

PENERAPAN ALGORITMA *NAIVE BAYES* UNTUK PREDIKSI WAKTU TUNGGU ALUMNI MENDAPATKAN PEKERJAAN PADA LEMBAGA PUSAT LAYANAN KARIR STM IK ROYAL

Tasyia Dita Aulia¹, Yessica Siagian¹, Pristiyanilicia Putri^{1*}

¹Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal

**email*: pristiyanilicia@gmail.com

Abstract: Getting a job is a goal by graduates after completing their studies. But in this process, getting a job inevitably requires waiting time. If one measure of success in the field of education of a tertiary institution is the number of graduates entering the workforce. So to implement this goal, alumni data is needed that can be used in decision making. Existing tracer study data at the STM IK Royal Career Service Center (LPLK) have not been utilized optimally by the campus. If we dig deeper, we will get some knowledge from this data by applying data mining techniques. Therefore, it is clear that using data mining to predict alumni waiting time is important for a university to determine its strategy. The purpose of this study is to predict the waiting time for alumni to get a job using data mining by applying the Naive Bayes algorithm. Training and testing data were taken from 193 data tracer study alumni of STM IK Royal in 2021 who had graduated and were working. The criteria used to predict alumni waiting time are gender, major, year of entry, year of graduation, GPA, and length of time getting a job. The results of modeling using the Naive Bayes algorithm produce an accuracy of 94%.

Keyword: data mining; naive bayes; waiting time

Abstrak: Mendapatkan pekerjaan merupakan tujuan oleh lulusan setelah menyelesaikan studi mereka. Tetapi dalam proses ini, mendapatkan pekerjaan pasti membutuhkan waktu tunggu. Jika salah satu ukuran keberhasilan di bidang pendidikan suatu perguruan tinggi adalah banyaknya lulusan yang masuk ke dunia kerja. Maka untuk menerapkan tujuan itu, maka diperlukan data alumni yang dapat digunakan dalam pengambilan keputusan. Data *tracer study* yang ada di Lembaga Pusat Layanan Karir (LPLK) STM IK Royal belum dimanfaatkan secara optimal oleh kampus. Jika kita gali lebih dalam, kita akan mendapatkan suatu pengetahuan dari data ini dengan menerapkan teknik *data mining*. Oleh karena itu, jelas bahwa menggunakan *data mining* untuk memprediksi waktu tunggu alumni penting bagi suatu perguruan tinggi untuk menentukan strateginya. Tujuan dari penelitian ini adalah untuk memprediksi waktu tunggu alumni mendapatkan pekerjaan menggunakan *data mining* dengan menerapkan algoritma *naive bayes*. Data *training* dan *testing* diambil dari 193 data *tracer study* alumni STM IK Royal tahun 2021 yang sudah lulus dan bekerja. Kriteria yang digunakan untuk memprediksi waktu tunggu kerja alumni yaitu jenis kelamin, jurusan, tahun masuk, tahun lulus, IPK, dan lama mendapatkan pekerjaan. Hasil pemodelan dengan menggunakan algoritma *naive bayes* menghasilkan akurasi sebesar 94%.

Kata Kunci: data mining; naive bayes; waktu tunggu

PENDAHULUAN

Perkembangan ilmu pengetahuan di bidang komputer dan teknologi semakin pesat dari waktu ke waktu di berbagai bidang kehidupan yang dapat menghasilkan informasi baru, guna memenuhi kebutuhan informasi. Teknologi informasi telah menjadi kebutuhan hidup yang sangat penting dan dapat mempermudah pekerjaan manusia. Oleh karena itu, teknologi informasi dapat digunakan dalam berbagai bidang. Salah satu contoh pemanfaatan teknologi yaitu pada bidang *data mining* [1].

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Royal Kisaran merupakan perguruan tinggi berbasis komputer di Kabupaten Asahan sejak tahun 2011. Setiap tahunnya meluluskan ratusan alumni yang siap bekerja dan mengisi lapangan pekerjaan.

Sejauh ini di Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Royal sudah melakukan pelacakan terhadap status pekerjaan alumni. Pelacakan alumni perlu dilakukan untuk mengetahui bagaimana status pekerjaan alumni. Namun dari pelacakan tersebut belum pernah dilakukan proses *data mining* untuk memprediksi waktu tunggu alumni mendapatkan pekerjaan.

Berdasarkan latar belakang diatas, peneliti ingin melakukan penelitian dengan menerapkan sistem yang dapat memprediksi waktu tunggu alumni mendapatkan pekerjaan menggunakan metode *Naive Bayes* agar dapat mempermudah pihak lembaga dalam melakukan pendataan waktu tunggu alumni mendapatkan pekerjaan.

Berikut rangkuman riset terdahulu yang sudah dicoba antara lain: "Implementasi *naive bayes* Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan [2]", penelitian ini diusulkan penerapan algoritma *naive bayes classification* untuk melakukan prediksi waktu tunggu alumni dalam memperoleh pekerjaan. 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.

Riset berikutnya yang dicoba dengan judul "Perkiraan Masa Tunggu Alumni Mendapatkan Pekerjaan Menggunakan Metode Prediksi *Data Mining* Dengan Algoritma *Naive Bayes Classifier* [3]", Tujuan yang ingin dicapai pada penelitian ini adalah memprediksi perkiraan masa tenggang atau lama waktu alumni UMY mendapatkan pekerjaan setelah menyelesaikan studi S1. *Algoritma Naive Bayes* dalam memprediksi masa tenggang atau lama alumni mendapatkan pekerjaan memiliki tingkat *accuracy* dari *performance Vector* yaitu 71%, *class precision* yaitu CEPAT 68%, LAMBAT 75%, dan untuk *class recall* yaitu CEPAT 83,02% sedangkan LAMBAT 57,45%.

Pada penelitian dengan judul "Prediksi Masa Tunggu Kerja Alumni Menggunakan *Naive Bayes Classifier* Pada Program Studi Ilmu Komputer Universitas Nusa Cendana [4]", Pada penelitian ini, penulis menggunakan metode *NBC* karena metode ini hanya membutuhkan *training* data dalam jumlah yang kecil untuk proses klasifikasi. Metode ini akan diimplementasikan dengan bantuan aplikasi *RapidMiner 7.3* untuk menghasilkan prediksi masa tunggu kerja dan tingkat akurasi dari hasil prediksi yang diperoleh. Atribut yang digunakan pada penelitian ini adalah Masa Studi, Indeks

Prestasi Kumulatif (IPK), keikutsertaan organisasi, dan penguasaan kompetensi. Hasil prediksi masa tunggu kerja yang diperoleh, disajikan dalam bentuk *confusion matrix* dengan tingkat akurasi sebesar 81,82%.

Penelitian yang dilakukan bertujuan untuk mempelajari dan mengimplementasikan *data mining* dengan metode *naive bayes* dalam memprediksi waktu tunggu alumni mendapatkan pekerjaan. Membantu alumni di STMIK Royal sudah lulus agar memiliki gambaran perihal prediksi waktu tunggu mendapatkan pekerjaan. Merancang aplikasi berbasis *web* untuk memprediksi waktu tunggu alumni mendapatkan pekerjaan dengan menggunakan metode *naive bayes* pada Lembaga Pusat Layanan Karir STMIK Royal.

METODE

Berbicara soal *big data*, istilah yang paling umum digunakan adalah *data mining* dan *knowledge discovery in the database (KDD)*. Proses *KDD* sendiri digunakan untuk mengambil informasi dari *database* yang sudah ada dimana *database* tersebut memiliki tabel-tabel berbeda yang saling terhubung. Proses *KDD* dan *data mining* sering digunakan secara bergantian, meskipun memiliki arti yang berbeda, proses *KDD* dan *data mining* saling terkait. [5]. Adapun tahapan dari proses *KDD*, yaitu :

1. *Data selection*

Pemilihan data harus dilakukan sebelum dimulainya tahap *data mining* di *KDD*. Data terpilih yang digunakan dalam proses *data mining* disimpan dalam *file* terpisah dari basis data operasional.

2. *Pre-processing / cleaning*

Sebelum memulai proses *data mining*, perlu dilakukan pembersihan data di *KDD*. Proses pembersihan melibatkan penghapusan data yang sama, memeriksa data yang bertentangan, dan mengoreksi kesalahan data.

3. *Transformation*

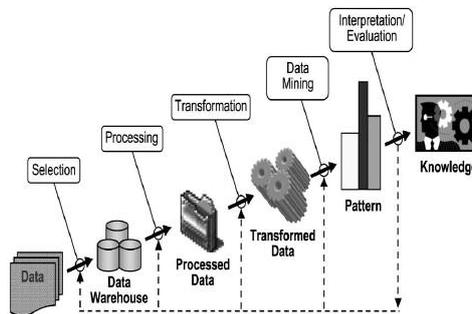
coding adalah perubahan data terpilih sehingga data tersebut cocok untuk proses *data mining*. Proses pengkodean *KDD* adalah proses kreatif dan sangat bergantung pada jenis atau model data yang diambil dari *database*.

4. *Data mining*

Data mining adalah proses pencarian pola atau informasi menarik pada data terpilih dengan menggunakan teknik atau metode tertentu. Teknik atau algoritma *data mining* sangat berbeda. Memilih metode atau algoritma yang tepat sangat bergantung pada tujuan dan keseluruhan proses *KDD*.

5. *Interpretation / evaluation*

Model data yang dibuat oleh proses *data mining* harus disajikan dalam bentuk yang gampang dipahami. Langkah ini merupakan bagian dari proses *KDD* yang disebut *interpretation*. Pada langkah ini, kami memeriksa apakah model atau informasi yang didapatkan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Proses *Knowledge Discovery in Database (KDD)* dapat dilihat pada gambar 1 berikut ini:



Gambar 1. Proses *Data Mining*

Persamaan dari *Teorema Bayes* adalah :

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

Keterangan :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class spesifik*

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

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

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

$P(X)$: Probabilitas X

Untuk menjelaskan metode *naive bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut.

HASIL DAN PEMBAHASAN

Berdasarkan data *training* yang di dapat, peneliti mengambil 30% dari 193 data *training* untuk dijadikan data *testing* pada sistem prediksi waktu tunggu alumni mendapatkan pekerjaan. Data *testing* tersebut dapat dilihat pada tabel 1 berikut ini:

Transformasi Data

Mentransformasi atau mengubah data ke dalam bentuk yang lebih sesuai untuk proses *data mining*. Data akan diubah ke dalam format yang dapat memudahkan proses untuk memprediksi waktu tunggu alumni mendapatkan pekerjaan. Data transformation dapat dilihat pada tabel 2 berikut ini:

Tabel 2. Data Transformasi IPK

IPK	Transformasi
3,05 – 3,33	Interval 1
3,34 – 3,62	Interval 2
3,63 – 3,91	Interval 3

Penerapan Metode Klasifikasi Algoritma *Naive Bayes*

Dalam proses perhitungan algoritma *naive bayes* ada beberapa tahapan proses yang harus di lalui adalah sebagai berikut:

Menghitung jumlah *Class/Label*

$$P(\text{Prediksi} = \text{Cepat}) = 123/193$$

$$P(\text{Prediksi} = \text{Lambat}) = 70/193$$

Jenis Kelamin

$$P(\text{Jenis Kelamin} = \text{Perempuan} \mid \text{Prediksi} = \text{Cepat}) = 82/123$$

$$P(\text{Jenis Kelamin} = \text{Perempuan} \mid \text{Prediksi} = \text{Lambat}) = 40/70$$

$$P(\text{Jenis Kelamin} = \text{Laki-Laki} \mid \text{Prediksi} = \text{Cepat}) = 41/123$$

$$P(\text{Jenis Kelamin} = \text{Laki-Laki} \mid \text{Prediksi} = \text{Lambat}) = 30/70$$

Jurusan

$$P(\text{Jurusan} = \text{Sistem Informasi} \mid \text{Prediksi} = \text{Cepat}) = 105/123$$

$$P(\text{Jurusan} = \text{Sistem Informasi} \mid \text{Prediksi} = \text{Lambat}) = 59/70$$

$$P(\text{Jurusan} = \text{Sistem Komputer} \mid \text{Prediksi} = \text{Cepat}) = 18/123$$

$$P(\text{Jurusan} = \text{Sistem Komputer} \mid \text{Prediksi} = \text{Lambat}) = 11/70$$

Indeks Prestasi Kumulatif (IPK)

$$P(\text{IPK} = \text{Interval 1} \mid \text{Prediksi} = \text{Cepat}) = 23/123$$

$$P(\text{IPK} = \text{Interval 1} \mid \text{Prediksi} = \text{Lambat}) = 12/70$$

$$P(\text{IPK} = \text{Interval 2} \mid \text{Prediksi} = \text{Cepat}) = 52/33$$

$$P(\text{IPK} = \text{Interval 2} \mid \text{Prediksi} = \text{Lambat}) = 36/70$$

$$P(\text{IPK} = \text{Interval 3} \mid \text{Prediksi} = \text{Cepat}) = 48/123$$

$$P(\text{IPK} = \text{Interval 3} \mid \text{Prediksi} = \text{Lambat}) = 22/70$$

Waktu Tunggu (Bulan)

$$P(\text{Waktu Tunggu} = 0-5 \mid \text{Prediksi} = \text{Cepat}) = 122,9/123$$

$$P(\text{IPK} = \text{Interval} = 0-5 \mid \text{Prediksi} = \text{Lambat}) = 1/70$$

$$P(\text{Waktu Tunggu} = \geq 6 \mid \text{Prediksi} = \text{Cepat}) = 0,1/123$$

$$P(\text{IPK} = \text{Interval} = \geq 6 \mid \text{Prediksi} = \text{Lambat}) = 69/70$$

Kesesuaian Pekerjaan Dengan Bidang IT

$$P(\text{Kesesuaian Pekerjaan Dengan Bidang IT} = \text{Sesuai} \mid \text{Prediksi} = \text{Cepat}) = 70/123$$

$$P(\text{Kesesuaian Pekerjaan Dengan Bidang IT} = \text{Sesuai} \mid \text{Prediksi} = \text{Lambat}) = 41/70$$

$$P(\text{Kesesuaian Pekerjaan Dengan Bidang IT} = \text{Tidak} \mid \text{Prediksi} = \text{Cepat}) = 53/123$$

$$P(\text{Kesesuaian Pekerjaan Dengan Bidang IT} = \text{Tidak} \mid \text{Prediksi} = \text{Lambat}) = 29/70$$

Dari perhitungan probabilitas tersebut, maka dapat di rekap pada tabel 3 sebagai berikut:

Tabel 3. Data Rekap Perhitungan

Variabel	Sub Variabel	Cepat	Lambat	Probabilitas	
				Cepat	Lambat
		123	70	0,6373	0,3627
Jenis Kelamin	Perempuan	82	40	0,667	0,571
	Laki-Laki	41	30	0,333	0,429
Jurusan	Sistem Informasi	105	59	0,854	0,843
	Sistem Komputer	18	11	0,146	0,157
	Interval 2	52	36	0,423	0,514
	Interval 3	48	22	0,390	0,314
Waktu Tunggu (Bulan)	0-5	122,9	1	0,999	0,014
	>=6	0,1	69	0,001	0,986
Kesesuaian Pekerjaan Dengan Bidang IT	Sesuai	70	41	0,569	0,586
	Tidak	53	29	0,431	0,414

Mengalikan semua hasil probabilitas variabel setiap klasifikasi

Lulus Cepat

$$\begin{aligned}
 & P(\text{Jenis Kelamin} = \text{Laki-Laki} \mid \text{Prediksi} = \text{Cepat}) * P(\text{Jurusan} = \text{Sistem Informasi} \mid \\
 & \text{Prediksi} = \text{Cepat}) * P(\text{IPK} = \text{Interval 1} \mid \text{Prediksi} = \text{Cepat}) * P(\text{Waktu Tunggu} = \text{0-5} \\
 & \text{bulan} \mid \text{Prediksi} = \text{Cepat}) * P(\text{Kesesuaian Pekerjaan Dengan Bidang IT} = \text{Sesuai} \mid \\
 & \text{Prediksi} = \text{Cepat}) * P(\text{Cepat}) \\
 & = 41/123 * 105/123 * 23/123 * 122,9/123 * 70/123 * 123/193 \\
 & = 0,333 * 0,854 * 0,187 * 0,999 * 0,569 * 0,6373 \\
 & = 0,0192829
 \end{aligned}$$

Lulus Lambat

$$\begin{aligned}
 & P(\text{Jenis Kelamin} = \text{Laki-Laki} \mid \text{Prediksi} = \text{Lambat}) * P(\text{Jurusan} = \text{Sistem Informasi} \mid \\
 & \text{Prediksi} = \text{Lambat}) * P(\text{IPK} = \text{Interval 1} \mid \text{Prediksi} = \text{Lambat}) * P(\text{Waktu Tunggu} = \\
 & \text{0-5 bulan} \mid \text{Prediksi} = \text{Lambat}) * P(\text{Kesesuaian Pekerjaan Dengan Bidang IT} = \text{Sesuai} \\
 & \mid \text{Prediksi} = \text{Lambat}) * P(\text{Lambat}) \\
 & = 30/70 * 59/70 * 12/70 * 1/70 * 41/70 * 70/193 \\
 & = 0,429 * 0,843 * 0,171 * 0,014 * 0,586 * 0,3627 \\
 & = 0,0001879
 \end{aligned}$$

Menghitung nilai probabilitas untuk mendapatkan hasil prediksi

$$\text{Probabilitas dapat kerja cepat} = \frac{0,0192829}{(0,0192829 + 0,0001879)}$$

$$= 0,9903483$$

$$= \frac{0,0192829}{0,0194786}$$

$$\text{Probabilitas dapat kerja lambat} = \frac{0,0001879}{(0,0001879 + 0,0192829)}$$

$$= \frac{0,0001879}{0,0194709}$$

$$= 0,0096517$$

Penentuan Prediksi

Prediksi mendapatkan pekerjaan cepat = 0,9903483

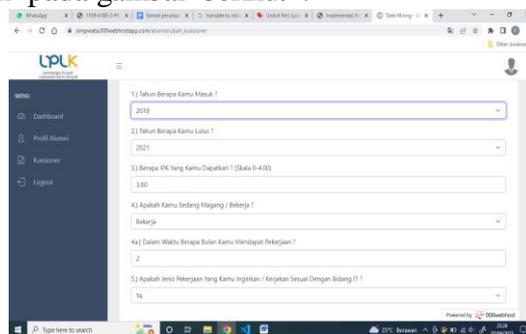
Prediksi mendapatkan pekerjaan lambat = 0,0096517

Jadi, prediksi alumni dengan data tersebut adalah mendapatkan pekerjaan cepat. Berdasarkan perhitungan prediksi data *sheet* dan data *testing* menggunakan algoritma *naive bayes* tingkat akurasi yang didapat menggunakan model *Confusion Matrix* sebesar 94% , dan nilai eror sebesar 6%.

Tabel 4. *Confusion Matrix*

Aktual	Prediksi	
	Dapat kerja cepat	Dapat kerja lambat
Dapat kerja cepat	12	0
Dapat kerja lambat	1	6

Untuk memudahkan alumni dalam memprediksi waktu tunggu mendapatkan pekerjaan, penelitian ini menghasilkan aplikasi prediksi berbasis *web* aplikasi yang dapat diakses kapan saja dan di mana saja. Alumni harus masuk mengakses aplikasi ini, seperti yang ditunjukkan pada gambar berikut :



Gambar 2. Pertanyaan Prediksi

Gambar 2 menunjukkan pertanyaan kepada alumni untuk membuat prediksi waktu tunggu. Setelah proses prediksi selesai, alumni akan mengetahui hasil prediksi menggunakan Algoritma *naive bayes* yang ditampilkan dalam Gambar 3.

Nama	NIM	Jenis Kelamin	Jurusan	Tahun Masuk	Tahun Lulus	IPK	Status	Waktu Tunggu	Ketersediaan Pekerjaan Dengan Bidang IT
YULIA	16220466	Perempuan	Sistem Informasi	2016	2021	3.83	Bekerja	4 Bulan	Ya
NINETH									

Selamat Berdasarkan Data yang Telah Kamu Inputkan, Kamu Terprediksi Mendapatkan Pekerjaan Dengan Rentang Waktu yang Cepat

Follow @lplk@stmikroyal Untuk Mendapatkan Informasi Seperti Lowongan Kerja, Seminar/Workshop, Dan Mengajar

Gambar 3 . Tampilan Hasil Prediksi

Proses prediksi dengan menggunakan Algoritma *naive bayes* dapat dilihat pada Gambar 3, yang merupakan proses utama prediksi.

SIMPULAN

Berdasarkan pembahasan menggunakan algoritma *naive bayes* dan confusion matrix, didapatkan nilai akurasi sebesar 92%, yang termasuk dalam kategori Klasifikasi Baik. Data Mining dengan *naive bayes* merupakan metode yang bagus dan bisa diterapkan untuk memprediksi waktu tunggu alumni mendapatkan pekerjaan berdasarkan pengalaman sebelumnya.

DAFTAR PUSTAKA

- [1] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [2] I. M. B. Adnyana, “Implementasi Naïve Bayes Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan,” *Semin. Nas. Teknol. Komput. Sains*, vol. 1, no. 1, pp. 131–134, 2020, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sainteks/article/view/418>
- [3] 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, doi: 10.18196/st.212225.
- [4] N. D. Rumlaklak *et al.*, “PREDIKSI MASA TUNGGU KERJA ALUMNI MENGGUNAKAN NAÏVE BAYES CLASSIFIER PADA PROGRAM STUDI ILMU KOMPUTER UNIVERSITAS NUSA,” vol. 10, no. 2, pp. 143–150, 2022, doi: 10.35508/jicon.v10i2.7426.
- [5] M. Intan Pratiwi Hant, “Data Mining Technique Using Naïve Bayes Algorithm To Predict Shopee Consumer Satisfaction Among Millennial Generation,” vol. 3, no. 4, pp. 829–838, 2022.