

IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFICATION UNTUK KLASIFIKASI KELULUSAN MAHASISWA TEPAT WAKTU (STUDI KASUS: PROGRAM STUDI SISTEM INFORMASI UNIVERSITAS TELKOM)

Novegha Yustira¹, Deden Witarsyah², Edi Sutoyo³

^{1,2,3} Universitas Telkom, Bandung

noveghayustira@telkomuniversity.ac.id¹, dedenwitarsyah@telkomuniversity.ac.id²,
edisutoyo@telkomuniversity.ac.id³

Abstrak

Memprediksi prestasi akademik mahasiswa sangat penting bagi penyelenggara pendidikan karena program strategis tersebut dapat direncanakan dalam meningkatkan atau mempertahankan kinerja mahasiswa selama masa studi di perguruan tinggi. Universitas Telkom merupakan salah satu universitas swasta terbaik di Indonesia. Dapat dikatakan sebagai salah satu universitas terbaik karena Universitas Telkom dapat mencetak lulusan-lulusan yang berkompeten dan memiliki integritas yang baik ditambah lagi para lulusan dari Universitas Telkom memiliki daya saing yang baik di tingkat nasional maupun internasional. Maka dengan mengamati tingkat kelulusan menggunakan *Data Mining* dapat membawa kontribusi yang cukup besar bagi lembaga pendidikan, dalam upaya peningkatan kompetensi kurikulum pada Perguruan Tinggi yang diharapkan hasil *Data Mining* dapat menjadikan referensi standar kurikulum sebagai bentuk peningkatan kompetensi lulusan. Setiap kumpulan atau gudang data dapat memberikan pengetahuan penting yang menjadi informasi yang sangat berharga bagi perguruan tinggi. Pada perguruan tinggi, suatu sistem informasi dapat digunakan untuk memperoleh informasi yang menunjang setiap pada pengambilan suatu keputusan.

Karenanya dibuatlah sistem klasifikasi kelulusan mahasiswa tepat waktu menggunakan model waterfall untuk menyelesaikan permasalahan tersebut. Sehingga klasifikasi kelulusan mahasiswa tepat waktu bisa lebih efektif sesuai dengan informasi yang diberikan oleh sistem yang dibuat. *Data Mining* dengan algoritma *Naïve Bayes*. Metode metode *Naïve Bayes* mengklasifikasikan data menjadi 2 menjadi Tepat dan Terlambat. Penelitian ini menghasilkan akurasi *Naïve Bayes* sebesar 75 %. Sistem pendeteksi tingkat kelulusan mahasiswa ini dapat dikembangkan lagi sesuai kebutuhan sebuah universitas seperti menambah modul-modul yang belum ada seperti pengelolaan rencana studi.

Kata Kunci: *Naïve Bayes*, *Data Mining*, Algoritma, Prediksi

Abstract

Predicting student academic achievement is very important for education providers because these strategic programs can be planned to improve or maintain student performance during their studies in college. Telkom University is one of the best private universities in Indonesia. It can be said to be one of the best universities because Telkom University can produce graduates who are competent and have good integrity plus graduates from Telkom University have good competitiveness at the national and international levels. So by observing the graduation rate using Data Mining can bring a considerable contribution to educational institutions, in an effort to increase curriculum competence in Higher Education, it is hoped that the results of Data Mining can make reference to curriculum standards as a form of increasing graduate competence. Each data collection or warehouse can provide important knowledge which is invaluable information for higher education. In tertiary institutions, an information system can be used to obtain information that supports any decision making. Therefore, a Classification system for student graduation is made on time using the waterfall model to solve these problems. So that the Classification of student graduation on time can be more effective according to the information provided by the system created. Data Mining with the Naïve Bayes algorithm. The Naïve Bayes method classifies data into 2 into Right and Late. This study resulted in an accuracy of 75% Naïve Bayes. This student graduation rate detection system can be developed

further according to the needs of a university, such as adding additional modules that do not exist such as managing study plans.

Keywords: Naïve Bayes, Data Mining, Algoritma, Prediction

1. Pendahuluan

Isi pendahuluan sebaiknya mengandung latar belakang, tujuan, rumusan masalah dan metoda penelitian, yang dipaparkan secara tersirat (*implicit*). Tidak perlu menggunakan *subtitle* dalam format pendahuluan. [10 pts]

Memprediksi prestasi akademik siswa sangat penting bagi suatu lembaga pendidikan karena program strategis tersebut dapat direncanakan dalam meningkatkan atau mempertahankan kinerja siswa selama masa studi di Lembaga [6]. Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Kualitas dari sebuah perguruan tinggi selain dapat dilihat dari rata-rata lama lulusannya mendapatkan pekerjaan juga dapat dilihat dari rata-rata masa atau lama studi dari mahasiswanya. Setiap perguruan tinggi akan mempunyai variasi lama studi dari mahasiswanya yang berbeda-beda [9]. Lama studi dari mahasiswa sangatlah penting bagi mahasiswa itu sendiri dan tentunya bagi program studi dalam rangka untuk menentukan langkah-langkah yang diperlukan agar perjalanan studi mahasiswa menjadi lancar. Permasalahan lama studi setiap mahasiswa biasanya disebabkan atau dipengaruhi oleh banyak faktor, sehingga perlu dibuat sistem prediksi kelulusan mahasiswa untuk mengetahui faktor apa saja yang dapat mempengaruhi kelulusan mahasiswa [9].

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data Mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis yang menemukan keteraturan, pola, dan hubungan dalam dataset

berukuran besar [12]. *Data Mining* sendiri memiliki beberapa teknik salah satunya klasifikasi. Klasifikasi merupakan salah satu proses pada *Data Mining* yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar.

Dengan demikian dibutuhkan suatu metode Klasifikasi dengan Algoritma Naive Bayes yang dapat memprediksi dengan akurat tingkat kelulusan tepat waktu pada mahasiswa, maka dibuat suatu sistem pendeteksi kelulusan pada perguruan tinggi menggunakan metode *Naïve Bayes*. Beberapa penelitian sejenis tentang metode klasifikasi *Naïve Bayes* diantaranya adalah Perkiraan Waktu Studi Mahasiswa Menggunakan Metode Klasifikasi dengan Algoritma Naive Bayes dengan menggunakan metode *Data Mining* dengan algoritma *Naïve Bayes* menyebutkan bahwa sistem yang dirancang adalah sistem prediksi tingkat kelulusan mahasiswa menggunakan metode *Data Mining* dengan algoritma *Naïve Bayes* [4]. Kriteria yang digunakan dalam memprediksi tingkat kelulusan adalah NIM, IPK, SKS, kategori kelulusan (tepat, terlambat). Sedangkan dalam penelitiannya Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan *Data Mining* Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya dengan memanfaatkan salah satu metode algoritma klasifikasi yaitu Naive Bayes, akan dihasilkan pola – pola berdasarkan probabilitas pada tiap atribut yang dapat digunakan untuk menentukan apakah mahasiswa lulus tepat waktu atau tidak[9]. Dari data mahasiswa yang terkumpul berjumlah 1.354, data kemudian dilakukan pre-procesing untuk proses mining pada sistem yang dikembangkan dengan basis website dengan memanfaatkan Weka CLI.

Dengan demikian, meskipun di atas telah disebutkan adanya penelitian dengan tema yang serupa dengan penelitian yang peneliti lakukan, akan tetapi mengingat subjek, objek dan tempat penelitian yang berbeda, maka untuk mengatasi

permasalahan diatas maka diterapkan metode *Naïve Bayes Classification*. Dengan menerapkan sistem ini universitas dapat mengetahui lebih dini apakah mahasiswa akan terlambat dalam menyelesaikan masa studinya.

2. Studi Terkait

2.1 Data Mining

Data Mining disebut juga Knowledge Discovery in Database (KDD) didefinisikan sebagai ekstraksi informasi potensial. Dalam *Data Mining* terdapat dua pendekatan metode pelatihan implisit dan tidak dikenal dari sekumpulan data. *Data Mining* biasanya mengolah data dari database dengan ukuran yang besar. Dari data tersebut dilakukan pencarian pola atau tren sesuai dengan tujuan dari penerapan *Data Mining* tersebut. Hasil dari pengolahan *Data Mining* tersebut selanjutnya dapat digunakan untuk pengambilan keputusan maupun analisis yang dibutuhkan[18]. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. *Data Mining* merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [7]. Proses Knowledge Discovery in Database melibatkan hasil proses *Data Mining* (proses pengekstrak kecenderungan suatu pola data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami, yaitu:

- a. *UnSupervised learning*, metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data.
- b. *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai *output* atau label selama proses *training*.

Ada beberapa teknik yang dimiliki *Data Mining* berdasarkan tugas yang bisa dilakukan,

setiap teknik memiliki algoritma masing-masing. Teknik dalam *Data Mining* terbagi menjadi enam kategori, yaitu:

a. Deskripsi

Para peneliti biasanya mencoba menemukan cara untuk mendeskripsikan pola dan trend yang tersembunyi dalam data.

b. Estimasi

Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori.

c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi dimasa depan).

d. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e. Klastering

Clustering lebih ke arah pengelompokan record, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.

f. Asosiasi

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

2.2 Sistem

Beberapa pengertian sistem menurut para ahli adalah sebagai berikut. sistem didefinisikan sebagai serangkaian tindakan yang saling berhubungan dan berkaitan untuk melakukan dan mencapai tugas bersama-sama[8].

Sistem adalah suatu rangkaian yang terdiri dari dua atau lebih komponen yang saling berhubungan dan saling berinteraksi satu sama lain untuk mencapai tujuan dimana sistem biasanya terbagi dalam sub sistem yang lebih kecil yang mendukung sistem yang lebih besar[1].

Dari definisi di atas yang telah dijelaskan di atas, peneliti menyimpulkan sistem adalah kumpulan dari beberapa komponen yang saling terkait satu sama lainnya dalam satu rangkaian yang mempunyai tujuan yang telah ditetapkan, dimana sistem meliputi proses masukan (*input*) dan menghasilkan pengeluaran (*output*) yang sesuai dari sistem tersebut.

2.3 Informasi

Informasi ibarat denyut nadi yang selalu berdetak di dalam tubuh manusia, maksud istilah kalimat tersebut yaitu, informasi sangat berperan penting pada suatu organisasi. Informasi dapat didefinisikan sebagai berikut :

Informasi adalah fakta yang telah diolah dengan cara tertentu yang menggambarkan suatu kejadian nyata untuk diolah agar dapat dipahami dan digunakan dalam pengambilan suatu keputusan[2].

Informasi adalah data yang telah dikelola dan diproses untuk memberikan arti dan memperbaiki proses pengambilan keputusan[1].

Informasi adalah data yang telah diproses kedalam suatu bentuk yang mempunyai arti bagi si penerima dan mempunyai nilai nyata dan terasa bagi keputusan saat itu atau keputusan mendatang[5].

Jadi, peneliti dapat menarik kesimpulan bahwa informasi adalah data yang telah diolah sedemikian rupa sehingga dapat berguna bagi penerimanya dan dapat meningkatkan pengetahuan seseorang yang menggunakan data tersebut, dan informasi juga dapat dijadikan bahan pertimbangan untuk pengambilan keputusan.

2.4 Sistem Informasi

Sistem merupakan kelompok unsur yang erat hubungannya dengan yang lain, yang berfungsi bersama untuk mencapai tujuan tertentu. Suatu sistem terdiri dari unsur, komponen, atau variabel yang terorganisir, saling berinteraksi, saling bergantung satu sama lain, dan terpadu [13]. Sistem dapat didefinisikan sebagai kumpulan dari elemen yang berupa data, jaringan kerja dari prosedur yang saling berhubungan, sumber daya manusia, teknologi software dan

hardware yang saling berinteraksi sebagai kesatuan untuk mencapai tujuan tertentu[3].

Sedangkan informasi dapat didefinisikan sebagai hasil dari pengolahan data dalam suatu bentuk yang lebih berguna dan berarti bagi penerimanya, menggambarkan kejadian yang nyata untuk membantu penerima informasi dalam mengambil keputusan. Informasi merupakan data yang telah diolah untuk digunakan dalam pengambilan keputusan [13]. Informasi dikatakan tepat waktu jika dihasilkan pada saat yang diperlukan. Relevansi informasi adalah ketika informasi berhubungan dengan kepentingan pengambilan keputusan yang telah direncanakan.

Sehingga dapat disimpulkan bahwa sistem informasi adalah sistem yang menyediakan informasi dengan cara sedemikian rupa sehingga bermanfaat bagi penerima. Secara lebih detil, sistem informasi dapat didefinisikan sebagai seperangkat entitas yang terdiri dari hardware, software dan *brainware* yang saling bekerjasama untuk menyediakan data yang diolah sehingga berguna dan bermanfaat bagi penerima data tersebut.

2.5 Naïve Bayes Classification

Naïve Bayesian Classification (NBC) adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap pengujian/ klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan vocabulary, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel, sedapat mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen. Pada tahap pengujian/ klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi [17]. Semua klasifikasi Bayes merupakan klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Untuk klasifikasi Bayes sederhana dikenal sebagai *Naïve Bayesian Classifier* yang dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari

atribut-atribut lain. Asumsi ini disebut class conditional independence yang dibuat untuk memudahkan dalam perhitungan [15].

Teorema Bayes merupakan teorema yang mengacu konsep probabilitas bersyarat (Tan et al., 2016). Teorema Bayes dapat dinotasikan pada persamaan berikut:

$$P(X) = \frac{P(c)P(c)}{P(x)} \quad (1)$$

Keterangan

x	:	Data dengan class yang belum diketahui
c	:	Hipotesis data merupakan suatu class spesifik
P(c x)	:	Probabilitas hipotesis berdasar kondisi (posteriori probability)
P(c)	:	Probabilitas hipotesis (prior probability)
P(x c)	:	Probabilitas berdasarkan kondisi pada hipotesis
P(x)	:	Probabilitas c

2.6 Penelitian Serupa

Penelitian dengan judul Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi) [16]. *Naive Bayes* dapat melakukan pengklasifikasian dengan metode probabilitas dan statistik, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Nilai Presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode *Naive Bayes* Clasification. Implementasi *Naive Bayes* menggunakan aplikasi WEKA dapat menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus. Pengelompokan Pilihan Lulus dilakukan berdasarkan atribut terpilih yaitu Prodi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata.

Penelitian dengan judul “OPTIMALISASI ALGORITMA C4.5 MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI KELULUSAN SISWA SMKN 2 CIMAHI” [14]. Pada angkatan tahun 2015 yang berjumlah 300 siswa, siswa yang tidak lulus pada ujian nasional

berjumlah 1 siswa, pada angkatan tahun 2016 yang berjumlah 299 siswa, siswa yang tidak lulus pada ujian nasional berjumlah 1 siswa, pada angkatan tahun 2017 yang berjumlah 302 siswa 100% siswa lulus tepat waktu, namun pada tahun yang akan datang masih belum diketahui apakah siswa akan lulus 100% atau tidak. Maka dari itu penelitian ini dilakukan untuk memprediksi kelulusan siswa dengan menggunakan algoritma C4.5+ algoritma genetika. Algoritma C4.5 mempunyai tingkat accuracy sebesar 99,78%, precision 99,78%, recall 100% dan execution time 1 second. Sedangkan algoritma C4.5 + algoritma genetika mempunyai tingkat accuracy sebesar 99,78%, precision 99,78%, recall 100% AUC 0,500, dan execution time 36 second dan setelah dilakukan pengujian T-test antara algoritma C4.5 dan algoritma C4.5 + algoritma genetika dikatakan signifikan jika $\alpha=0.050$. Dalam hasil penelitian ini didapat $\alpha=1.000$.

Dalam penelitian berjudul “Prediksi Rating Film Menggunakan Metode *Naive Bayes*”[10]. Agar suatu film dapat terus berkembang, tentunya membutuhkan penilaian-penilaian dari para penikmat film, untuk mengetahui selera film yang sesuai dengan para penikmat film. Untuk itu dibutuhkan analisis agar dapat mengetahui bagaimana minat penikmat film yaitu dengan membuat penilaian-penilaian yang nantinya digunakan untuk mengetahui rating suatu film menggunakan metode *Naive Bayes* yaitu metode yang melakukan pendekatan statistika yang fundamental dalam pengenalan pola (pattern recognition). Pendekatan ini didasarkan pada kuantifikasi trade-off antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan resiko yang ditimbulkan dalam keputusan-keputusan tersebut. Metode tersebut merupakan salah satu metode dari *Data Mining*, dengan atribut yang sudah ditentukan, yaitu meliputi genre film, aktor film, bahasa, warna, durasi film, negara, dan lainnya yang dapat digunakan sebagai tolak ukur sutradara untuk membuat film.

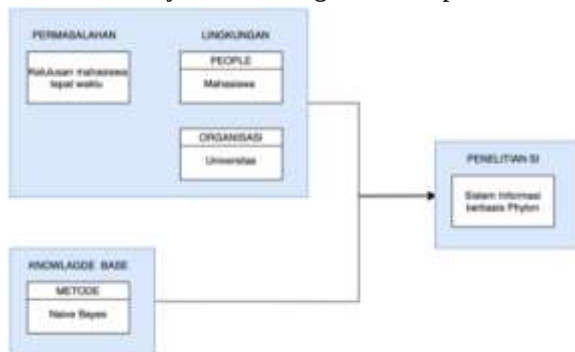
Berdasarkan beberapa penelitian yang telah dijelaskan diatas maka dapat disimpulkan bahwa penelitian ini menggunakan metode *Naive Bayes* untuk mengelompokkan mahasiswa yang dapat lulus tepat waktu berdasarkan rekam data

akademik yang dimiliki sehingga dapat dijadikan acuan oleh perencana akademik dalam membuat keputusan strategis untuk mahasiswa agar dapat lulus tepat waktu.

3. Metode Penelitian

Model konseptual menunjukkan hubungan logis antara variabel-variabel penelitian yang telah diidentifikasi pada tahapan sebelumnya. Variabel yang digunakan adalah variabel pada aspek keberlanjutan (sosial, lingkungan dan ekonomi) dan kinerja industri kecil dan menengah [11]. Kerangka konseptual dibangun berdasarkan teori yang sudah ada maupun dokumen-dokumen penelitian terdahulu sehingga terintegrasi sebagai satu kesatuan.

Berikut ini model konseptual dari penelitian ini akan dijelaskan oleh gambar 1 seperti berikut.

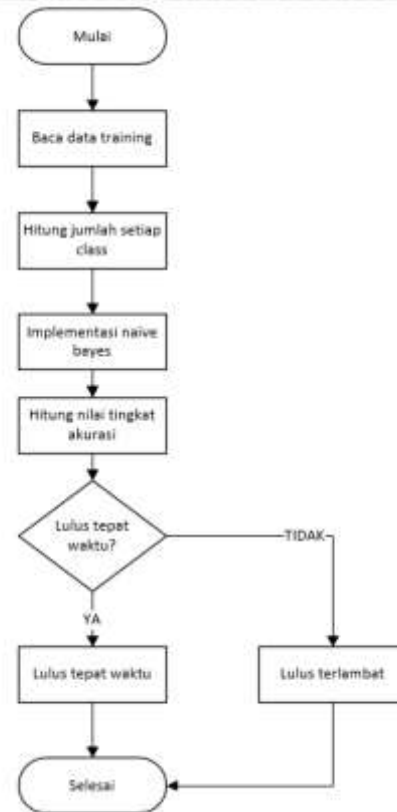


Gambar 1 Model Konseptual

Dari Gambar 1 dapat disimpulkan bahwa dalam pembuatan permodelan prediksi tingkat kelulusan mahasiswa di program studi Sistem Informasi Universitas Telkom didasari oleh permasalahan yang terdapat pada lingkungan program studi Sistem Informasi Universitas Telkom adalah tingkat kelulusan mahasiswa yang tepat waktu. Pelaku yang terlibat dalam lingkungan tersebut adalah mahasiswa dan perencana akademik. Perancang akademik nantinya akan membuat keputusan dengan menggunakan data rekam akademik mahasiswa. Dalam melakukan penelitian ini peneliti menggunakan metode algoritma *Naïve Bayes* classification. Penelitian ini akan menghasilkan sebuah permodelan prediksi tingkat kelulusan mahasiswa.

3.1 Perancangan

Dalam membangun model prediksi untuk penelitian ini terdapat dua tahap proses yang saling berhubungan yaitu tahap *training* dan tahap testing yang digambarkan dalam gambar 2 berikut:



Gambar 2 Flowchart

Dari tahapan tersebut, dilakukan perhitungan Algoritma Naive Bayes menggunakan data *training* dan Data testing sebagai data uji.

Tahap Terakhir adalah implementasi. Implementasi merupakan tahap pengujian dari program aplikasi yang berhasil dibuat, kemudian dijalankan di komputer yang berbeda, setelah itu dilakukan analisis dari hasil program aplikasi yang dijalankan di komputer yang berbeda. Pengujian juga dilakukan dengan hasil yang akan menampilkan hasil klasifikasi dari kelulusan mahasiswa.

3.2 Analisa Studi kasus

Mahasiswa yang mengikuti kegiatan pendidikan akademik untuk program sarjana dengan beban belajar mahasiswa paling sedikit 144 sks dengan batas maksimal 160 sks termasuk

tugas akhir. Dalam kurikulum akademik yang sudah dirancang mahasiswa diharapkan agar dapat menyelesaikan masa studinya selama 3,5 – 4 tahun, tapi perlu diingat sekarang rata-rata tingkat kelulusan mahasiswa lebih dari 4 tahun bahkan sampai dengan 7 tahun yang mana hal tersebut dianggap terlambat dalam menyelesaikan masa studinya.

Masalah ini dapat memberikan dampak yang merugikan bagi pihak universitas dan juga mahasiswa. Minimnya jumlah mahasiswa yang lulus tepat waktu menyebabkan turunnya kualitas dan mutu universitas, selain itu mahasiswa yang terlalu lama menempuh masa studi memiliki kecenderungan terkena ancaman Drop-Out (DO). Mahasiswa dengan kecenderungan yang demikian umumnya adalah mahasiswa berkebutuhan khusus yang memerlukan bimbingan lebih dini sehingga mereka butuh untuk dibina agar dapat lulus tepat waktu. Oleh sebab itu, penting untuk memberikan perhatian lebih terhadap mahasiswa-mahasiswa yang berpotensi lulus terlambat.

Pentingnya universitas untuk mengetahui mahasiswanya yang terlambat dalam menyelesaikan masa studinya, agar dapat menjadi bahan evaluasi bagi pihak instansi dalam sistem akademiknya dan untuk memberikan solusi kepada mahasiswa yang mengalami keterlambatan sehingga mahasiswa tersebut dapat menyelesaikan masa studinya dengan tepat waktu.

Pada tahap ini dilakukan analisis terhadap studi kasus yang terdapat di Program Studi Sistem Informasi Universitas Telkom. Dalam penelitian ini penulis menggunakan rekam data akademik mahasiswa program studi Sistem Informasi Universitas Telkom mulai dari angkatan 2014 hingga 2016 seperti yang dijelaskan pada tabel IV-1. Data tersebut diperoleh dari bagian Sistem Informasi (SISFO) Universitas Telkom dalam bentuk format CSV pada awal bulan maret tahun 2019. Dan sumber data tersebut akan menghasilkan *output* berupa suatu model yang dapat digunakan oleh perencana akademik dalam hal ini ketua program studi (kaprodi) untuk memprediksi kelulusan mahasiswa serta menghasilkan *output* berupa data yang dapat digunakan pada aplikasi dashboard.

Dalam melakukan transformasi data, sumber data yang akan digunakan harus dibersihkan (cleansing) terlebih dahulu sebelum diklasifikasikan berdasarkan atribut yang sesuai menjadi dua kelompok kelulusan yaitu tepat waktu dan terlambat. Sebelum melakukan klasifikasi data, diperlukan untuk menentukan atribut yang akan digunakan untuk proses klasifikasi berdasarkan tujuan penelitian. Penentuan atribut ini dilakukan berdasarkan referensi dari penelitianpenelitian serupa yang pernah dilakukan. Terdapat dua jenis atribut yang ditentukan dalam klasifikasi data, yang pertama yaitu atribut dependen yang nilainya terikat berdasarkan nilai atribut-atribut lainnya. Atribut dependen yang digunakan dalam penelitian ini adalah status kelulusan mahasiswa yaitu tepat waktu dan terlambat. Atribut kedua yaitu atribut independen yang nilainya tidak terkait dengan nilai atribut lainnya atau bisa disebut atribut predictor.

3.3 Perancangan Data

Dalam penelitian ini menggunakan rekam data mahasiswa program studi Sistem Informasi Universitas Telkom mulai dari angkatan 2014 hingga 2019 sejumlah 2239 records. Data yang digunakan sebagai data *input* berasal dari data mentah berupa data induk mahasiswa (students background) dan data nilai akademik mahasiswa. Data *input* tersebut akan dibagi menjadi dua bagian sebagai berikut.

1. Pembentukan Model Prediksi

Data yang digunakan untuk proses pembentukan model prediksi ini adalah data mahasiswa angkatan 2014 hingga 2016 yang berjumlah 1054 *records*. Kemudian data tersebut dibagi menjadi dua bagian yaitu:

- a. *Training Set*. *Training Set* merupakan data yang digunakan dalam proses pelatihan sistem dan terdiri dari pasangan *data input* dan *data target*. Dari keseluruhan data yang didapatkan tersebut, 80% bagian akan digunakan sebagai *data training*.
- b. *Testing Set*. *Testing Set* merupakan data yang digunakan untuk menguji kemampuan sistem dan juga terdiri dari pasangan *data input* dan *data target*.

Data testing yang digunakan berjumlah 20% dari data yang diperoleh.

2. Implementasi Model Prediksi

Data yang digunakan dalam proses pengujian model prediksi yang telah didapatkan adalah data mahasiswa angkatan 2018 hingga 2019 yang berjumlah 1.015 records.

3.4 Implementasi Naïve Bayes

Data Training yang digunakan adalah 10 orang mahasiswa yang ada pada fakultas Rekayasa Industri dan jurusan Sistem Informasi, seperti yang ada pada tabel 1.

Tabel 1 Data Training

No	NIM	IPK	SKS	Status
1	12171254	3,17	BELUM MEMENUHI SYARAT	tepat
2	12171244	3,14	TELAH MEMENUHI SYARAT	tepat
3	12174499	2,50	TELAH MEMENUHI SYARAT	tepat
4	12179313	2,89	BELUM MEMENUHI SYARAT	tepat
5	12175334	3,30	TELAH MEMENUHI SYARAT	tepat
6	12172772	1,80	TELAH MEMENUHI SYARAT	terlambat
7	12179852	1,90	BELUM MEMENUHI SYARAT	terlambat
8	12175333	2,40	TELAH MEMENUHI SYARAT	terlambat
9	12175313	2,65	BELUM MEMENUHI SYARAT	tepat
10	2016230051	3,30	TELAH MEMENUHI SYARAT	tepat

Data testing yang digunakan adalah sebagai berikut:

Tabel 2 Data Testing

No	NIM	IPK	SKS	Status
1	12171254	3,17	TIDAK LENGKAP	TEPAT

Probabilitas dari data kelulusan adalah sebagai berikut:

$$P(C_i) = P(\text{Status Kelulusan} = \text{"Tepat"}) = 7/10$$

$$P(C_i) = P(\text{Status Kelulusan} = \text{"Terlambat"}) = 3/10$$

Setelah mendapatkan nilai probabilitas dari data training diatas selanjutnya adalah menghitung probabilitas dari masing-masing kriteria.

$$P(\text{IPK} = \text{"B"} | \text{STATUS KELULUSAN} = \text{TEPAT}) = 4/7$$

$$P(\text{IPK} = \text{"B"} | \text{STATUS KELULUSAN} = \text{TERLAMBAT}) = 1/3$$

Berdasarkan perhitungan menggunakan Naïve Bayes mahasiswa dengan nama Itasia Nuryanti memiliki status kelulusan "TEPAT" karena berdasarkan hasil nilai probabilitas kelulusan memiliki nilai 0.0653.

4. Hasil dan Pembahasan

Data yang akan dilatihkan maupun diujikan telah dibagi dan di tiap percobaannya menggunakan Naïve Bayes. Data fitur yang telah dibagi tiap percobaannya dilakukan klasifikasi menggunakan Naïve Bayes.

Tabel 3 Hasil Percobaan

Data Aktual	Data Prediksi	
	Tepat	Terlambat
Tepat	12	3
Terlambat	2	3

Pada percobaan ini digunakan dataset yang telah displit secara random diperoleh data training 80% dan Data testing 20% yang akan diklasifikasikan menggunakan model pembelajaran Naïve Bayes maka hasil yang didapatkan berupa confusion matrix yang kemudian didapatkan hasil 75%.

5. Kesimpulan

Dari hasil penelitian yang telah dilakukan didapatkan sejumlah atribut yang dapat digunakan untuk prediksi ketepatan kelulusan mahasiswa. Atribut tersebut meliputi Fakultas, jurusan, IPK (Indeks Prestasi Kumulatif), dan IPS (Indeks Prestasi semester. Kemudian dengan menggunakan Naïve Bayes digunakan sebagai algoritma untuk prediksi tingkat kelulusan yang mendapatkan hasil 75%.

Referensi

- [1] B, Marshall Romney, dan Steinbart, P. J. (2015). *Sistem Informasi, Edisi Sembilan, Buku Satu, diterjemahkan: Deny Arnos Kwary dan Dewi Fitriasari*. Jakarta: Salemba Empat.
- [2] Hafizar. (2017). Perancangan sistem informasi pendataan karyawan pada perusahaan jasa berbasis web. *SENSI*, 3.
- [3] Hamidini, M. D. (2017). Analisis dan Perancangan Sistem Informasi Pembahasan Secara Praktis dengan Contoh Kasus. Yogyakarta: Deepublish.
- [4] Handayani, L., & Maulida, E. L. (2015). Perkiraan Waktu Studi Mahasiswa Menggunakan Metode Klasifikasi Dengan Algoritma Naive Bayes. *Seminar Nasional Teknologi Informasi Komunikasi Dan Industri*, (November), 138–147.
- [5] Hutahaean, J. (2015). *Konsep Sistem Informasi*. Jakarta: Deepublish.
- [6] Ibrohim, D. R. Z. (2006). PREDICTING STUDENTS' ACADEMIC PERFORMANCE: COMPARING ARTIFICIAL. *Annual SAS Malaysia Forum*.
- [7] Idris, M. (2019). Implementasi Data Mining Dengan Algoritma Naive Bayes Untuk Memprediksi Angka Kelahiran. *Pelita Informatika: Informasi Dan Informatika*, 7(3), 160–167.
- [8] Mulyati, M., Tarmizi, R., & Panugali, A. (2018). Sistem Informasi Absensi Berbasis Web Pada Badan Penanggulangan Bencana Daerah Kota Tangerang. *ICIT Journal*, 4(2), 117–127.
- [9] Pambudi, R. D., Supianto, A. A., & Setiawan, N. Y. (2019). Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2196, 3(3), 2194–2200.
- [10] Pratiwi, R. W., & Nugroho, Y. S. (2017). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *DutaCom Journal*, 12(1), 91–108.
- [11] Print, I., Online, I., Fandeli, H., Hasan, A., & Amrina, E. (2020). Dampak : Jurnal Teknik Lingkungan Universitas Andalas Model Konseptual Pengaruh Keberlanjutan terhadap Kinerja Industri Kecil dan Menengah. 01, 15–24.
- [12] Rahwali, Rasti Hansun, Seng Wiratama, Y. widya. (2017). Prediksi Kelayakan Masuk Penjurusan IPA Siswa Sekolah Menengah Atas Menggunakan C4.5 (Studi Kasus: SMA Tarakanita Gading Serpong). *Jurnal Telematika*, 10(2), 12–26.
- [13] Setiawan, Ira Anggraeni. Hendro, T. P. N. D. (2017). Klasifikasi Artikel Berita Menggunakan Metode Text Mining Dan Naive Bayes Classifier.
- [14] Sugianto, C. A., & Pujiyanita, N. S. (2019). OPTIMALISASI ALGORITMA C4 . 5 MENGGUNAKAN ALGORITMA GENETIKA Abstrak. 10(02), 46–51.
- [15] Sulaksono, J., & Darsono. (2015). Sistem pakar penentuan penyakit gagal jantung menggunakan metode naive bayes classifier. *Seminar Nasional Teknologi Informasi Dan Multimedia 2015*, 6–8.
- [16] Syarli, S., & Muin, A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22–26.
- [17] Tan, Y., Shenoy, P. P., Chan, M. W., & Romberg, P. M. (2016). On construction of hybrid logistic regression-na ve bayes model for classification. *Journal of Machine Learning Research*, 52(2016), 523–534.
- [18] Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>