

Implementasi Naïve Bayes Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan

I Made Budi Adnyana

Program Studi Sistem Informasi, ITB STIKOM Bali, Denpasar, Indonesia

Email: budi.adnyana@stikom-bali.ac.id

Abstrak—STIKOM Bali mengimplementasikan sistem Tracer Study berbasis website untuk melakukan pelacakan dan mendapatkan umpan balik dari alumni. Salah satu aspek penilaian dari sistem Tracer Study STIKOM Bali adalah perihal waktu tunggu alumni dalam memperoleh pekerjaan. Data yang diperoleh dari penilaian ini belum diolah secara optimal oleh pihak STIKOM Bali. Jika digali lebih mendalam, maka bisa didapatkan suatu pola untuk memprediksi waktu tunggu alumni dalam memperoleh pekerjaan. Pada penelitian ini mencoba menerapkan salah satu teknik data mining, yaitu teknik klasifikasi dengan menggunakan algoritma Naïve Bayes Classification. Naïve Bayes adalah metode klasifikasi yang menghitung sekumpulan probabilitas dari kombinasi nilai dalam dataset. Data sampel diperoleh dari sistem Tracer Study STIKOM Bali, dengan jumlah data 1240 terdiri dari 7 atribut bertipe kategorikal. Uji coba dilakukan dengan menggunakan aplikasi WEKA dengan parameter 10 *folds cross-validation*. Hasil uji coba menunjukkan tingkat akurasi yang dihasilkan sebesar 48.629%, dimana dari 1240 data uji, 603 data berhasil diklasifikasi dengan benar.

Kata Kunci: Prediksi, Klasifikasi, Alumni, Naive Bayes

1. PENDAHULUAN

Dirjen Pembelajaran dan Kemahasiswaan mewajibkan perguruan tinggi untuk mengadakan Tracer Study setiap tahun secara reguler guna memenuhi keperluan data akreditasi, perbaikan proses pembelajaran, serta pengembangan kurikulum di perguruan tinggi [1]. Ristekdikti telah menerapkan standar penilaian yang sama untuk seluruh perguruan tinggi, namun setiap perguruan tinggi mempunyai strategi tersendiri untuk menciptakan lulusan yang berkualitas. STIKOM Bali telah menjalankan sistem Tracer Study berbasis website. Sistem ini berguna untuk melacak dan mendapatkan umpan balik dari alumni. Pada sistem ini terdapat beberapa survey mengenai situasi alumni khususnya dalam hal pekerjaan dan pemanfaatan kompetensi yang diperoleh selama kuliah di STIKOM Bali.

Salah satu aspek penilaian dari sistem Tracer Study STIKOM Bali adalah perihal waktu tunggu alumni dalam memperoleh pekerjaan. Memperoleh pekerjaan adalah salah satu hal yang ingin dicapai oleh lulusan setelah menyelesaikan kuliahnya. Namun pada prosesnya, tentu membutuhkan waktu tunggu dalam memperoleh pekerjaan tersebut. Dimana salah satu tolak ukur keberhasilan suatu perguruan tinggi dalam bidang pendidikan adalah terserapnya lulusan di dunia kerja [2]. Masing-masing perguruan tinggi tentu mempunyai strategi yang berbeda-beda untuk meningkatkan kualitas lulusannya agar memenuhi kebutuhan dunia kerja saat ini. Untuk menjalankan strategi tersebut, pihak perguruan tinggi tentunya membutuhkan suatu informasi dari data alumni, yang dapat digunakan untuk mendukung pengambilan keputusan.

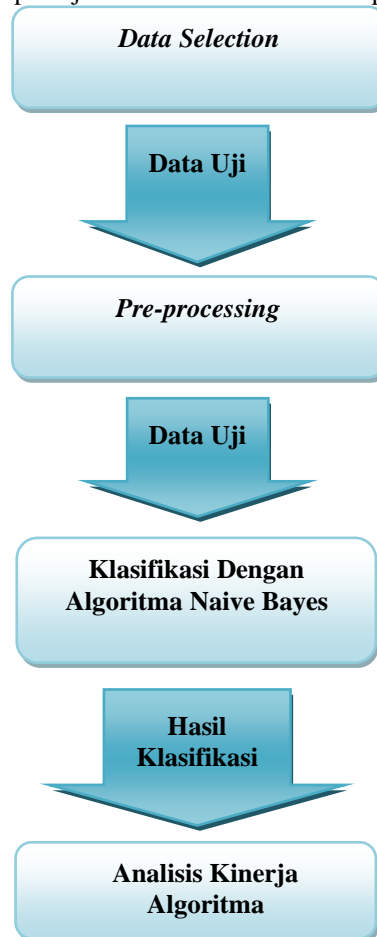
Data Tracer Study STIKOM Bali ini belum dimanfaatkan secara optimal oleh pihak manajemen kampus. Jika digali lebih mendalam, maka bisa didapatkan suatu pola atau pengetahuan dari data ini dengan menerapkan teknik data mining. Data mining dapat digunakan untuk memecahkan masalah dengan cara menganalisa data yang ada dalam database Tracer Study STIKOM Bali. Data mining yang diterapkan dalam dunia pendidikan dikenal dengan istilah EDM (Educational Data Mining) [3]. Pada penelitian yang diusulkan ini mencoba menerapkan teknik klasifikasi untuk memprediksi waktu tunggu alumni dalam memperoleh pekerjaan. Lama waktu tunggu alumni dalam memperoleh pekerjaan ini dapat dipengaruhi oleh berbagai faktor, seperti program studi/konsentrasi alumni, IPK alumni, jenis kelamin, status kelas, lama studi, dan sebagainya.

Data Mining merupakan proses yang menerapkan teknik matematika, statistik, kecerdasan buatan, serta machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [4]. Salah satu metode analisis yang terdapat dalam data mining adalah klasifikasi. Klasifikasi merupakan suatu teknik pengelompokan atau pengkategorian data, dimana data tersebut memiliki class atau label. Algoritma-algoritma yang digunakan untuk penyelesaian klasifikasi dikategorikan kedalam supervised learning. Klasifikasi ini juga dapat digunakan untuk meramalkan kecenderungan data pada masa depan

Salah satu teknik klasifikasi yang populer digunakan adalah Naïve Bayes Classification (NBC). NBC merupakan algoritma klasifikasi yang dapat digunakan untuk melakukan prediksi terhadap probabilitas keanggotaan suatu class dengan menerapkan teorema Bayes. Beberapa penelitian terkait yang menggunakan metode NBC untuk melakukan prediksi, seperti prediksi prestasi siswa [5] dan prediksi pembelian kendaraan [6]. Selain itu, NBC juga digunakan untuk prediksi autisme dini, berdasarkan perbandingan gen individu dengan gen autis yang ada di dataset. Hasil uji cobanya menunjukkan Naïve Bayes sangat efisien dalam melakukan proses klasifikasi gen autis [7]. Berdasarkan uraian diatas, maka pada penelitian ini diusulkan penerapan algoritma Naïve Bayes Classification untuk melakukan prediksi waktu tunggu alumni dalam memperoleh pekerjaan.

2. METODOLOGI PENELITIAN

Penelitian ini terdiri dari beberapa tahapan dimulai dari data selection, data preprocessing, proses klasifikasi dengan algoritma Naïve Bayes, dan terakhir analisis kinerja dari hasil uji coba klasifikasi algoritma Naïve Bayes terhadap data waktu tunggu alumni dalam memperoleh pekerjaan. Gambaran umum alur penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Gambaran umum penelitian

Data Selection adalah suatu proses seleksi data atau pemilihan himpunan data dari sekumpulan data. Kegiatan yang dilakukan pada tahap ini adalah pemilihan data alumni STIKOM Bali serta atribut-atributnya yang relevan terhadap proses analisis yang akan dikerjakan berikutnya. Data hasil seleksi ini lebih lanjut akan digunakan sebagai data uji, disimpan terpisah dengan database operasional STIKOM Bali.

Preprocessing bertujuan untuk pembersihan data, integrasi data, dan transformasi data agar data alumni yang akan digunakan pada proses klasifikasi lebih tepat dan akurat. Tahap *preprocessing* ini dilakukan dengan beberapa langkah, seperti mengisi data yang hilang/kosong, mengurangi data *noise* yang mengandung kesalahan atau data yang *outlier*, dan menangani data yang tidak konsisten

Tahapan klasifikasi menggunakan algoritma Naive Bayes dilakukan dengan *tools* aplikasi WEKA. Aplikasi ini merupakan salah satu aplikasi yang populer digunakan pada data mining karena kehandalannya. Hasil uji coba klasifikasi menggunakan aplikasi WEKA selanjutnya dianalisis tingkat akurasi dan kesalahannya. Pengukuran tingkat akurasi menggunakan pendekatan *Correctly Classified Instances*, sedangkan tingkat kesalahan menggunakan pendekatan Mean Absolute Error (MAE)

3. ANALISA DAN PEMBAHASAN

3.1 Data Selection dan Preprocessing

Dataset yang digunakan pada penelitian ini adalah berupa sekumpulan data lulusan atau alumni STIKOM Bali, data ini diperoleh langsung dari sistem Tracer Study yang ditangani oleh bagian Kemahasiswaan STIKOM Bali. Data yang digunakan adalah lulusan jenjang S1 yang lulus 5 tahun terakhir. Dari hasil seleksi diperoleh sebanyak 1240 data alumni untuk perhitungan proses data mining. Data ini terdiri dari 7 buah atribut bertipe *categorical*, yaitu: Program Studi, Jenis Kelamin, Status Kelas, Konsentrasi, Lama Studi, IPK, dan Lama Memperoleh Pekerjaan. Output dari model yang dikembangkan ini adalah berupa klasifikasi “Lama Memperoleh Pekerjaan”, yang dibagi menjadi 4 kelas, yaitu : “0 – 2 Bulan”, “3 – 6 Bulan”, “7 – 12 Bulan”, dan “Lebih Dari 12 Bulan”. Format dataset alumni yang digunakan sebagai data uji ditunjukkan pada Tabel 1.

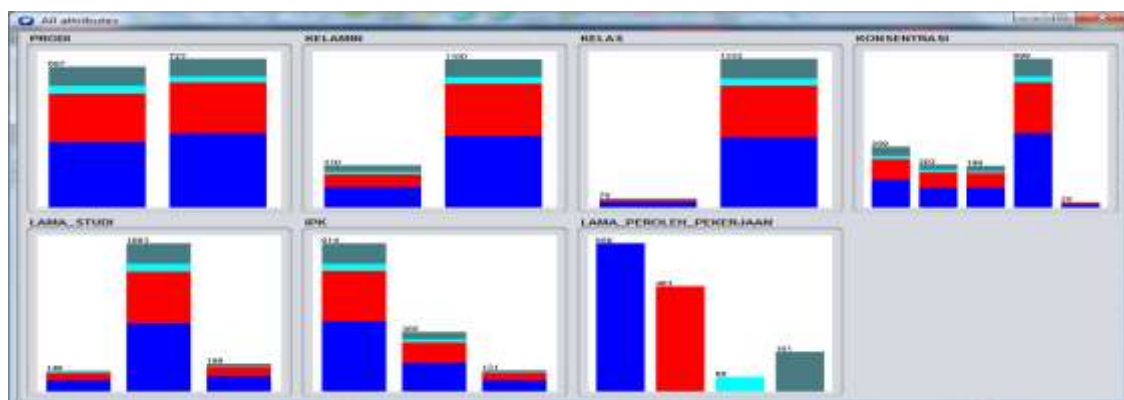
Tabel 1. Format Dataset Alumni

Nama Atribut	Type	Value
Program Studi	Categorical	<ul style="list-style-type: none"> Sistem Komputer Sistem Informasi
Jenis Kelamin	Categorical	<ul style="list-style-type: none"> Laki Perempuan
Status Kelas	Categorical	<ul style="list-style-type: none"> Reguler Karyawan
Konsentrasi	Categorical	<ul style="list-style-type: none"> NTCS RADE ES GM IS
Lama Studi	Categorical	<ul style="list-style-type: none"> Kurang dari 4 tahun 4 – 5 tahun Lebih dari 5 tahun
IPK	Categorical	<ul style="list-style-type: none"> IPK 2.50 – 2.99 IPK 3.00 – 3.49 IPK 3.50 – 4.00
Lama Memperoleh Pekerjaan	Categorical	<ul style="list-style-type: none"> 0 – 2 bulan 3 – 6 bulan 7 – 12 bulan Lebih dari 12 bulan

Agar data uji yang diperoleh ini dapat digunakan pada aplikasi WEKA, maka format data uji tersebut perlu dirubah kedalam bentuk ARFF (Attribute Relation File Format). Proses konversi format data ini dapat dilakukan pada aplikasi WEKA menggunakan fitur ArffViewer. Tampilan data sampel pada ArffViewer dapat dilihat pada Gambar 2. Visualisasi data sampel untuk setiap atributnya dapat dilihat pada Gambar 3.

No	1. PROGRAM STUDI	2. JENIS KELAMIN	3. STATUS KELAS	4. KONSENTRASII	5. LAMA STUDI	6. IPK	7. LAMA PEROLEH PEKERJAAN
70	SISTEM KOMPUTER	LAKI	KARYAWAN	RADE	4 - 5 Tahun	IPK 3.00 - 3.49	3 - 6 Bulan
71	SISTEM KOMPUTER	LAKI	KARYAWAN	RADE	Lebih Dari 5 Tahun	IPK 3.00 - 3.49	3 - 6 Bulan
72	SISTEM INFORMASI	LAKI	KARYAWAN	IS	Kurang Dari 4 Tahun	IPK 3.00 - 3.49	Lebih Dari 12 Bulan
73	SISTEM INFORMASI	LAKI	KARYAWAN	IS	4 - 5 Tahun	IPK 3.00 - 3.49	Lebih Dari 12 Bulan
74	SISTEM KOMPUTER	LAKI	KARYAWAN	NTCS	Lebih Dari 5 Tahun	IPK 2.50 - 2.99	Lebih Dari 12 Bulan
75	SISTEM INFORMASI	LAKI	KARYAWAN	IS	4 - 5 Tahun	IPK 3.50 - 4.00	7 - 12 Bulan
76	SISTEM KOMPUTER	LAKI	KARYAWAN	NTCS	Lebih Dari 5 Tahun	IPK 2.50 - 2.99	7 - 12 Bulan
77	SISTEM KOMPUTER	LAKI	KARYAWAN	NTCS	Kurang Dari 4 Tahun	IPK 3.00 - 3.49	7 - 12 Bulan
78	SISTEM KOMPUTER	LAKI	KARYAWAN	NTCS	Kurang Dari 4 Tahun	IPK 3.00 - 3.49	7 - 12 Bulan
79	SISTEM INFORMASI	PEREMPUAN	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
80	SISTEM INFORMASI	PEREMPUAN	REGULER	ES	4 - 5 Tahun	IPK 2.50 - 2.99	0 - 2 Bulan
81	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.50 - 4.00	0 - 2 Bulan
82	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
83	SISTEM INFORMASI	PEREMPUAN	REGULER	ES	Kurang Dari 4 Tahun	IPK 2.50 - 2.99	0 - 2 Bulan
84	SISTEM INFORMASI	LAKI	REGULER	ES	Lebih Dari 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
85	SISTEM INFORMASI	LAKI	REGULER	ES	Kurang Dari 4 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
86	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
87	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
88	SISTEM INFORMASI	PEREMPUAN	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
89	SISTEM INFORMASI	PEREMPUAN	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
90	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.50 - 4.00	0 - 2 Bulan
91	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.50 - 4.00	0 - 2 Bulan
92	SISTEM INFORMASI	PEREMPUAN	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
93	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
94	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan
95	SISTEM INFORMASI	PEREMPUAN	REGULER	ES	4 - 5 Tahun	IPK 2.50 - 2.99	0 - 2 Bulan
96	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.50 - 4.00	0 - 2 Bulan
97	SISTEM INFORMASI	LAKI	REGULER	ES	4 - 5 Tahun	IPK 3.00 - 3.49	0 - 2 Bulan

Gambar 2. Data sampel pada ArffViewer



Gambar 3. Visualisasi data sampel

3.2 Klasifikasi dengan algoritma Naive Bayes

Naive Bayes adalah metode klasifikasi dengan menggunakan statistic dan probabilitas yang dikemukakan oleh Thomas Bayes, yaitu untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa lampau sehingga dikenal dengan

teorema Bayes. Teorema tersebut dikombinasikan dengan "naive" yang mengasumsikan kondisi antar atribut adalah independen [8].

Uji coba algoritma tersebut dilakukan dengan menggunakan tools *Classify* pada aplikasi WEKA. Pembagian data sampel menggunakan 10 *folds cross-validation*, dimana data sampel dibagi menjadi beberapa kombinasi dengan komposisi 9/10 sebagai *data training* dan 1/10 sebagai *testing*. Berikut dibawah ini adalah hasil uji coba klasifikasi Naïve Bayes dengan menggunakan aplikasi WEKA.

Correctly Classified Instances	603	48.629 %
Incorrectly Classified Instances	637	51.371 %
Kappa	0.1248	
MAE	0.3	
RMSE	0.3919	
RAE	92.6134 %	
RRSE	97.4086 %	
Total Number of Instances	1240	

Berikut dibawah ini adalah *confussion matrix* untuk klasifikasi menggunakan Naive Bayes.

```
=== Confusion Matrix ===
a      b      c      d      <-- classified as
397    144    0      0 | a = 0 - 2 Bulan
255    205    2      0 | b = 3 - 6 Bulan
104    76     1      0 | c = 7 - 12 Bulan
42     14     0      0 | d = Lebih Dari 12 Bulan
```

Dari hasil uji coba diatas, berdasarkan pada nilai *Correctly Classified Instances* menunjukkan bahwa tingkat akurasi yang dihasilkan adalah 48.629% dengan nilai Kappa statistic 0.1248, Sedangkan nilai *Incorrectly Classified Instances* adalah 51.371%. Hal ini menunjukkan dari 1240 data uji, 603 data berhasil diklasifikasi dengan benar, sedangkan 637 data diklasifikasi dengan tidak benar. Tingkat kesalahan (error) dapat dilihat dari pendekatan MAE (Mean Absolute Error) bernilai 0.3.

4 KESIMPULAN

Berdasarkan hasil penelitian diatas dapat diambil kesimpulan: data sampel diperoleh langsung dari sistem Tracer Study STIKOM Bali. Setelah melalui tahap *data selection* dan *preprocessing*, diperoleh 1240 data uji yang terdiri dari 7 atribut bertipe kategorikal. Proses klasifikasi dengan algoritma Naive Bayes menggunakan aplikasi WEKA, dengan parameter pengujian 10 *folds cross-validation*. Berdasarkan hasil uji coba menunjukkan tingkat akurasi yang dihasilkan adalah sebesar 48.629%, dari 1240 data uji hanya 603 data yang berhasil diklasifikasi dengan benar. Sedangkan tingkat kesalahan ditunjukkan dengan MAE (Mean Absolute Error) yang bernilai 0.3.

REFERENCES

- [1] Ristekdikti. (2019). Diakses pada Pelaksanaan Tracer Study Tingkat Perguruan Tinggi: <http://belmawa.ristekdikti.go.id/2016/08/02/pelaksanaan-tracer-study-tingkat-perguruan-tinggi/>
- [2] Amrinda, G A. (2018). Analisis Klasifikasi Waktu Tunggu Kerja Dengan Metode Support Vector Machine Dan Naïve Bayes Classification (Studi Kasus : Waktu tunggu kerja Alumni Universitas Islam Indonesia). Yogyakarta : Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Islam Indonesia
- [3] Ryan S.J.D. Baker and Kalinana Yacef. (2009) : "The State of Educational Data Mining in 2009: A Review and Future Visions,"
- [4] Turban, E., dkk. (2005). Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset.
- [5] U. Pujianto, E. N. Azizah, A. S. Damayanti, "Naive Bayes using to predict students' academic performance at faculty of literature", 2017 5th International Conference on Electrical Electronics and Information Engineering (ICEEIE), pp. 163-169, 2017.
- [6] F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah and C. B. Harahap, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 2018, pp. 1-5.
- [7] R. Reeta, G. Pavithra, V. Priyanka and J. S. Raghul, "Predicting Autism Using Naive Bayesian Classification Approach," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 0109-0113.
- [8] Jiawei Han, Micheline Kamber, Jian Pei. (2011). Data Mining: Concepts And Techniques 3rd Edition. Urbana-Champaign: Morgan Kaufmann