapat yakni FP mendapatkan 5, FN mendapatkan 14, TP mendapatkan 12 dan TN mendapatkan 13. Dari data tersebut terdapat 25 data set atau prediksi yang sesuai dengan fakta data dan 19 data yang tidak sesuai dapat dilihat pada tabel sebagai berikut.

Tabel 2. Tabel Confusion Matrix

| SBD | PBD | RPL | APS | SPK | Jurusan | Prediksi | Confusion Matrix |
|-----|-----|-----|-----|-----|---------|----------|------------------|
| A | A | A | A | B | RPL | SC | FP |
| C | B | E | C | C | RPL | RPL | TN |
| B | A | A | A | D | RPL | RPL | TN |
| A | A | A | A | C | SC | SC | TP |
| A | A | A | A | B | SC | SC | TP |
| C | A | A | A | B | RPL | RPL | TN |
| B | A | A | A | B | RPL | RPL | TN |
| B | A | C | A | D | RPL | SC | FP |
| C | A | B | B | C | RPL | RPL | TN |
| C | B | A | A | E | RPL | RPL | TN |
| B | A | A | A | B | SC | RPL | FN |
| B | A | A | A | B | SC | RPL | FN |
| B | A | A | A | B | RPL | RPL | TN |
| B | A | B | A | C | RPL | SC | FP |
| B | A | A | B | B | RPL | RPL | TN |
| B | A | A | B | C | RPL | RPL | TN |
| A | A | A | A | B | SC | SC | TP |
| A | A | A | A | B | SC | SC | TP |
| A | A | A | A | B | SC | SC | TP |
| B | A | A | A | C | RPL | RPL | TN |
| A | A | A | A | C | RPL | SC | FP |
| B | A | A | A | C | RPL | RPL | TN |
| A | A | A | A | B | SC | SC | TP |
| A | A | A | A | B | SC | SC | TP |
| B | B | A | A | C | RPL | RPL | TN |
| A | A | A | A | B | SC | SC | TP |
| A | A | B | A | B | SC | SC | TP |
| A | A | A | A | B | SC | SC | TP |
| B | A | A | A | B | SC | RPL | FN |
| E | B | E | A | B | SC | RPL | FN |
| A | A | A | A | B | SC | SC | TP |
| E | B | B | B | B | RPL | RPL | TN |
| B | B | B | A | B | SC | RPL | FN |
| D | B | C | A | C | SC | RPL | FN |
| B | A | B | A | B | SC | RPL | FN |
| B | A | C | B | B | SC | RPL | FN |
| A | A | B | A | B | SC | SC | TP |
| A | A | B | A | B | SC | SC | TP |
| B | A | B | A | B | SC | RPL | FN |
| B | B | A | A | B | SC | RPL | FN |
| E | B | B | B | B | SC | RPL | FN |
| E | C | B | B | C | RPL | SC | FP |
| B | B | C | B | B | SC | RPL | FN |
| B | B | B | B | B | SC | RPL | FN |
| B | B | D | D | B | SC | RPL | FN |



Gambar 6. Nilai Akurasi Confusion Matrix

Gambar 6 mempresentasikan hasil dari data set yang telah dihitung menggunakan pengujian Confusion Matrix mendapatkan nilai akurasi sebesar 57% dari 44 data set yang telah diuji hal ini menunjukkan bahwa memiliki tingkat akurasi yang tergolong rendah antara tingkat kedekatan nilai prediksi dengan nilai aktual atau fakta sedangkan nilai presisi mendapatkan hasil 71% dari 44 data set yang telah diuji hal ini menunjukkan bahwa tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem tergolong tinggi dan untuk hasil recall didapat nilai sebesar 46% dari 44 data set yang telah diuji hal ini menunjukkan bahwa tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi masih tergolong rendah.

V. KESIMPULAN

Berdasarkan penelitian Penerapan Metode *Naïve Bayes* Dalam Prediksi Penentuan Jurusan Mahasiswa TI dapat disimpulkan hasil sebagai berikut:

1. Cara merancang dan membangun terhadap penerapan metode naïve bayes dalam prediksi penentuan jurusan mahasiswa TI berbasis *Web* adalah dengan menyiapkan data set dan data training yang diperlukan untuk diolah. Selanjutnya disiapkan *database* untuk menampung seluruh data set dan data training yang ada. Untuk membuat aplikasi yang dapat mengolah data set dan data training yang berbeda didalamnya, diperlukan adanya *level* pengguna dengan kombinasi pemanfaatan fitur relasi tabel dalam *database*.
2. Hasil pengujian akurasi terhadap penerapan metode naïve bayes dalam prediksi penentuan jurusan mahasiswa TI berbasis *Web* menggunakan *Confusion Matrix* mendapatkan nilai akurasi sebesar 57% sedangkan nilai presisi mendapatkan hasil 71% dan untuk hasil recall didapat nilai sebesar 46% dari 44 data set yang telah diuji hal ini menunjukkan bahwa tingkat akurasi sistem terbilang cukup sedangkan hasil presisi terbilang tinggi dan untuk recall terbilang rendah.

REFERENSI

- [1] Ruslan and A. Gani, *Bimbingan Penjurusan*. Bandung: Angkasa, 1986.
- [2] F. A. Hermawati, *Data Mining*. Yogyakarta: Andi Offset, 2013.
- [3] Hartatik, "Klasifikasi Konsentrasi Penjurusan Mahasiswa Universitas AMIKOM Yogyakarta," *J. Ilm. DASI*, vol. 18, no. 1, pp. 57–63, 2017.
- [4] H. Heryono and A. Kardianawati, "Implementasi Metode Naive Bayes Untuk Klasifikasi Kredit Motor," *JOINS (Journal Inf. Syst.*, vol. 3, no. 1, pp. 10–21, 2018.
- [5] I. Rish, "An empirical study of the naive Bayes classifier," *IJCAI 2001 Work. Empir. methods Artif. Intell.*, vol. 3, no. 22, 2001.
- [6] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf. Sci. (Ny)*, 2016.