

Application of Data Mining with Classification Methods for Promotion of New Student Admissions at Muhammadiyah University of Sidoarjo Using Web-Based Naïve Bayes Algorithm

Penerapan Data Mining Dengan Metode Klasifikasi Pada Promosi Penerimaan Mahasiswa Baru Universitas Muhammadiyah Sidoarjo Menggunakan Algoritma Naïve Bayes Berbasis Web

Vianti Widyasari¹, Arif Senja Fitrani.S.Kom.,M.Kom²
{viantiw@umsida.ac.id¹, asfjim@umsida.ac.id²}

Program Studi Informatika, Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo,
Indonesia^{1,2}

Abstract. *The University of Muhammadiyah Sidoarjo (UMSIDA) is one of Indonesia's superior and innovative private colleges in developing IPTEKS based on Islamic values for community welfare. Umsida that has stood long enough with the number of students received in each year is quite a lot. Each new school year opening, this private college regularly organizes new student admissions (PMB) activities. Admission for new students (PMB) at Umsida can be done at pmb.umsida.ac.id.*

Therefore, research aims to create data mining applications classification method with the algorithm Naïve Bayes. This research uses the classification method used to measure accuracy level. To predict the promotion of new students receiving Muhammadiyah Sidoarjo University (UMSIDA) can be done using the Naïve Bayes algorithm with 7 predefined variables. Offline and online predictor of the dataset of 2601 data is divided into 2 as many as 70% of 2000 Training data and as much as 30% from 601 of Testing data.

Keywords - Data mining; classification; admission of new students; Naïve Bayes

Abstrak. *Universitas Muhammadiyah Sidoarjo (UMSIDA) merupakan salah satu Perguruan Tinggi Swasta di Indonesia yang Unggul dan Inovatif dalam mengembangkan IPTEKS berdasarkan nilai-nilai islam untuk kesejahteraan masyarakat. Umsida yang telah berdiri cukup lama dengan jumlah mahasiswa yang diterima di tiap tahun cukup banyak. Setiap pembukaan tahun ajaran baru, perguruan tinggi swasta ini rutin menyelenggarakan kegiatan penerimaan mahasiswa baru (PMB). Pendaftaran mahasiswa baru (PMB) di Umsida bisa dilakukan di pmb.umsida.ac.id.*

Oleh karena itu, penelitian bertujuan untuk membuat aplikasi data mining Metode Klasifikasi dengan algoritma Naïve Bayes. Penelitian ini menggunakan metode klasifikasi yang digunakan untuk mengukur tingkat akurasi. Untuk memprediksi promosi penerimaan mahasiswa baru Universitas Muhammadiyah Sidoarjo (UMSIDA) dapat dilakukan dengan menggunakan Algoritma Naïve Bayes dengan 7 variabel yang sudah ditentukan. Hasil prediksi offline dan online dari dataset sebanyak 2601 data dibagi menjadi 2 yaitu sebanyak 70% dari 2000 data Training dan sebanyak 30% dari 601 dari data Testing.

Kata Kunci - Data Mining; Klasifikasi; Penerimaan Mahasiswa Baru; Naïve Bayes

I. PENDAHULUAN

Universitas Muhammadiyah Sidoarjo (UMSIDA) merupakan salah satu Universitas Swasta di Indonesia Perguruan Tinggi Unggul dan Inovatif dalam pengembangan IPTEKS berdasarkan nilai-nilai islam untuk kesejahteraan masyarakat. Umsida yang telah berdiri cukup lama dengan jumlah mahasiswa yang diterima di tiap tahun cukup banyak. Setiap pembukaan tahun ajaran baru, perguruan tinggi swasta ini rutin menyelenggarakan kegiatan penerimaan mahasiswa baru (PMB). Pendaftaran mahasiswa baru (PMB) di Umsida bisa dilakukan di pmb.umsida.ac.id. atau bisa juga datang langsung ke kampus 1 Umsida di Jl. Mojopahit 666B Sidoarjo (depan RSUD Sidoarjo). (Pedoman Akademik Universitas Muhammadiyah Sidoarjo).

Informasi menjadi dasar dari pertimbangan calon mahasiswa baru menentukan keputusan yang tepat terhadap sejumlah alternatif pilihan yang tersedia. Informasi yang lengkap dan jelas mengenai perguruan tinggi yang hendak dituju oleh calon mahasiswa baru menjadi kebutuhan yang terpenuhi.

Beberapa strategi yang digunakan untuk menarik minat calon mahasiswa baru agar mendaftar di Umsida, yaitu strategi promosi baik secara online dan offline. Namun dalam memilih strategi yang tepat untuk promosi itu sangat

penting dalam menarik minat calon mahasiswa baru agar lebih tertarik untuk mendaftar. Oleh karena itu, dalam penelitian ini peneliti ingin menerapkan data mining untuk membantu dalam memprediksi pola dalam mempromosikan UMSIDA.

Dengan menggunakan Metode Klasifikasi Algoritma Naive Bayes yang dimana Algoritma ini secara umum memiliki tingkat akurasi yang cukup tinggi dan cepat dalam mengelola atribut bertipe diskret ataupun numerik. Algoritma Naive Bayes membaca semua data training dari memori atau yang sudah tersimpan. Data training yang diambil adalah data dari Penrimaan Mahasiswa Baru Universitas Muhammadiyah Sidoarjo khususnya data mahasiswa Universitas Muhammadiyah Sidoarjo yaitu tahun 2018. Data ini akan dikelola menggunakan Algoritma Naive Bayes, sehingga Klasifikasi ini dapat memudahkan dalam mengambil keputusan dan kebijakan yang tepat dan efektif dalam pola promosi untuk calon mahasiswa baru.

Data mining adalah adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual[1] . Data mining sangat perlu dilakukan terutama dalam mengelola data yang sangat besar untuk memudahkan aktifitas recording suatu transaksi dan untuk proses data warehousing agar dapat memberikan informasi yang akurat bagi penggunaanya.

II. Metode

A. Data Mining

Data mining merupakan sebuah proses teknik statistik, matematika yang digunakan untuk ekstraksi data atau mencari pola informasi menarik dalam sebuah data berukuran besar yang terpilih dengan menggunakan metode atau algoritma tertentu[1].

B. Pengelompokkan Data Mining

Ada banyak metode atau fungsi data mining yang bisa digunakan untuk menemukan, menggali dan menambang pengetahuan. Ada enam fungsi utama data mining, yaitu [2]:

- Deskripsi (description)
untuk memberi gambaran secara ringkas bagi sekumpulan data yang jumlahnya sangat besar dan banyak jenisnya.
- Estimasi (estimation)
untuk menerangkan sebuah nilai yang belum diketahui, misal menerangkan penghasilan seseorang ketika informasi mengenai orang tersebut diketahui.
- Prediksi
untuk memperkirakan nilai masa mendatang, misal memprediksi penerimaan mahasiswa baru satu tahun ke depan.
- Klasifikasi (classification)
- Klustering
Atau pengklusteran ini meruokan pengelompokan record atau kasus yang memiliki kemiripan data.
- Asosiasi
Bertujuan untuk mengidentifikasi hubungan antar atribut yang muncul dalam suatu waktu.

C. Tahapan Data Mining

Tahap data mining bersifat interaktif, tahapan data mining dibagi menjadi beberapa tahap, yaitu [3] :

- Pembersihan data (Data Cleaning)
Merupakan suatu proses dimana membersihkan atau menghilangkan integrasi data (Data Integration) merupakan penggabungan data dari berbagai database menjadi satu database yang baru[10].
- Seleksi data (Data Selection)
Merupakan penyeleksian data yang tidak sesuai didatabase. Hanya data yang sesuai untuk analisis yang diambil.
- Transformasi data (Data Transformation)
Merupakan pengubahan format data yang sudah diedit agar sesuai dengan proses data mining.
- Proses Mining
Merupakan proses paling utama saat metode yang dipakai untuk menentukan pengetahuan berharga yang tersembunyi dari data.

- Evaluasi pola (Pattern Evaluation)
Merupakan tahap identifikasi pola yang unik ke dalam knowledge base yang ditentukan.
- Presentasi pengetahuan (Knowledge Presentation)
Merupakan penyajian pengetahuan dan visualisasi dari metode yang digunakan untuk mendapatkan pengetahuan yang di peroleh pengguna.

D. Metode Klasifikasi

Metode klasifikasi merupakan metode yang mempelajari sekumpulan data yang dapat menghasilkan suatu pengklasifikasian data baru. Proses klasifikasi pada teknik data mining adalah suatu himpunan data yang dapat menghasilkan model klasifikasi (fungsi target). Jadi, memerlukan dataset pada himpunan untuk proses klasifikasi. Dataset yang digunakan yaitu atribut dan fitur menggunakan data training dan data testing. Teknik klasifikasi menghitung semua data training dan klasifikasi secara lokal memperhitungkan sebagian data training[4].

E. Algoritma Naïve Bayes

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network[9]. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar[5]. Naive Bayes memiliki persamaan seperti berikut [6].

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

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi hipotesis H

P(X) : Probabilitas X

Untuk menjelaskan suatu metode naïve bayes maka perlu diketahui dulu proses klasifikasinya memerlukan petunjuk untuk menentukan sebuah class yang cocok untuk dianalisis.

F. Kebutuhan Program

Web adalah sebuah kumpulan halaman yang diawali dengan halaman muka yang berisikan informasi, iklan, serta program aplikasi[7].

PHP adalah pemrograman interpreter yaitu proses penerjemahan beris kode sumber menjadi kode mesin yang dimengerti computer secara langsung pada saat baris kode dijalankan. PHP disebut sebagai pemrograman Server Side Programing, hal ini dikarenakan seluruh prosesnya dijalankan pada server[7].

MySQL merupakan suatu program database server dimana perangkat lunak tersebut mampu untuk digunakan sebagai transaksi menerima dan mengirim dengan waktu yang singkat pengguna dengan jumlah yang banyak sesuai standar SQL (structured Query Language) yaitu bahasa pemrograman database. MySQL dapat diakses oleh banyak pengguna dan juga membatasi akses berdasarkan privilege (hak user) secara bersamaan[8].

Flowchart merupakan bagan alur yang mirip dengan bagan alur sistem, yaitu untuk menggambarkan prosedur didalam sistem[7].

III. HASIL DAN PEMBAHASAN

A. Lokasi Dan Penelitian

Penelitian ini dilakukan pada bulan September - Desember 2019. Penelitian dilakukan setelah mendapat persetujuan dari dosen pembimbing. Data yang dipergunakan sebagai penelitian adalah data Penerimaan Mahasiswa Baru tahun 2018 (PMB). Penelitian ini dilakukan di Laboratorium Informatika ICT Center Pagerwaja Gelam Kec. Candi Kabupaten Sidoarjo Jawa Timur 61271 Kampus 2 Universitas Muhammadiyah Sidoarjo. Data yang digunakan dalam penelitian ini adalah NIM, Tempat Lahir, Jenis Kelamin, SLTA, Jarak Pendaftaran, Pekerjaan, Pekerjaan Wali, dan Sumber Informasi.

B. Alat dan Bahan Penelitian

Tahap penelitian ini penulis menggunakan alat bantu untuk membuat dan menganalisa sistem yang akan dirancang sebagai berikut:

Perangkat Keras

- Laptop Asus CORE i3
- Processor Inte ® Core™ i3-5010U CPU @ 2.10GHz 2.10GHz
- RAM 4.00 GB
- System Type 64 bit
- Layar 15 inch
- Sublime
- Hypertext Preprocessor (PHP)
- Windows 10 Pro 64 bit
- Xampp
- phpmySQL

C. Pengumpulan Data

Pencarian dan pengumpulan data dilaksanakan terlebih dahulu sebelum melakukan penelitian. Berikut adalah kegiatan yang dilaksanakan dalam pencarian dan pengumpulan data yang mendukung penelitian ini.

D. Studi Literatur

Studi literatur untuk penelitian yang akan menggunakan metode pencarian data dari buku, jurnal, browsing dan internet, yang berhubungan dengan data Penerimaan Mahasiswa Baru (PMB). Teori yang digunakan berasal dari buku jurnal dan penelitian terdahulu yang mendukung penyelesaian masalah pada penelitian yang akan dilakukan.

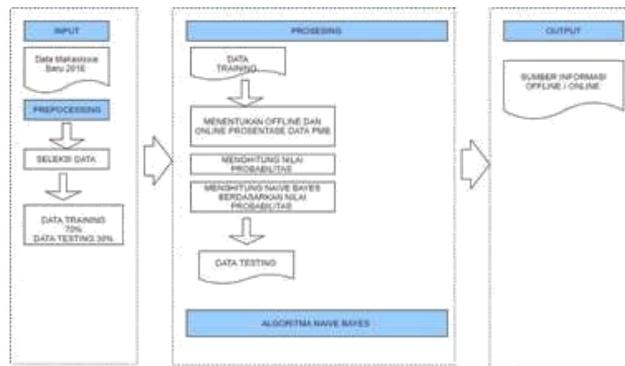
E. Observasi

Metode pengumpulan data dilakukan dengan penelitian di Kantor Penerimaan Mahasiswa Baru Universitas Muhammadiyah Sidoarjo. Sehingga, dengan adanya data yang diperoleh dapat di implementasikan kedalam sistem aplikasi yang ingin dibuat.

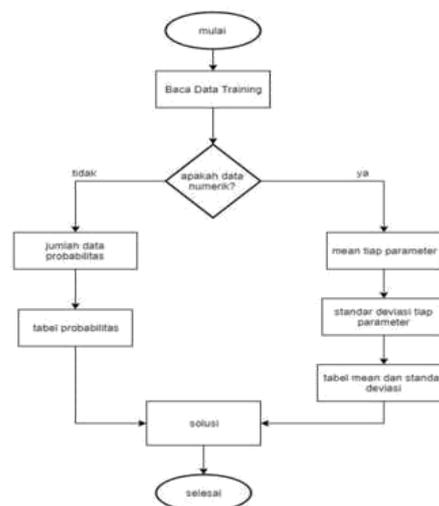
F. Pembahasan

Pada penelitian ini peneliti menggunakan metode klasifikasi naive bayes untuk menguji data komentar dari system E-Comolaint Universitas Muhammadiyah Sidoarjo. Pemilihan metode ini dikarenakan pada beberapa jurnal yang telah disebut pada BAB II menerangkan bahwa metode naive bayes adalah metode klasik dan sederhana namun bisa menunjukkan hasil akurasi perhitungan yang baik.

G. Diagram Klasifikasi



Gambar 1. Diagram Klasifikasi



Gambar 2. Flowchart Data Training

Keterangan dari gambar 2. adalah :

- Membaca data training
- Menghitung jumlah dan probabilitas
- Mendapatkan nilai bayes berdasarkan probabilitas
- Menghasilkan solusi



Gambar 3. Flowchart Data Testing

Keterangan dari gambar 3. adalah

- Input data testing
- Menentukan kelas prediksi

H. Hasil Penelitian

Aplikasi ini dibuat dengan tujuan untuk mengetahui Promosi Data Mahasiswa Penerimaan Mahasiswa Baru (PMB) tahun 2018 di Muhammadiyah Sidoarjo, yaitu dengan sumber informasi meliputi online dan offline. Dataset dibagi menjadi 2 bagian yaitu data training dan data testing. Pada BAB III sudah dijelaskan mulai dengan penginputan data sampai contoh perhitungan manual dataset dengan metode klasifikasi algoritma Naïve Bayes. Dari Perhitungan Manual di Bab III yang mengambil dataset sebanyak 100 data. Data untuk training sebanyak 70 data dari dataset yaitu 30 data. Dari hasil perhitungan manual menghasilkan data jumlah salah dan benar dengan algoritma naïve bayes. Kemudian menghasilkan akurasi kebenaran sebanyak 70% dan 30% salah. Adapun data yang didapat yaitu Penerimaan Mahasiswa Baru Tahun 2018 yang sudah dipilah dan dikategorikan sehingga menjadi 2.601 dataset, data ini dibagi menjadi 2 yaitu 70% yang dijadikan data training yaitu sebanyak 1.821 data dan 30% data testing yaitu sebanyak 780 data. data testing dimana data tersebut merupakan 30% data dari dataset setelah dilakukan pembagian data menjadi 780 data dari 2601 data.

I. Data

Data yang digunakan sebanyak 2.601 data. Dataset dibagi menjadi dua data yaitu Data Training dan Data Testing. Data Training digunakan oleh algoritma klasifikasi untuk membentuk sebuah model classifier. Sedangkan Data Testing digunakan untuk mengukur sejauh mana classifier menentukan data dengan benar.

J. Pre Processing

Data yang belum diproses merupakan data yang masih mentah, maka perlu dipersiapkan terlebih dahulu agar dapat dipakai dalam melakukan proses data mining. Pre Processing merupakan penggabungan data untuk memenuhi hasil klasifikasi pada promosi penerimaan mahasiswa baru. Teknis tersebut dapat di pre processing dalam menyiapkan data ialah dngan membuang duplikasi data, membuang data inkosistensi, dan memperbaiki pada kesalahan data. Maka data dari awal promosi Penerimaan Mahasiswa Baru di Universitas Muhammadiyah Sidoarjo ialah 2.601 data, karena ada data tersebut diseleksi ada data yang kurang atau kosong, dan ada data yang sama. Jadi data promosi penerimaan mahasiswa baru yang dipakai untuk proses klasifikasi dengan algoritma Naive Bayes ialah 100 data dengan 7 atribut dan 1 kelas prediksi (sumber informasi).

IV. Kesimpulan dan Saran

Untuk melihat pengaruh promosi terhadap masuknya mahasiswa baru dapat dilakukan dengan metode Klasifikasi Algoritma Naïve Bayes dengan menggunakan jumlah atribut sebanyak 7 atribut dan 1 class prediksi yaitu sumber informasi (Offline dan Online). Prediksi strategi promosi pada Universitas Muhammadiyah Sidoarjo dengan data yang dikategorikan menjadi 3 yaitu dataset (2.601, 1000, dan 500), dengan pembagian 70% data training dan 30% data testing. Dataset 1000 instance menempati hasil terbaik dari 3 dataset yang lain yaitu dengan nilai prosentase benar sebesar 78%. Dari 3 kategori dataset menunjukkan bahwa jumlah dataset tidak mempengaruhi class prediction dalam bentuk nilai prosentase benar.

REFERENSI

- [1] N. RUMINI, "PREDIKSI KEGAGALAN SISWA DALAM DATA MINING DENGAN MENGGUNAKAN METODE NAÏVE BAYES," 2019.
- [2] M. S. MUSTAFA, M. R. RAMADHAN, AND A. P. THENATA, "IMPLEMENTASI DATA MINING UNTUK EVALUASI KINERJA AKADEMIK MAHASISWA MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER," *CREAT. INF. TECHNOL. J.*, VOL. 4, NO. 2, P. 151, 2018.
- [3] S. M. RIDWAN MUJIB, SUYONO HADI, "PENERAPAN DATA MINING UNTUK EVALUASI KINERJA AKADEMIK MAHASISWA MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER," 2013.
- [4] A. SALEH, "49-104-1-SM(1)," VOL. 2, NO. 3, PP. 207–217, 2015.
- [5] A. JANANTO, "ALGORITMA NAIVE BAYES UNTUK Mencari PERKIRAAN WAKTU STUDI MAHASISWA," *TEKNOL. INF. DIN.*, VOL. 18, NO. 1, PP. 9–16, 2013.
- [6] A. ASRONI, N. MAHARTY ALI, AND S. RIYADI, "PERKIRAAN MASA TUNGGU ALUMNI MENDAPATKAN PEKERJAAN MENGGUNAKAN METODE PREDIKSI DATA MINING DENGAN ALGORITMA NAIVE BAYES CLASSIFIER," *SEMESTA TEK.*, VOL. 21, NO. 2, PP. 189–197, 2018.
- [7] E. COSTS, I. N. PT, AND P. NORTH, "APPLICATION OF WEB-BASED ELECTRIC VOLTAGE INSTALLATION," VOL. 4, NO. APRIL, PP. 32–40, 2019.
- [8] A. B. PUTRA AND S. NITA, "PERANCANGAN DAN PEMBANGUNAN SISTEM INFORMASI E-LEARNING BERBASIS WEB (STUDI KASUS PADA MADRASAH ALIYAH KARE MADIUN)," NO. 2017, PP. 81–85, 2019.
- [9] A. S. FITRIANI, T. INFORMATIKA, F. TEKNIK, AND U. M. SIDOARJO, "PENERAPAN DATA MINING MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES UNTUK MEMPREDIKSI PARTISIPASI PEMILIHAN GUBERNUR," VOL. 3, NO. 2, PP. 98–104, 2019.
- [10] M. SIMANJUNTAK, "IMPLEMENTASI ALGORITMA MERKLE HELLMAN UNTUK KEAMANAN DATABASE," VOL. 4, NO. 1, PP. 46–50, 2019.