

3272-8657-1-SM

by jiko unkhair

Submission date: 17-Jul-2021 04:26AM (UTC-0400)

Submission ID: 1610537573

File name: 3272-8657-1-SM.docx (171.2K)

Word count: 3499

Character count: 20900

PENERAPAN ALGORITMA NAÏVE BAYES DENGAN BACKWARD ELIMINATION UNTUK PREDIKSI WAKTU TUNGGU ALUMNI MENDAPATKAN PEKERJAAN

Abimanyu Widhiantoyo¹, Betha Nurina Sari², Dadang Yusup³

30
1,2,3 Program Studi Teknik Informatika, Faku 28 Ilmu Komputer, Universitas Singaperbangsa Karawang
Email: ¹abimanyu.widhiantoyo17032@student.unsika.ac.id, ²betha.nurina@staff.unsika.ac.id,
³dadang.dyf@staff.unsika.ac.id

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

Perguruan tinggi memiliki peranan yang signifikan dalam pengembangan kualitas pendidikan manusia. Merancang kurikulum dan strategi pendidikan yang tepat dapat menghasilkan lulusan yang berkualitas. *Tracer Study* menjadi salah satu metode untuk melacak status pekerjaan alumni setelah lulus dari pendidikannya. Fasilkom Unsika adalah salah satu fakultas yang ada di Universitas Singaperbangsa Karawang. Dari banyaknya jumlah lulusan yang dihasilkan, sejauh ini di Fasilkom Unsika belum pernah dilakukan pelacakan terhadap status pekerjaan alumni. Oleh karena itu pelacakan perlu dilakukan untuk nantinya dilakukan proses *Data Mining*. Dari proses *Data Mining* kemudian dihasilkan suatu pengetahuan. Penelitian ini bertujuan untuk memprediksi waktu tunggu alumni mendapatkan pekerjaan dengan menggunakan algoritma *Naïve Bayes* dan dibandingkan dengan algoritma *Naïve Bayes* dengan fitur seleksi *Backward Elimination*. Metodologi *Data Mining* yang digunakan yaitu *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Penelitian menggunakan kelas label CEPAT dan LAMBAT dengan menerapkan sembilan skenario *K-Folds Cross Validation*. Hasilnya menunjukkan bahwa algoritma *Naïve Bayes* dengan fitur seleksi *Backward Elimination* merupakan performa terbaik dengan nilai *Accuracy* 68,52% dan *Kappa* 0,370. Kesimpulan dari penelitian ini yaitu algoritma *Naïve Bayes* dengan fitur seleksi *Backward Elimination* terbukti dapat meningkatkan hasil evaluasi pada prediksi waktu tunggu alumni mendapatkan pekerjaan.

Kata kunci: *Data Mining, Naïve Bayes, CRISP-DM, Prediksi*

IMPLEMENTATION OF NAÏVE BAYES WITH BACKWARD ELIMINATION FOR PREDICTING THE WAITING TIME FOR ALUMNI EMPLOYMENT

Abstract

Universities have a significant role in developing the quality of human education. Designing the right curriculum and educational strategy can produce quality graduates. *Tracer Study* is one of the method to track alumni employment status after graduating from education. Fasilkom Unsika is one of the faculties at Singaperbangsa Karawang University. From the large number of graduates produced, so far at Fasilkom Unsika, no tracking has been carried out on the employment status of alumni. Therefore, tracking needs to be done to later carry out the *Data Mining* process. From the *Data Mining* process, knowledge is then generated. This study aims to predict the waiting time for alumni employment using the *Naïve Bayes* algorithm and compared against the *Naïve Bayes* algorithm with the *Backward Elimination* selection feature. The *Data Mining* methodology used is *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. The study used to FAST and SLOW label classes by applying nine scenarios of *K-Folds Cross Validation*. The results show that the *Naïve Bayes* algorithm with the *Backward Elimination* selection feature achieves the best performance with an *Accuracy* value of 68.52% and *Kappa* 0.37. The conclusion of this study is that the *Naïve Bayes* algorithm with the *Backward Elimination* selection feature is proven to be able to improve the evaluation results in predicting the waiting time for alumni to get a job.

Keywords: *Data Mining, Naïve Bayes, CRISP-DM, Prediction*

1. PENDAHULUAN

Perguruan tinggi memiliki peranan yang signifikan dalam pengembangan kualitas pendidikan manusia. Merancang kurikulum dan strategi

pendidikan yang tepat dapat menghasilkan lulusan yang berkualitas. Secara khusus, perguruan tinggi perlu mengidentifikasi potensi karir bagi mahasiswa. Kemampuan untuk memprediksi waktu tunggu

alumni dalam memperoleh pekerjaan dapat menjadi pengetahuan bagi perguruan tinggi. Pengetahuan tersebut dapat digunakan pihak perguruan tinggi untuk mengevaluasi kinerja dari kurikulum dan strategi pendidikan yang dijalankan selama ini [1].

Setiap perguruan tinggi memiliki strategi dan caranya tersendiri untuk dapat membentuk lulusan yang siap bersaing di dunia kerja. *Tracer Study* menjadi salah satu metode dan hal penting bagi perguruan tinggi untuk melacak jejak alumni setelah lulus dari pendidikannya sampai ke dunia kerja. *Tracer Study* tersendiri menjadi alat penilaian dan juga dijadikan sebagai salah satu syarat kelengkapan akreditasi oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) [2].

Fakultas Ilmu Komputer (Fasilkom) adalah salah satu fakultas yang ada di Universitas Singaperbangsa Karawang (Unsika). Terhitung dari tahun 2012-2020, Fasilkom Unsika telah menghasilkan 1335 lulusan. Dari banyaknya jumlah lulusan tersebut, sejauh ini di Fasilkom Unsika belum pernah dilakukan pelacakan terhadap status pekerjaan alumni saat ini. Oleh karena itu pelacakan perlu dilakukan untuk mengetahui bagaimana status pekerjaan alumni saat ini. Dari pelacakan tersebut nantinya dilakukan proses *Data Mining* untuk kemudian dihasilkan suatu pengetahuan.

Penelitian dengan algoritma *Naive Bayes* telah dilakukan pada penelitian sebelumnya oleh Asroni dkk untuk perkiraan masa tunggu alumni mendapatkan pekerjaan. [3] Penelitian tersebut menggunakan metode *Data Mining* dan algoritma *Naive Bayes* untuk melakukan klasifikasi. Didapat kesimpulan bahwa algoritma *Naive Bayes* dapat digunakan untuk memprediksi masa tenggang alumni mendapat pekerjaan dengan tingkat *Accuracy* dan *Performance Vector* yaitu 71%.

Penelitian selanjutnya oleh Maricar dkk yang perbandingan akurasi *Naive Bayes* dan *KNN* untuk meramalkan status pekerjaan alumni ITB STIKOM Bali. [4] Dihasilkan kesimpulan bahwa metode *Naive Bayes* maupun *KNN* mampu menjadi solusi untuk klasifikasi atau prediksi terhadap data alumni.

Penelitian lainnya oleh Adyana tentang implementasi algoritma *Naive Bayes* untuk memprediksi waktu tunggu alumni dalam memperoleh pekerjaan. [5] Penelitian tersebut menggunakan metode *Naive Bayes* dengan proses *data selection* dan *preprocessing* pada dataset. Evaluasi dihasilkan nilai akurasi sebesar 48,629% dengan parameter pengujian *10-Folds Cross Validation*.

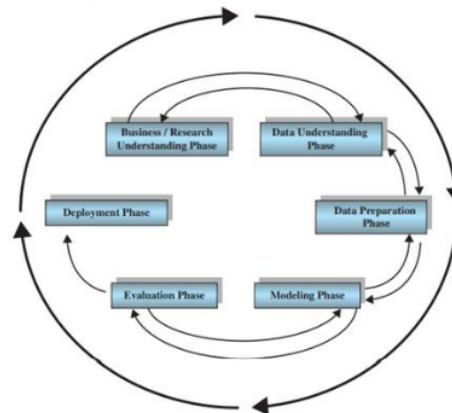
Penelitian ini menerapkan fitur seleksi *Backward Elimination* dan algoritma *Naive Bayes*. *Backward Elimination* merupakan salah satu metode seleksi fitur yang dapat mengoptimalkan akurasi model. Metode ini bekerja dengan cara pemilihan mundur kemudian melakukan pengujian terhadap seluruh atribut-atribut yang dianggap tidak sesuai dan tidak terlalu berpengaruh terhadap pemodelan [6].

Tujuan dari penelitian ini untuk memperoleh tingkat akurasi terbaik antara pemodelan algoritma *Naive Bayes* dengan algoritma *Naive Bayes + Backward Elimination*. Selain itu juga dilakukan analisis terkait atribut paling berpengaruh terhadap proses pemodelan.

2. METODOLOGI PENELITIAN

Penelitian dilakukan dengan menggunakan data status pekerjaan alumni Fasilkom Unsika yang diperoleh melalui proses pelacakan dengan formulir daring yang disebar kepada 1335 alumni dimulai tanggal 22 April 2021 sampai dengan 20 Mei 2021.

Metodologi penelitian yang digunakan sesuai dengan tahapan proses *Cross-Industry Standard Process for Data Mining* (CRISP-DM) sebagai berikut [7]:



Gambar 1. Proses CRISP-DM [8]

2.1. Business Understanding (Pemahaman Bisnis)

Terhitung sejak tahun 2012-2020 Fasilkom Unsika telah menghasilkan ribuan lulusan, dari banyaknya jumlah lulusan tersebut, sejauh ini di Fasilkom Unsika belum pernah dilakukan pelacakan terhadap status pekerjaan alumni saat ini. Penelitian ini bertujuan untuk memprediksi waktu tunggu alumni Fasilkom Unsika mendapatkan pekerjaan. Atribut yang menjadi label yaitu atribut "Lama Waktu Mendapatkan Pekerjaan" yang terbagi ke dalam dua kelas yaitu kelas "CEPAT" dan kelas "LAMBAT".

2.2. Data Understanding (Pemahaman Data)

Data diperoleh melalui pelacakan dengan menggunakan formulir daring yang disebar keseluruhan alumni. Data terkumpul sebanyak 34,53% dari keseluruhan alumni yang terdiri dari 461 *record* data dengan 19 atribut di dalamnya.

2.3. Data Preparation (Persiapan Data)

Pada tahap ini dilakukan proses seleksi data, pembersihan data, transformasi data, dan format data. Berikut ini tabel seleksi data:

5

Tabel 1. Seleksi Data

No	Atribut	Keterangan
1	Nama	×
2	NPM	×
3	Tempat Lahir	×
4	Tanggal Lahir	×
5	Jenis Kelamin	✓
6	Angkatan	✓
7	Tahun Lulus	✓
8	IPK	✓
9	Lama Studi	✓
10	Riwayat Organisasi/Kepanitiaan	✓
11	Pengalaman Magang/Kerja	✓
12	Memiliki Sertifikasi Keahlian	✓
13	Status Pekerjaan	×
14	Lama Waktu Mendapatkan Pekerjaan	×
15	Bidang Pekerjaan	✓
16	Jenis Perusahaan	✓
17	Posisi Pekerjaan	×
18	Kota Tempat Bekerja	✓
19	Kesesuaian Bidang Pekerjaan dengan TI	✓

Keterangan:

- ✓ = Atribut terpilih
- × = Atribut tidak terpilih

Terdapat dua atau lebih data yang menjadi representasi rangkap dari entitas yang sama disebut duplikasi data. Oleh karena itu perlu dilakukan perbandingan terhadap duplikasi data kemudian menghapus salah satunya [9]. Data yang hilang adalah keadaan dimana ada variabel pada atribut yang tidak memiliki nilai untuk dianalisis dan menjadi bias sehingga memungkinkan hasil dari analisis menjadi tidak valid. Salah satu cara untuk mengatasi data yang hilang adalah dengan menghapusnya [10]. Penyimpangan pada data yang bisa terlalu tinggi atau terlalu rendah dibanding dengan data lainnya disebut *data outlier* [11].

Saat dianalisis, pada *dataset* terdapat 4 duplikasi data dan 1 *data outlier*. Selain itu juga terdapat *missing value* pada 7 atribut dengan *missing value* terbesar yaitu 172 *missing value* pada atribut Jenis Perusahaan, Kota Tempat Bekerja, dan Kesesuaian Bidang Pekerjaan dengan TI. Setelah melalui tahap pembersihan data, data dikurangi dari 461 *record* menjadi 270 *record*. Proses dilakukan dengan menghapus duplikasi data, data yang hilang, dan *data outlier*.

Transformasi data dilakukan karena adanya salah penulisan dan juga *data imbalanced*. Klasifikasi yang hanya berjalan pada kelas mayoritas dapat menyebabkan penurunan tingkat akurasi yang disebut juga data dengan kelas yang tidak seimbang (*data imbalanced*) [12]. Transformasi dilakukan pada atribut Kota Tempat Bekerja, Lama Studi, dan Lama Waktu Mendapatkan Pekerjaan.

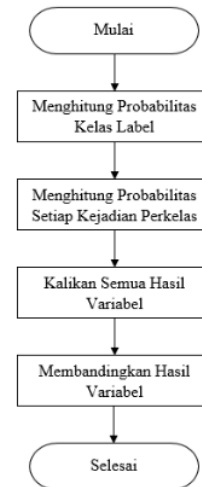
Format data akhir yang telah siap digunakan untuk proses pemodelan yaitu sebanyak 270 *record* data dengan 12 atribut utama dan 1 atribut label.

2.4. Modeling (Pemodelan)

Penerapan algoritma *Naïve Bayes* dilakukan pada tahap pemodelan terhadap dataset yang telah diolah sebelumnya pada fase *data preparation Naïve Bayes* dapat beroperasi lebih baik dalam kasus dunia nyata yang kompleks dan tidak harus memiliki data latih yang banyak dalam proses klasifikasi untuk menentukan parameter ataupun pola. Berikut ini persamaan dari teorema Bayes [13]:

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

Tahapan proses *Naïve Bayes* yaitu pada Gambar 2 berikut [14]:



Gambar 2. Proses *Naïve Bayes*

2.5. Evaluation (Evaluasi)

Hasil dari evaluasi yang didapat nantinya akan ditentukan model mana yang memiliki hasil terbaik. *K-Fold Cross Validation* dapat digunakan untuk meraih perkiraan yang tidak bias dari kinerja model pada jumlah data yang terbatas [15]. Pada proses klasifikasinya, masing-masing bagian akan membagi data secara acak ke dalam k bagian dengan ukuran yang sama. Secara umum penggunaan $k = 10$ atau *10-Folds Cross Validation* sangat dianjurkan karena merupakan jumlah *fold* terbaik untuk uji validitas [16]. *Confusion Matrix* digunakan sebagai alat ukur untuk mengetahui kinerja dari algoritma. Pengujian menggunakan 9 skenario uji yaitu *2-folds*, *3-folds*, ..., *10-folds*. Dari 9 skenario nantinya dipilih skenario yang memiliki hasil kinerja terbaik [17].

2.6. Deployment (Penyebaran)

Disajikan hasil pemodelan algoritma *Naïve Bayes* yang dibandingkan dengan pemodelan algoritma *Naïve Bayes + Backward Elimination*. Analisis juga dilakukan terhadap atribut yang berpengaruh terhadap pemodelan

3. HASIL DAN PEMBAHASAN

Pemodelan terhadap *dataset* dilakukan dengan menggunakan algoritma *Naïve Bayes* dengan tahapan sebagai berikut:

3.1. Hasil

1. Menghitung Probabilitas Kelas Label

Setelah melalui beberapa proses pengolahan data, data yang digunakan yaitu sebesar 270 record data yang terdiri dari total 12 atribut dan 1 kelas label dengan data kategori kelas CEPAT berjumlah 142 dan data kategori kelas LAMBAT berjumlah 128. Probabilitas tiap kelas disajikan pada tabel 2 di bawah ini:

Kelas	Jumlah	Probabilitas Kelas P(H)	
		CEPAT	LAMBAT
CEPAT	142	0,526	
LAMBAT	128		0,474

2. Menghitung Probabilitas Setiap Kejadian Perkelas

Untuk mencari nilai probabilitas setiap kejadian perkelas yaitu dengan menghitung jumlah atribut pada kelas CEPAT dan kelas LAMBAT, kemudian dibagi dengan jumlah kelas yang ada. Nilai probabilitas setiap kejadian perkelas sebagai berikut:

a. Probabilitas atribut Jenis Kelamin dapat dilihat pada tabel 3:

Jenis Kelamin	Probabilitas P(X H)	
	CEPAT	LAMBAT
Laki-Laki	0,845	0,781
Perempuan	0,155	0,219

b. Probabilitas atribut Angkatan dapat dilihat pada tabel 4:

Angkatan	Probabilitas P(X H)	
	CEPAT	LAMBAT
2009	0,056	0,016
2010	0,042	0,008
2011	0,127	0,086
2012	0,183	0,102
2013	0,218	0,141
2014	0,183	0,258
2015	0,106	0,195
2016	0,085	0,195

c. Probabilitas atribut Tahun Lulus dapat dilihat pada tabel 5:

Tahun Lulus	Probabilitas P(X H)	
	CEPAT	LAMBAT
2013	0,042	0,016
2014	0,035	0,008
2015	0,070	0,047
2016	0,021	0,039
2017	0,113	0,094
2018	0,225	0,273
2019	0,296	0,297
2020	0,197	0,227

d. Probabilitas atribut IPK dapat dilihat pada tabel 6:

IPK	Probabilitas P(X H)	
	CEPAT	LAMBAT
Sampel Uji = 3,00	0,472	0,333

e. Probabilitas atribut Lama Studi dapat dilihat pada tabel 7:

Lama Studi	Probabilitas P(X H)	
	CEPAT	LAMBAT
3,5 – 4 Tahun	0,437	0,633
> 4 Tahun	0,563	0,367

f. Probabilitas atribut Riwayat Organisasi/Kepanitiaan dapat dilihat pada tabel 8:

Riwayat Organisasi/Kepanitiaan	Probabilitas P(X H)	
	CEPAT	LAMBAT
1	0,373	0,328
2	0,155	0,164
3	0,246	0,203
4	0,134	0,172
5	0,092	0,133

g. Probabilitas atribut Pengalaman Magang/Kerja dapat dilihat pada tabel 9:

Pengalaman Magang/Kerja	Probabilitas P(X H)	
	CEPAT	LAMBAT
Ada	0,761	0,625
Tidak	0,239	0,375

h. Probabilitas atribut Memiliki Sertifikasi Keahlian dapat dilihat pada tabel 10:

Memiliki Sertifikasi Keahlian	Probabilitas P(X H)	
	CEPAT	LAMBAT
Ada	0,535	0,578
Tidak	0,465	0,422

- i. Probabilitas atribut Bidang Pekerjaan dapat dilihat pada tabel 11:

Tabel 11. Probabilitas Bidang Pekerjaan

Bidang Pekerjaan	Probabilitas (P(XIH))	
	CEPAT	LAMBAT
Pegawai Swasta	0,915	0,969
Pegawai Negeri Sipil	0,085	0,031

- j. Probabilitas atribut Jenis Perusahaan dapat dilihat pada tabel 12:

Tabel 12 Probabilitas Jenis Perusahaan

Jenis Perusahaan	Probabilitas (P(XIH))	
	CEPAT	LAMBAT
Lokal/Daerah	0,472	0,469
Nasional	0,324	0,367
Multinasional	0,204	0,164

- k. Probabilitas atribut Kota Tempat Bekerja dapat dilihat pada tabel 13:

Tabel 13. Probabilitas Kota Tempat Bekerja

Kota Tempat Bekerja	Probabilitas (P(XIH))	
	CEPAT	LAMBAT
Karawang	0,718	0,500
Luar Karawang	0,282	0,500

- l. Probabilitas atribut Kesesuaian Bidang Pekerjaan dengan TI dapat dilihat pada tabel 14:

Tabel 14. Probabilitas Kesesuaian Bidang Pekerjaan dengan TI

Kesesuaian Bidang Pekerjaan dengan TI	Probabilitas (P(XIH))	
	CEPAT	LAMBAT
1	0,331	0,195
2	0,246	0,383
3	0,423	0,422

3. Kalikan Hasil Variabel

Dari perhitungan probabilitas yang telah dilakukan sebelumnya dapat digunakan untuk melakukan proses prediksi pada data uji seperti pada tabel 15 dengan cara mengalikan setiap atribut perkelas.

Tabel 15. Sampel Data Uji

No	Atribut	Keterangan
1	Jenis Kelamin	Laki-Laki
2	Angkatan	2012
3	Tahun Lulus	2018
4	IPK	3,00
5	Lama Studi	> 4 Tahun
6	Riwayat Organisasi/Kepanitiaan	3
7	Pengalaman Magang/Kerja	Ada
8	Memiliki Sertifikasi Keahlian	Ada
9	Bidang Pekerjaan	Pegawai Swasta
10	Jenis Perusahaan	Nasional
11	Kota Tempat Bekerja	Karawang

12	Kesesuaian Bidang Pekerjaan dengan TI	3
13	Lama Waktu Mendapatkan Pekerjaan	?

1. Kelas CEPAT

$$\begin{aligned}
 &= P(\text{CEPAT}) \times P(\text{Jenis Kelamin} = \text{Laki-Laki} \mid \text{CEPAT}) \times P(\text{Angkatan} = 2012 \mid \text{CEPAT}) \times P(\text{Tahun Lulus} = 2018 \mid \text{CEPAT}) \times P(\text{IPK} = 3,00 \mid \text{CEPAT}) \times P(\text{Lama Studi} = > 4 \text{ Tahun} \mid \text{CEPAT}) \times P(\text{Riwayat Organisasi/Kepanitiaan} = 3 \mid \text{CEPAT}) \times P(\text{Pengalaman Magang/Kerja} = \text{Ada} \mid \text{CEPAT}) \times P(\text{Memiliki Sertifikasi Keahlian} = \text{Ada} \mid \text{CEPAT}) \times P(\text{Bidang Pekerjaan} = \text{Pegawai Swasta} \mid \text{CEPAT}) \times P(\text{Jenis Perusahaan} = \text{Nasional} \mid \text{CEPAT}) \times P(\text{Kota Tempat Bekerja} = \text{Karawang} \mid \text{CEPAT}) \times P(\text{Kesesuaian Bidang Pekerjaan dengan TI} = 3 \mid \text{CEPAT}) \\
 &= 0,526 \times 0,845 \times 0,183 \times 0,225 \times 0,472 \times 0,563 \times 0,246 \times 0,761 \times 0,535 \times 0,915 \times 0,324 \times 0,718 \times 0,423 \\
 &= 4,38562\text{E-}05 \\
 &= 0,0000438562
 \end{aligned}$$

2. Kelas LAMBAT

$$\begin{aligned}
 &= P(\text{LAMBAT}) \times P(\text{Jenis Kelamin} = \text{Laki-Laki} \mid \text{LAMBAT}) \times P(\text{Angkatan} = 2012 \mid \text{LAMBAT}) \times P(\text{Tahun Lulus} = 2018 \mid \text{LAMBAT}) \times P(\text{IPK} = 3,00 \mid \text{LAMBAT}) \times P(\text{Lama Studi} = > 4 \text{ Tahun} \mid \text{LAMBAT}) \times P(\text{Riwayat Organisasi/Kepanitiaan} = 3 \mid \text{LAMBAT}) \times P(\text{Pengalaman Magang/Kerja} = \text{Ada} \mid \text{LAMBAT}) \times P(\text{Memiliki Sertifikasi Keahlian} = \text{Ada} \mid \text{LAMBAT}) \times P(\text{Bidang Pekerjaan} = \text{Pegawai Swasta} \mid \text{LAMBAT}) \times P(\text{Jenis Perusahaan} = \text{Nasional} \mid \text{LAMBAT}) \times P(\text{Kota Tempat Bekerja} = \text{Karawang} \mid \text{LAMBAT}) \times P(\text{Kesesuaian Bidang Pekerjaan dengan TI} = 3 \mid \text{LAMBAT}) \\
 &= 0,474 \times 0,781 \times 0,102 \times 0,273 \times 0,333 \times 0,367 \times 0,203 \times 0,625 \times 0,578 \times 0,969 \times 0,367 \times 0,500 \times 0,422 \\
 &= 6,93232\text{E-}06 \\
 &= 0,00000693232
 \end{aligned}$$

4. Membandingkan Hasil Variabel

Nilai probabilitas terbesar dari hasil perhitungan pada kelas label akan dijadikan sebagai masukan data uji tersebut [18]. Dari perhitungan manual sebelumnya dengan mengalikan hasil variabel dengan menggunakan sampel data uji, didapat hasil pada kelas CEPAT dengan nilai probabilitas 4,38562E-05, dan hasil pada kelas LAMBAT dengan nilai probabilitas 6,93232E-06. Apabila dibandingkan dari perhitungan *Naïve Bayes* tersebut, nilai probabilitas tertinggi yaitu pada kelas prediksi (P | CEPAT). Maka diperoleh kesimpulan bahwa data uji tersebut masuk ke dalam kelas prediksi lama waktu mendapatkan pekerjaan CEPAT.

3.2. Pembahasan

Berdasarkan tabel 16 di bawah, maka skenario ke-sembilan dengan *10-Folds Cross Validation* merupakan skenario terbaik dengan nilai *Accuracy* sebesar 67,78% dan *Kappa* 0,356.

Tabel 16 Perbandingan Evaluasi *Naive Bayes*

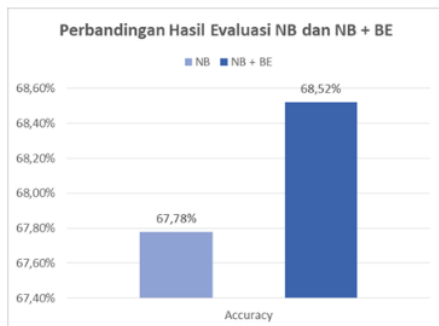
Skenario	k-folds	Accuracy	Kappa
1	2	64,81%	0,293
2	3	64,81%	0,294
3	4	64,07%	0,282
4	5	64,81%	0,296
5	6	67,41%	0,346
6	7	65,19%	0,302
7	8	65,93%	0,315
8	9	67,41%	0,346
9	10	67,78%	0,356

Berdasarkan tabel 17 di bawah, maka skenario ke-sembilan dengan *10-Folds Cross Validation* merupakan skenario terbaik dengan memiliki nilai evaluasi *Accuracy* sebesar 68,52% dan *Kappa* 0,370.

Tabel 17. Perbandingan Evaluasi *Naive Bayes* + *Backward Elimination*

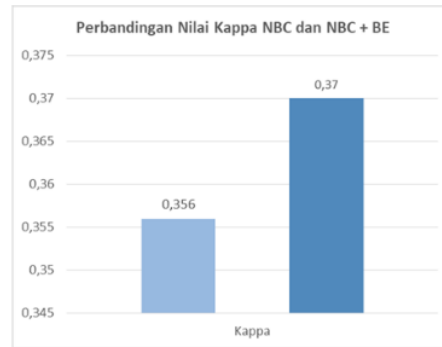
Skenario	k-folds	Accuracy	Kappa
1	2	68,52%	0,370
2	3	67,78%	0,354
3	4	67,04%	0,341
4	5	67,78%	0,354
5	6	67,41%	0,349
6	7	67,41%	0,348
7	8	68,52%	0,370
8	9	67,78%	0,354
9	10	68,52%	0,370

Evaluasi dengan *10-Folds Cross Validation* menggunakan nilai *Accuracy* menghasilkan model *Naive Bayes* dengan fitur seleksi *Backward Elimination* mampu meningkatkan kinerja pemodelan berdasarkan nilai *Accuracy*. Dapat dilihat perbandingannya pada gambar 3 berikut ini.



Gambar 3. Perbandingan Hasil Evaluasi

Evaluasi dengan *10-Folds Cross Validation* menggunakan nilai *Kappa* menghasilkan model *Naive Bayes* dengan fitur seleksi *Backward Elimination* mampu meningkatkan kinerja pemodelan berdasarkan nilai *Kappa* yang meningkat dari 0,356 menjadi 0,370 seperti pada gambar 4 berikut.



Gambar 4. Perbandingan Nilai Kappa

Tabel 18 di bawah menunjukkan bahwa pada fitur seleksi *Backward Elimination*, atribut yang dianggap tidak berpengaruh pada pengujian *10-Folds Cross Validation* yaitu atribut IPK dan atribut Riwayat Organisasi/Kepanitiaan.

Tabel 18. Atribut Berpengaruh *Backward Elimination*

Atribut	Weight
Jenis Kelamin	1
Angkatan	1
Tahun Lulus	1
IPK	0
Lama Studi	1
Riwayat Organisasi/Kepanitiaan	0
Pengalaman Magang/Kerja	1
Memiliki Sertifikasi Keahlian	1
Bidang Pekerjaan	1
Jenis Perusahaan	1
Kota Tempat Bekerja	1
Kesesuaian Bidang Pekerjaan dengan TI	1
Lama Waktu Mendapatkan Pekerjaan	33

4. KESIMPULAN

Dari penelitian yang telah dilakukan dapat disimpulkan bahwa:

- Hasil dari semua pengujian yang telah dilakukan dengan 9 skenario uji menggunakan *K-Fold Cross Validation* diketahui bahwa pada pemodelan *Naive Bayes* dihasilkan performa terbaik dengan *10-Folds Cross Validation*. Pengujian menghasilkan nilai evaluasi *Accuracy* sebesar 67,78% dan *Kappa* 0,356.
- Pada pemodelan *Naive Bayes* dengan fitur seleksi *Backward Elimination* menghasilkan nilai evaluasi tertinggi dengan menggunakan *10-Folds Cross Validation*. Menghasilkan nilai evaluasi *Accuracy* sebesar 68,52% dan *Kappa* 0,370.
- Algoritma *Naive Bayes* dengan Fitur seleksi *Backward Elimination* terbukti dapat meningkatkan hasil evaluasi pada prediksi waktu tunggu alumni mendapatkan pekerjaan.

5. DAFTAR PUSTAKA

- [1] C. D. Casuat dan E. D. Festijo, "Identifying the Most Predictive Attributes Among Employability Signals of Undergraduate Students", *16th IEEE International Colloquium on Signal Processing & its Applications*, 2020.
- [2] DIKTI. Tracer Study, "Tentang Tracer Study", 12 Februari 2021. [Online]. Available: <http://tracerstudy.kemdikbud.go.id/index.php/sitpout>.
- [3] Asroni, N. M. Ali dan S. Riyadi, "Perkiraan Masa Tunggu Alumni Mendapatkan Pekerjaan Menggunakan Metode Prediksi Data Mining Dengan Algoritma Naive Bayes Classifier", *Semesta Teknika*, pp. 189-197, 2014.
- [4] M. A. Maricar dan D. Pramana, "Perbandingan Akurasi Naive Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali", *Jurnal Sistem dan Informatika (JSI)*, pp. 16-22, 2019.
- [5] I. M. B. Adnyana, "Implementasi Naive Bayes Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan", *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, pp. 131-134, 2010.
- [6] R. Resmiati dan T. Arifin, "Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination", *SISTEMASI: Jurnal Sistem Informasi*, pp. 381-393, 2021.
- [7] D. M. A. North, *Data Mining for the Masses*, Second Edition, 2016.
- [8] D. T. Larose dan C. D. Larose, *Discovering Knowledge in Data an Introduction to Data Mining 2nd Edition*, Hoboken, New Jersey: John Wiley & Sons, Inc, 2014.
- [9] A. G. Lazuardy dan H. Setiaji, "Data Cleansing pada Data Rumah Sakit", *Proceeding SINTAK*, pp. 242-248, 2019.
- [10] H. A. Chusna dan A. T. Rumiati, "Penerapan Metode K-Means dan Fuzzy C-Means untuk Pengelompokan Sekolah Menengah Pertama (SMP) di Indonesia Berdasarkan Standar Nasional Pendidikan (SNP)", *JURNAL SAINS DAN SENI ITS*, pp. 216-223, 2020.
- [11] I. L. Uruilal, C. J. Supit dan T. Jansen, "Prediksi Banjir di Sungai Ranowanko Kecamatan Amurang Kabupaten Minahasa Selatan", *Jurnal Sipil Statik*, pp. 167-174, 2020.
- [12] E. Ardiningtyas dan P. H. Prima Rosa, "ANALISIS BALANCING DATA UNTUK MENINGKATKAN AKURASI DALAM KLASIFIKASI", *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, pp. 24-28, 2011.
- [13] M. Sadikin, R. Rosnelly, Roslina, T. S. Gunawan dan Wanayumini, "Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier dan C4.5", *Jurnal Media Informatika Budidarma*, pp. 1100-1009, 2020.
- [14] D. Alita, I. Sari, A. R. Isnain dan Styawati, "Penerapan Naive Bayes Classifier untuk Pendukung Keputusan Penerimaan Beasiswa", *JMSI*, vol. II, no. 1, pp. 17-23, 2021.
- [15] M. F. Aziz, S. Defiyanti dan B. N. Sari, "Perbandingan Algoritma Cart dan K-Nearest Neighbor untuk Prediksi Luas Lahan Panen Tanaman Padi di Kabupaten Karawang", *Jurnal TAM (Technology Acceptance Model)*, pp. 74-78, 2018.
- [16] M. R. Fanani, "Algoritma Naive Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa", *Jurnal DISPROTEK*, pp. 13-22, 2020.
- [17] S. P. Nabila, N. Ulinnuha dan A. Yusuf, "Model Prediksi Kelulusan Tepat Waktu dengan Metodefuzzy C-Means dan K-Nearest Neighbors Menggunakan Data Registrasi Mahasiswa", *Jurnal Ilmiah NERO*, pp. 39-47, 2021.
- [18] D. Sartika dan D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian", *Jatiji*, pp. 151-161, 2017.

ORIGINALITY REPORT

23%
SIMILARITY INDEX

20%
INTERNET SOURCES

9%
PUBLICATIONS

9%
STUDENT PAPERS

PRIMARY SOURCES

1 ejurnal.stmik-budidarma.ac.id 1 %
Internet Source

2 Submitted to Universitas Brawijaya 1 %
Student Paper

3 Ahmad Fauzi, Fanny Fatma Wati, Indah Sulistyowati, Muhammad Faitsuallah Akbar, Eka Rahmawati, Ratna Kurnia Sari. "Penerapan Metode Machine Learning Dalam Memprediksi Keberhasilan Panggilan Telemarketing Menjual Produk Bank", Indonesian Journal on Software Engineering (IJSE), 2020 1 %
Publication

4 jsi.stikom-bali.ac.id 1 %
Internet Source

5 Submitted to Universitas Dian Nuswantoro 1 %
Student Paper

6 journal.uad.ac.id 1 %
Internet Source

7	Luthfia Rahman, Noor Akhmad Setiawan, Adhistya Erna Permanasari. "Feature selection methods in improving accuracy of classifying students' academic performance", 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017	1 %
Publication		
8	jurnal.untan.ac.id	1 %
Internet Source		
9	Ruhmi Sulaehani. "PREDIKSI KEPUTUSAN KLIEN TELEMARKETING UNTUK DEPOSITO PADA BANK MENGGUNAKAN ALGORITMA NAIVE BAYES BERBASIS BACKWARD ELIMINATION", ILKOM Jurnal Ilmiah, 2016	1 %
Publication		
10	sistemasi.ftik.unisi.ac.id	1 %
Internet Source		
11	www.ejurnal.stmik-budidarma.ac.id	1 %
Internet Source		
12	Selcuk Ogutcu. "Student Performance Score Prediction Using Artificial Neural Network with the Support of Exploratory Factor Analysis and Clustering", 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP), 2020	1 %
Publication		

13	www.scribd.com Internet Source	1 %
14	ejournal.unisnu.ac.id Internet Source	1 %
15	ejurnal.its.ac.id Internet Source	1 %
16	prosiding.seminar-id.com Internet Source	1 %
17	jurnal-stmik.muralinggau.ac.id Internet Source	1 %
18	Submitted to Dominion High School Student Paper	1 %
19	Submitted to University of Rijeka Student Paper	1 %
20	id.scribd.com Internet Source	1 %
21	journal.umy.ac.id Internet Source	<1 %
22	hes-gotappointment-newspaper.icu Internet Source	<1 %
23	nero.trunojoyo.ac.id Internet Source	<1 %
24	repository.mercubuana.ac.id Internet Source	<1 %

25 ejournal.unsrat.ac.id Internet Source <1 %

26 norma.ncirl.ie Internet Source <1 %

27 text-id.123dok.com Internet Source <1 %

28 Submitted to Universitas Sebelas Maret Student Paper <1 %

29 journal.thamrin.ac.id Internet Source <1 %

30 123dok.com Internet Source <1 %

31 Febri Hadi, Dodi Guswandi. "Penentuan Penerimaan Mahasiswa Baru Pascasarjana Menggunakan Simple Additive Weighting (SAW)", Indonesian Journal of Computer Science, 2019 Publication <1 %

32 Submitted to STMIK STIKOM Bali Student Paper <1 %

33 docplayer.info Internet Source <1 %

34 jurnal.ideaspublishing.co.id Internet Source <1 %

jurnal.untad.ac.id

35

Internet Source

<1 %

36

nuhrison.blogspot.com

Internet Source

<1 %

37

repository.its.ac.id

Internet Source

<1 %

38

Desi Ratna Sari, Dedy Hartama, Irfan Sudahri Damanik, Anjar Wanto. "Penerapan Metode Naive Bayes dalam Memprediksi Kepuasan Mahasiswa Terhadap Cara Pengajaran Dosen", Prosiding Seminar Nasional Riset Information Science (SENARIS), 2019

Publication

<1 %

39

ejournal.unkhair.ac.id

Internet Source

<1 %

40

es.scribd.com

Internet Source

<1 %

41

jurnal.ubl.ac.id

Internet Source

<1 %

42

repository.unpas.ac.id

Internet Source

<1 %

43

core.ac.uk

Internet Source

<1 %

44

ejournal.akprind.ac.id

Internet Source

<1 %

45 ojs.cahayasurya.ac.id

Internet Source

<1 %

46 Perani Rosyani, Oke Hariansyah. "Pengenalan Citra Bunga Menggunakan Segmentasi Otsu Treshold dan Naïve Bayes", Jurnal Sistem dan Informatika (JSI), 2020

Publication

<1 %

47 doku.pub

Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off