

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

**Abimanyu Widhiantoyo<sup>1</sup>, Betha N Sari<sup>2</sup>, Dadang Yusup<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang  
Email: <sup>1</sup>abimanyu.widhiantoyo17032@student.unsika.ac.id, <sup>2</sup>betha.nurina@staff.unsika.ac.id,  
<sup>3</sup>dadang.dyf@staff.unsika.ac.id

(Naskah masuk: 1 Juli 2021, diterima untuk diterbitkan: 13 Agustus 2021)

### **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* meraih 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:** *Naïve Bayes, Backward Elimination, CRISP-DM, Waktu Tunggu Kerja Alumni*

## **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 the *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* 0.370. 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:** *Naïve Bayes, Backward Elimination CRISP-DM, Alumni Employment Waiting Time*

### **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. Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) menjadikan *Tracer Study* sebagai salah satu penilaian dan juga dijadikan sebagai salah satu syarat kelengkapan akreditasi [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 sebelumnya dengan algoritma *Naive Bayes* telah dilakukan 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 tentang 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*.

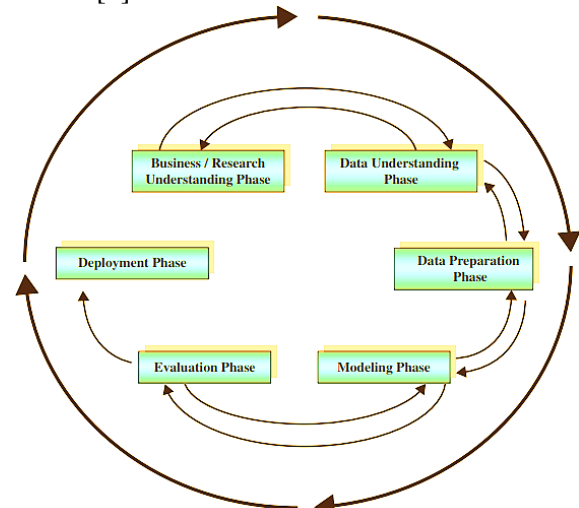
Penelitian ini menerapkan fitur seleksi *Backward Elimination* pada algoritma *Naive Bayes*. *Backward Elimination* merupakan salah satu metode seleksi fitur yang dapat mengoptimalkan akurasi model. Metode ini bekerja dengan cara pemilihan mundur, kemudian melakukan pengujian terhadap seluruh atribut-atribut 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. METODE 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:

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]. Hasil yang keliru seringkali ditemui pada banyak penelitian dengan analisis *data imbalanced*. Untuk melanjutkan proses analisis data, penanganan khusus perlu dilakukan [13].

Transformasi dilakukan pada atribut Kota Tempat Bekerja, Lama Studi, dan Lama Waktu Mendapatkan Pekerjaan.

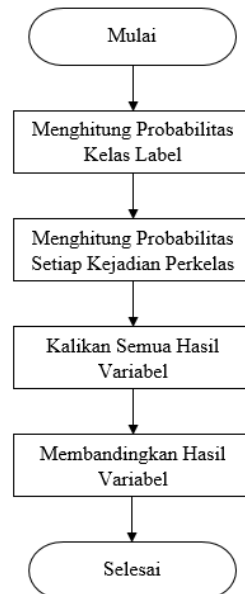
Format data akhir yang telah siap digunakan untuk proses pemodelan yaitu sebanyak 270 *record* data dengan 12 atribut utama dan 1 atribut label.

### 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 [14]:

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

Tahapan proses *Naïve Bayes* yaitu pada Gambar 2 berikut [15]:

Gambar 2. Proses *Naïve Bayes*

### 2.5. Evaluation (Evaluasi)

Model terbaik nantinya akan dipilih berdasarkan hasil evaluasi yang diperoleh. *K-Fold Cross Validation* dapat digunakan untuk memperoleh perkiraan yang tidak bias dari kinerja model pada jumlah data yang terbatas [16]. Pada proses klasifikasinya, masing-masing bagian akan membagi data secara acak ke dalam k bagian dengan ukuran yang sama. Penggunaan nilai k = 10 atau *10-Folds Cross Validation* secara umum sangat dianjurkan karena merupakan jumlah *fold* terbaik untuk uji validitas [17]. *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 [18].

## 2.6. Deployment (Penyebaran)

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

## 3. HASIL DAN PEMBAHASAN

Pemodelan terhadap *dataset* dilakukan menggunakan algoritma *Naive Bayes* dengan bantuan tools *RapidMiner Studio*. *RapidMiner Studio* dapat membantu aktivitas analisis prediktif dengan proses kerja dan penggunaan yang mudah. Begitu juga dengan proses *data mining* seperti data *cleansing*, *filtering*, *clustering*, dll yang mudah digunakan [19]. Tahapan proses *Naive Bayes* 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:

Tabel 2. Probabilitas Kelas Label

Kelas	Jumlah	Probabilitas
		Kelas P(H)
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:

Tabel 3. Probabilitas Jenis Kelamin

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:

Tabel 4. Probabilitas Angkatan

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:

Tabel 5. Probabilitas Tahun Lulus

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:

Tabel 6. Probabilitas IPK

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:

Tabel 7. Probabilitas Lama Studi

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:

Tabel 8. Probabilitas Riwayat Organisasi/Kepanitiaan

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:

Tabel 9. Probabilitas Pengalaman Magang/Kerja

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:

Tabel 10. Probabilitas Memiliki Sertifikasi Keahlian

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(X H))	
	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(X H))	
	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(X H))	
	CEPAT	LAMBAT
Karawang Luar	0,718	0,500
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(X H))	
	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

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

#### 2. Kelas LAMBAT

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

#### 4. Membandingkan Hasil Variabel

Nilai probabilitas terbesar dari hasil perhitungan pada kelas label akan dijadikan sebagai masukan data uji tersebut [20]. Dari perhitungan manual sebelumnya dengan mengkalikan 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 *Naïve Bayes*

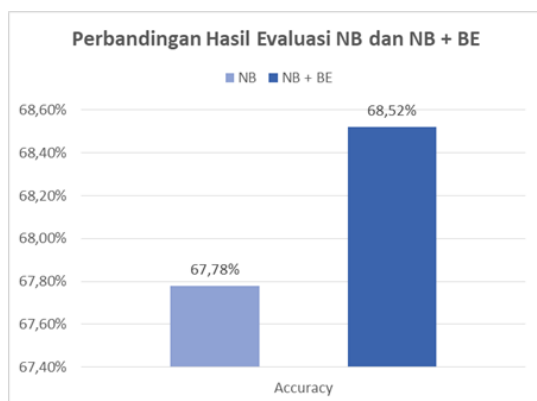
Skenario	k-folds	Accuracy	Kappa
1	2	64,81%	0,293
2	3	64,81%	0,294
3	4	64,07%	0,282
4	5	64,81%	0,296
5	6	67,41%	0,346
6	7	65,19%	0,302
7	8	65,93%	0,315
8	9	67,41%	0,346
9	10	67,78%	0,356

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 *Naïve 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

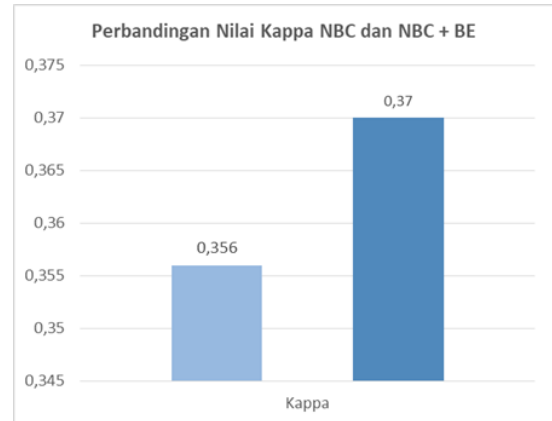
Evaluasi dengan *10-Folds Cross Validation* menggunakan nilai *Accuracy* menghasilkan model *Naïve Bayes* dengan fitur seleksi *Backward Elimination* mampu meningkatkan kinerja pemodelan berdasarkan nilai *Accuracy*. Dapat dilihat perbandingannya pada gambar 3 berikut ini.



Gambar 3. Perbandingan Hasil Evaluasi

Evaluasi dengan *10-Folds Cross Validation* menggunakan nilai *Kappa* menghasilkan model

*Naïve 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	1

### 4. KESIMPULAN

Dari penelitian yang telah dilakukan dapat disimpulkan bahwa:

1. Hasil dari semua pengujian yang telah dilakukan dengan 9 skenario uji menggunakan *K-Fold Cross Validation* diketahui bahwa pada pemodelan *Naïve Bayes* dihasilkan performa terbaik dengan *10-Folds Cross Validation*. Pengujian menghasilkan nilai evaluasi *Accuracy* sebesar 67,78% dan *Kappa* 0,356.
2. Pada pemodelan *Naïve 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.

3. 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. 2020. "Identifying the Most Predictive Attributes Among Employability Signals of Undergraduate Students", *16th IEEE International Colloquium on Signal Processing & its Applications*.
- [2] DIKTI. Tracer Study. 2021. "Tentang Tracer Study", Tersedia [<http://tracerstudy.kemdikbud.go.id/index.php/site/about>]. diakses 12 Februari 2021.
- [3] Asroni, N. M. Ali dan S. Riyadi. 2018. "Perkiraan Masa Tunggu Alumni Mendapatkan Pekerjaan Menggunakan Metode Prediksi Data Mining Dengan Algoritma Naive Bayes Classifier", *Semesta Teknika*, Vol. 21(2), pp. 189-197.
- [4] M. A. Maricar dan D. Pramana. 2019. "Perbandingan Akurasi Naive Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali", *Jurnal Sistem dan Informatika (JSI)*, Vol. 14(1), pp. 16-22.
- [5] I. M. B. Adnyana. 2020. "Implementasi Naive Bayes Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan", *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, pp. 131-134.
- [6] R. Resmiati dan T. Arifin. 2021. "Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination", *SISTEMASI: Jurnal Sistem Informasi*, Vol. 10(2), pp. 381-393.
- [7] D. M. A. North. 2016. *Data Mining for the Masses*, Second Edition.
- [8] D. T. Larose dan C. D. Larose. 2014. *Discovering Knowledge in Data an Introduction to Data Mining 2nd Edition*, Hoboken, New Jersey: John Wiley & Sons, Inc.
- [9] A. G. Lazuardy dan H. Setiaji. 2019. "Data Cleansing pada Data Rumah Sakit", *Proceeding SINTAK*, Vol.3, pp. 242-248.
- [10] H. A. Chusna dan A. T. Rumiati. 2020. "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*, Vol. 9(2), pp. 216-223.
- [11] I. L. Uruilal, C. J. Supit dan T. Jansen. 2020. "Prediksi Banjir di Sungai Ranowanko Kecamatan Amurang Kabupaten Minahasa Selatan", *Jurnal Sipil Statik*, Vol. 8(2), pp. 167-174.
- [12] Y. E. Ardiningtyas dan P. H. Prima Rosa. 2021. "Analisis Balancing Data untuk Meningkatkan Akurasi dalam Klasifikasi", *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, pp. 24-28.
- [13] A. Indrawati. 2021. "Penerapan Teknik Kombinasi Oversampling dan Undersampling untuk Mengatasi Permasalahan Imbalanced Dataset", *JIKO (Jurnal Informatika dan Komputer)*, Vol. 4(1), pp. 38-43.
- [14] M. Sadikin, R. Rosnelly, Roslina, T. S. Gunawan dan Wanayumini. 2020. "Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier dan C4.5", *Jurnal Media Informatika Budidarma*, Vol. 4(4), pp. 1100-1009.
- [15] D. Alita, I. Sari, A. R. Isnain dan Styawati. 2021. "Penerapan Naive Bayes Classifier untuk Pendukung Keputusan Penerima Beasiswa", *JDMSI*, Vol. 2(1), pp. 17-23.
- [16] M. F. Aziz, S. Defiyanti dan B. N. Sari. 2018. "Perbandingan Algoritma Cart dan K-Nearest Neighbor untuk Prediksi Luas Lahan Panen Tanaman Padi di Kabupaten Karawang", *Jurnal TAM (Technology Acceptance Model)*, Vol. 9(2), pp. 74-78.
- [17] M. R. Fanani. 2020. "Algoritma Naive Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa", *Jurnal DISPROTEK*, Vol. 11(1), pp. 13-22.
- [18] S. P. Nabila, N. Ulinuha dan A. Yusuf. 2021. "Model Prediksi Kelulusan Tepat Waktu dengan Metodefuzzy C-Means dan K-Nearest Neighbors Menggunakan Data Registrasi Mahasiswa", *Jurnal Ilmiah NERO*, Vol. 6(1), pp. 39-47.
- [19] V. Pynam, R. R. Spanadna and K. Srikanth. 2018. "An Extensive Study of Data Analysis Tools (Rapid Miner, Weka, R Tool, Knime, Orange)", *SSRG International Journal of Computer Science and Engineering*, Vol. 5(9) pp. 4-11.
- [20] D. Sartika dan D. I. Sensuse. 2017. "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian", *Jatiji*, Vol. 1(2), pp. 151-161.