



KLASIFIKASI MASA TUNGGU KERJA LULUSAN S1 MATEMATIKA UNAND MENGGUNAKAN *NAIVE BAYES* *CLASSIFIER*

YULMIATI^{1*}, IZZATI RAHMI H.G², FERRA YANUAR³

^{1,2,3}Jurusan Matematika, Fakultas MIPA, Universitas Andalas

*ayulunnie@gmail.com

ABSTRAK

Prodi S1 Matematika adalah salah satu prodi yang berakreditasi A di Universitas Andalas (Unand). Sebagian dari lulusan S1 Matematika Unand telah bekerja di instansi pemerintahan maupun swasta. Lulusan tersebut tentunya memiliki masa tunggu kerja yang berbeda-beda dan dipengaruhi oleh beberapa faktor. Faktor-faktor yang mempengaruhi masa tunggu kerja antara lain yaitu jenis kelamin, usia, masa studi, upah, keaktifan berorganisasi, IPK dan sebagainya. Oleh karena itu, penelitian ini melakukan klasifikasi masa tunggu kerja lulusan S1 Matematika Unand berdasarkan atribut yang terdiri dari faktor-faktor yang diduga mempengaruhi masa tunggu kerja lulusan S1 Matematika Unand. Metode klasifikasi yang digunakan adalah *Naive Bayes Classifier*. Dalam proses klasifikasi tidak semua atribut yang efektif dan relevan, sehingga perlu dilakukan teknik pemilihan atribut menggunakan *Information Gain* (IG). Berdasarkan nilai akurasi diketahui bahwa klasifikasi *Naive Bayes* menggunakan lima belas atribut adalah 0,8839 dan klasifikasi *Naive Bayes* menggunakan delapan atribut terpilih dari IG adalah 0,901786. Atribut-atribut yang terpilih dari IG adalah keaktifan berorganisasi, sumber utama biaya kuliah, kemampuan bersaing dengan lulusan perguruan tinggi lain, masa studi, kompetensi yang dikuasai saat lulus, IPK, jenis kelamin, dan melanjutkan studi.

Kata Kunci: masa tunggu kerja, klasifikasi, *Naive Bayes Classifier*, *Information Gain*

ABSTRACT

The S1 Mathematics study program is one of the A accredited study programs at Andalas University (Unand). Most of the Unand Mathematics S1 graduates have worked in government and private agencies. Certainly, the graduates have different waiting period to get the first job. The waiting period to get the first job is influenced by several factors. Factors that influence the waiting period to get the first job include gender, age, study period, wages, organizational activity, cumulative achievement index and so on. Therefore, this study classifies the waiting time to get the first job of Unand Mathematics S1 graduates based on attributes consisting of factors that are thought to affect the waiting time of Unand Mathematics S1 graduates to get the first job. The classification method used is Naive Bayes Classifier (NBC). In the classification process, not all attributes are effective and relevant, so it is necessary to do attribute selection techniques using Information Gain (IG). Based on the accuracy value it is known that the Naive Bayes classification using fifteen attributes is 0.8839 and the Naive Bayes classification using eight selected attributes from IG is 0.901786. The attributes selected from the IG are organizational activity, main sources of tuition fees, ability to compete with other tertiary graduates, study periods, competencies mastered upon graduation, cumulative achievement index, gender, and continuing studies.

Keywords: waiting time to get a job, classification, *Naive Bayes Classifier*, *Information Gain*

1 Pendahuluan

Salah satu indikator penilaian akreditasi perguruan tinggi dan program studi adalah persentase lulusan yang langsung bekerja. Indikator ini digunakan untuk mengukur lulusan yang memperoleh pekerjaan dengan masa tunggu kerja kurang dari 12 bulan. Dalam hal ini, masa tunggu kerja merupakan waktu yang dibutuhkan seseorang untuk mendapatkan pekerjaan pertama yang dimulai saat telah dinyatakan lulus [1].

Universitas Andalas adalah salah satu perguruan tinggi yang memiliki peringkat akreditasi A di Indonesia. Selain itu, sebagian besar Program Studi (Prodi) S1 Unand telah berakreditasi A. Salah satu prodi tersebut adalah Prodi S1 Matematika. Sampai akhir tahun 2019 terdapat lebih dari 600 orang lulusan S1 yang telah dihasilkan oleh Jurusan Matematika Unand. Sebagian lulusan S1 Matematika Unand telah bekerja di instansi pemerintahan maupun swasta. Lulusan tersebut tentunya memiliki masa tunggu kerja yang berbeda-beda.

Masa tunggu kerja dipengaruhi oleh beberapa faktor. Faktor-faktor yang mempengaruhi masa tunggu kerja antara lain adalah usia, pendidikan, pengalaman, status perkawinan dan upah [2]. Menurut Anjani, variabel yang mempengaruhi masa tunggu kerja adalah prestasi belajar, masa studi dan keaktifan berorganisasi [1]. Berbeda halnya dengan penelitian Putranto diketahui bahwa variabel yang mempengaruhi masa tunggu kerja adalah jumlah anggota keluarga, status SLTA, jenis pekerjaan yang dipilih dan motivasi dalam mencari kerja [3].

Penelitian ini melakukan klasifikasi masa tunggu kerja lulusan S1 Matematika Unand berdasarkan atribut yang terdiri dari faktor-faktor yang diduga mempengaruhi masa tunggu kerja. Metode yang digunakan adalah *Naive Bayes Classifier*. Metode *Naive Bayes* adalah metode pengklasifikasian peluang yang berakar pada kaidah Bayes untuk menghitung sekumpulan nilai peluang dengan menjumlahkan frekuensi dan kombinasi nilai dataset yang diberikan. Metode ini merupakan salah satu metode klasifikasi sederhana dan memiliki akurasi yang tinggi. Berdasarkan penelitian Handayani diketahui bahwa metode *Naive Bayes Classifier* mengklasifikasikan teks otomatis pengaduan dan pelaporan masyarakat dengan rata-rata akurasi sebesar 92,67% [4]. Selain itu pada penelitian Syarli diketahui bahwa akurasi *Naive Bayes* dalam mengklasifikasikan kelulusan mahasiswa baru di perguruan tinggi adalah 94% [5].

Pada saat melakukan teknik klasifikasi, seringkali terdapat banyak atribut yang digunakan. Dari atribut-atribut yang digunakan terdapat kemungkinan tidak semua atribut yang efektif dan relevan pada klasifikasi, terutama pada klasifikasi *Naive Bayes*. Oleh karena itu, pada penelitian ini akan digunakan metode *Naive Bayes Classifier* dalam dua tahap. Pada tahap pertama akan dilakukan klasifikasi *Naive Bayes* pada masa tunggu kerja lulusan S1 Matematika Unand dengan semua atribut yang digunakan pada penelitian. Kemudian, pada tahap kedua akan dilakukan lagi klasifikasi *Naive Bayes* dengan menggunakan atribut yang efektif dan relevan dalam klasifikasi yang dipilih melalui teknik pemilihan atribut *Information Gain*.

2 Tinjauan Pustaka

2.1 Metode *Naive Bayes Classifier*

Misalkan Y adalah suatu variabel kategori yang menyatakan *class* dan y adalah nilai dari Y . Kemudian, misalkan X_1, X_2, \dots, X_n adalah atribut-atribut pada data yang terdiri dari n atribut [6]. Berdasarkan kaidah Bayes diperoleh persamaan.

$$P(Y = y | X_1 = x_1, \dots, X_n = x_n) = \frac{P(Y = y) P(X_1 = x_1, \dots, X_n = x_n | Y = y)}{P(X_1 = x_1, \dots, X_n = x_n)} \quad (1)$$

Atribut $X_k, k = 1, 2, \dots, n$ diasumsikan saling bebas, maka berdasarkan persamaan (1) model peluang *Naive Bayes Classifier* adalah

$$P(Y = y|X_1 = x_1, \dots, X_n = x_n) \propto P(Y = y) \prod_{k=1}^n P(X_k = x_k|Y = y) \quad (2)$$

Untuk mendapatkan nilai peluang posterior diperlukan nilai peluang prior dan peluang bersyarat pada atribut X_k . Nilai peluang tersebut diperoleh dengan menduga peluang kejadian setiap variabel [7]. Untuk menduga peluang prior dilakukan dengan menghitung peluang kejadian Y yang dinyatakan sebagai berikut

$$P(Y = y) = \frac{\text{Banyak pengamatan di class } y}{\text{Total pengamatan}}$$

Kemudian untuk menduga peluang bersyarat pada atribut X_k dilakukan dengan menghitung peluang kejadian bersyarat pada masing-masing atribut X_k . Untuk menduga nilai peluang bersyarat *Naive Bayes* dengan atribut kontinu digunakan model peluang untuk sebaran normal. Karena nilai atribut numerik pada setiap *class* diasumsikan terdistribusi berdasarkan sebaran normal [8]. Kemudian, untuk menduga nilai peluang bersyarat *Naive Bayes* dengan atribut diskret dilakukan dengan menghitung peluang kejadian bersyarat untuk atribut X_k .

Misal x_k menyatakan nilai dari atribut X_k , maka nilai peluang bersyarat *Naive Bayes* untuk X_k bisa dihitung menggunakan rumus [9]

$$P(X_k = x_k|Y = y) = \frac{N(x_k, y)}{N(y)} \quad (3)$$

dimana

$N(x_k, y)$: Banyak pengamatan atribut X_k yang bernilai x_k pada *class* y

$N(y)$: Banyak pengamatan pada *class* y .

Berdasarkan penelitian J. Wu, terdapat kategori di atribut tertentu tidak terdapat pada suatu *class* yang menyebabkan peluang bersyarat dari atribut tersebut bernilai nol, sehingga pengklasifikasian *Naive Bayes* tidak bisa dilakukan [7]. Untuk mengatasi masalah ini digunakan *laplace smoothing*, sehingga persamaan (3) bisa dinyatakan sebagai berikut [7][9]

$$P(X_k = x_k|Y = y) = \frac{N(x_k, y) + 1}{N(y) + p} \quad (4)$$

dengan p menyatakan banyak nilai pada atribut X_k

Klasifikasi *Naive Bayes* yang dihasilkan mengikuti aturan keputusan MAP (*Maximum A Posteriori*) yang didefinisikan sebagai berikut [9]

$$\hat{Y} = \arg \max_{y=1, \dots, m} P(Y = y|X_1 = x_1, \dots, X_n = x_n) \quad (5)$$

2.2 Pemilihan Atribut

Salah satu teknik pemilihan atribut adalah Informasi Gain (IG). Informasi Gain adalah teknik pemilihan atribut yang paling sederhana dengan melakukan perbandingan atribut dan sering digunakan untuk dataset yang terdiri dari variabel-variabel kategori. IG berguna untuk mendeteksi atribut yang paling banyak memiliki informasi pada klasifikasi [10].

Misalkan D adalah himpunan data yang terdiri dari c *class*, Informasi *Gain* bisa dihitung dengan langkah-langkah sebagai berikut :

1. Menghitung ukuran ketidakpastian (*entropy*) pada *class* dengan rumus sebagai berikut [11]

$$info(D) = - \sum_{i=1}^c p_i \log_2(p_i) \quad (6)$$

dimana p_i adalah nilai peluang untuk *class* ke- i , c adalah jumlah *class* pada data dan untuk $p_i = 0$ maka $0 \log_2 0 = 0$ [12].

2. Menghitung *entropy* pada masing-masing atribut di D menggunakan rumus berikut [11]

$$info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times info(D_j) \quad (7)$$

dimana $info(D_j)$ adalah *entropy* pada kategori j , $|D_j|$ adalah jumlah pengamatan untuk kategori ke- j , $|D|$ adalah jumlah total pengamatan, A adalah atribut, dan v adalah banyak kategori pada A .

3. Menghitung *Information Gain* dengan rumus berikut [4]

$$Gain(A) = info(D) - info_A(D) \quad (8)$$

Atribut yang terpilih adalah atribut yang memiliki nilai gain yang tidak terlalu besar atau terlalu kecil. Oleh karena itu, atribut yang terpilih memiliki nilai gain yang lebih besar dari rata-rata nilai gain seluruh atribut [10].

2.3 Evaluasi Klasifikasi

Untuk melakukan evaluasi pada klasifikasi *Naive Bayes* digunakan matriks konfusi[3]. Matriks konfusi ini disajikan pada Tabel 1.

Tabel 1: Matriks Konfusi untuk j class

Class Asli (i)	Class dugaan (j)			
	Class 1	Class 2	...	Class j
Class 1	n_{11}	n_{12}	...	n_{1j}
Class 2	n_{21}	n_{22}	...	n_{2j}
\vdots	\vdots	\vdots	\vdots	\vdots
Class i	n_{i1}	n_{i2}	...	n_{ij}

Pada Tabel 1 diketahui bahwa n_{ij} menyatakan banyak pengamatan dari kelas asli i yang hasil klasifikasinya masuk ke kelas dugaan j . Dari matriks konfusi tersebut, bisa dihitung nilai akurasi dari klasifikasi. Untuk menghitung nilai akurasi tersebut digunakan rumus berikut

$$\text{Akurasi} = \frac{\text{jumlah pengamatan yang benar diklasifikasikan}}{\text{jumlah total pengamatan}}$$

3 Metode Penelitian

3.1 Data

Data yang digunakan pada penelitian ini adalah data penelitian *tracer study* tahun 2019/2020 lulusan S1 Jurusan Matematika Unand yang melibatkan 112 responden lulusan S1 Matematika Unand yang telah bekerja. Dari 112 responden tersebut, terdapat 89% responden yang memiliki masa tunggu kerja kurang dari 12 bulan dan 11% responden memiliki masa tunggu kerja sama dengan atau lebih dari 12 bulan. Masa tunggu kerja (Y) dijadikan sebagai variabel respon (*class*) yang terdiri dari 2 kategori yaitu masa tunggu kerja < 12 bulan ($y = 0$) dan masa tunggu kerja ≥ 12 bulan ($y = 1$). Kemudian, setiap responden akan diklasifikasikan berdasarkan 15 variabel prediktor (atribut) menggunakan *Naive Bayes*. Atribut yang digunakan pada penelitian ini ditunjukkan oleh Tabel 2.

Tabel 2: Atribut yang digunakan dalam penelitian

No	Atribut	Kategori
1.	Jenis kelamin (X_1)	1. Perempuan
		2. Laki-laki
2.	Daerah asal (X_2)	1. Padang, Sumatera Barat
		2. Luar Padang, Sumatera Barat
		3. Luar Provinsi Sumatera Barat
3.	Jalur masuk Jurusan S1 Matematika Unand (X_3)	1. Undangan (SNMPTN)
		2. Ujian Tulis (SBMPTN)
		3. Mandiri
4.	Sumber utama biaya saat kuliah (X_4)	1. Orang tua/keluarga
		2. Beasiswa
		3. Sebagian beasiswa
		4. Biaya sendiri
5.	Keaktifan Berorganisasi (X_5)	1. Sangat aktif
		2. Aktif
		3. Cukup aktif
		4. Kurang aktif
		5. Tidak aktif
6.	Pengalaman akademik (X_6) *	1. Ada
		2. Tidak ada
7.	IPK saat lulus kuliah (X_7)	1. $\leq 2,75$
		2. $> 2,75$ sampai ≤ 3
		3. > 3 sampai $\leq 3,5$
		4. $> 3,5$ sampai 4
8.	Masa studi (X_8)	1. Tepat waktu (≤ 48 bulan)
		2. Tidak tepat waktu (> 48 bulan)
9.	Melanjutkan studi (X_9)	1. Ya
		2. Tidak
10.	Mengetahui cara membuat CV untuk melamar pekerjaan (X_{10})	1. Ya
		2. Tidak
11.	Waktu mulai mencari pekerjaan (X_{11})	1. Sebelum lulus
		2. Setelah lulus
12.	Proses mencari kerja (X_{12})	1. Aktif (mencari sendiri)
		2. Pasif (koneksi)
13.	Kesesuaian bidang studi dengan pekerjaan (X_{13})	1. Sangat sesuai
		2. Sesuai
		3. Kurang sesuai
		4. Tidak sesuai
14.	Kompetensi yang dikuasai saat baru lulus kuliah (X_{14}) **	1. Sangat menguasai
		2. Menguasai
		3. Kurang menguasai
15.	Kemampuan bersaing dengan lulusan dari Perguruan Tinggi lain (X_{15}) ***	1. Sangat mampu
		2. Mampu
		3. Kurang mampu

Keterangan :

* Pengalaman akademik saat kuliah antara lain menjadi asisten praktikum atau asisten mata kuliah, membantu penelitian dosen, mengikuti lomba penelitian ilmiah mahasiswa, mengikuti lomba/olimpiade bidang ilmu, dan student mobility/short course/seminar internasional.

**Kompetensi yang dikuasai saat baru lulus kuliah antara lain pengetahuan di bidang ilmu, pengetahuan di luar bidang ilmu, kemampuan bahasa inggris, keterampilan komputer, kerjasama tim, berpikir kritis, mampu bekerja dalam tekanan, kemampuan berkomunikasi, kemampuan adaptasi, dan keterampilan riset.

***Kemampuan bersaing dengan lulusan dari Perguruan Tinggi lain merupakan persepsi dari lulusan S1 Matematika Unand apakah mampu atau tidak untuk bersaing dengan lulusan PT lain dalam dunia kerja.

3.2 Langkah-langkah Analisis

Langkah-langkah analisis yang dilakukan pada penelitian ini adalah :

1. Mengklasifikasikan setiap pengamatan menggunakan metode *Naive Bayes* dengan langkah-langkah sebagai berikut :
 - Menghitung peluang $P(Y = y)$.
 - Menghitung nilai peluang bersyarat untuk masing-masing atribut.
 - Menghitung nilai posterior untuk semua pengamatan.
 - Menentukan dugaan class dengan *Maximum A Posteriori* (MAP).
2. Menghitung nilai akurasi pada klasifikasi yang telah dilakukan.
3. Memilih atribut dengan menggunakan Information Gain.
4. Mengklasifikasikan kembali setiap pengamatan menggunakan *Naive Bayes* dengan atribut-atribut terpilih.
5. Menghitung kembali nilai akurasi pada klasifikasi.

4 Hasil dan Pembahasan

Pada penelitian ini, klasifikasi *Naive Bayes* dilakukan dalam dua tahapan. Klasifikasi *Naive Bayes* tahap pertama dilakukan dengan mengklasifikasikan masa tunggu kerja menggunakan 15 atribut yang terdiri dari faktor-faktor yang diduga mempengaruhi masa tunggu kerja lulusan S1 Matematika Unand. Peluang prior atau peluang kejadian untuk *class* Y adalah

- $P(Y = \text{masa tunggu} < 12 \text{ bulan}) = P(Y = 1) = \frac{100}{112} = 0,8928$
- $P(Y = \text{masa tunggu} \geq 12 \text{ bulan}) = P(Y = 0) = \frac{12}{112} = 0,1071$.

Langkah selanjutnya adalah menentukan peluang bersyarat untuk setiap atribut. Peluang bersyarat untuk setiap atribut ditentukan dengan menggunakan penduga peluang *Naive Bayes* dengan *laplace smoothing*. Setelah nilai peluang bersyarat diperoleh, maka bisa dihitung nilai posterior untuk masing-masing *class* masa tunggu kerja pada setiap pengamatan. Salah satu data pengamatan yang digunakan ditunjukkan oleh Tabel 3.

Tabel 3: Data pengamatan dengan 15 atribut

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	Y
1	2	2	2	2	1	3	1	2	1	1	1	2	2	1	1

Cara perhitungan posterior untuk data pada Tabel 3 adalah

- $P(Y = 1|X_1 = 1, \dots, X_{15} = 1) \propto P(Y = 1) \times \prod_{k=1}^{15} P(X_k = x_k|Y = 1)$
 $\propto 0,8928 \times 1,5843019 \times 10^{-5} = 1,4146 \times 10^{-5}$

- $$P(Y = 0|X_1 = 1, \dots, X_{15} = 1) \propto P(Y = 0) \times \prod_{k=1}^{15} P(X_k = x_k|Y = 0)$$

$$\propto 0,1071 \times 2,242022 \times 10^{-6} = 2,4022 \times 10^{-7}.$$

Setelah diperoleh nilai peluang posterior, langkah selanjutnya adalah menentukan hasil klasifikasi *Naive Bayes* dengan MAP. Dari hasil posterior diperoleh bahwa posterior pada *class* $Y = 1$ lebih besar dari posterior pada *class* $Y = 0$, sehingga berdasarkan aturan MAP hasil klasifikasi *Naive Bayes* untuk pengamatan pada Tabel 3 adalah *class* $Y = 1$.

Selanjutnya dilakukan evaluasi klasifikasi *Naive Bayes* menggunakan matriks konfusi. Matriks konfusi dari klasifikasi *Naive Bayes* tahap pertama ditunjukkan oleh Tabel 4.

Tabel 4: Matriks konfusi *Naive Bayes* tahap pertama

Class Asli	Class Dugaan	
	Y = 1	Y = 0
Y = 1	97	3
Y = 0	10	2

Berdasarkan Tabel 4 diperoleh akurasi dari klasifikasi *Naive Bayes* tahap pertama dengan menggunakan 15 atribut adalah

$$\text{Akurasi} = \frac{\text{jumlah pengamatan yang benar diklasifikasikan}}{\text{jumlah total pengamatan}} = \frac{97 + 2}{96 + 10 + 4 + 2} = \frac{99}{112} = 0,8839.$$

Pada klasifikasi *Naive Bayes* tahap kedua, atribut yang digunakan adalah atribut yang telah terpilih dari *Information Gain* (IG). Nilai IG dari semua atribut ditunjukkan oleh Tabel 5.

Tabel 5: Nilai IG semua atribut

Atribut	Nilai IG	Keterangan
X_5	0,064854	Keaktifan berorganisasi
X_4	0,02809	Sumber utama biaya saat kuliah
X_{15}	0,024329	Kemampuan bersaing dengan lulusan PT lain
X_8	0,018391	Masa studi S1
X_{14}	0,017605	Kompetensi yang dikuasai saat baru lulus
X_7	0,017508	Nilai IPK saat lulus kuliah
X_1	0,016992	Jenis Kelamin
X_9	0,016429	Melanjutkan studi
X_{11}	0,010605	Waktu mulai mencari pekerjaan
X_{10}	0,003776	Pengetahuan cara membuat CV
X_3	0,003236	Jalur masuk Jurusan S1 Matematika
X_2	0,002798	Daerah asal
X_{12}	0,002078	Proses mencari pekerjaan
X_{13}	0,001993	Kesesuaian bidang studi dengan pekerjaan
X_6	0,00083	Pengalaman akademik

Berdasarkan Tabel 5, diperoleh rata-rata nilai IG dari semua atribut adalah

$$\begin{aligned} \text{Rata - rata} &= \frac{\text{Gain}(X_1) + \dots + \text{Gain}(X_{15})}{15} = \frac{0,016992 + \dots + 0,024329}{15} \\ &= 0,015282933. \end{aligned}$$

Selanjutnya, akan dipilih atribut yang memiliki nilai IG lebih besar dari rata-rata nilai IG semua atribut. Atribut-atribut yang terpilih dari IG ditunjukkan oleh Tabel 6.

Tabel 6: Atribut-atribut terpilih dari IG

Atribut	Keterangan
X ₅	Keaktifan berorganisasi
X ₄	Sumber utama biaya saat kuliah
X ₁₅	Kemampuan bersaing dengan lulusan PT lain
X ₈	Masa studi S1
X ₁₄	Kompetensi yang dikuasai saat baru lulus
X ₇	Nilai IPK saat lulus kuliah
X ₁	Jenis Kelamin
X ₉	Melanjutkan studi

Berdasarkan Tabel 6 diketahui bahwa terdapat 8 atribut yang terpilih dari IG. Atribut-atribut tersebut digunakan untuk mengklasifikasikan masa tunggu kerja lulusan S1 Matematika Unand menggunakan *Naive Bayes*. Evaluasi klasifikasi *Naive Bayes* tahap kedua ditunjukkan oleh Tabel 7.

Tabel 7: Matriks konfusi *Naive Bayes* tahap kedua

Class Asli	Class Dugaan	
	Y = 1	Y = 0
Y = 1	99	1
Y = 0	10	2

Berdasarkan Tabel 7 diperoleh akurasi dari klasifikasi *Naive Bayes* tahap kedua yang menggunakan 8 atribut terpilih dari IG adalah 0,901786.

5 Kesimpulan

Berdasarkan klasifikasi masa tunggu kerja lulusan S1 Matematika yang telah dilakukan, diperoleh klasifikasi *Naive Bayes* tahap pertama yang menggunakan 15 atribut memiliki nilai akurasi sebesar 0,8839 dan klasifikasi *Naive Bayes* tahap kedua yang menggunakan 8 atribut terpilih dari *Information Gain* memiliki akurasi sebesar 0,901786. Atribut-atribut yang terpilih dari IG adalah keaktifan berorganisasi, sumber utama biaya kuliah, kemampuan bersaing dengan lulusan perguruan tinggi lain, masa studi, kompetensi yang dikuasai saat lulus, IPK saat lulus, jenis kelamin dan melanjutkan studi.

Daftar Pustaka

- [1] N. S. Anjani, "Pengaruh Prestasi Belajar, Masa Studi dan Keaktifan Berorganisasi Terhadap Masa Tunggu dan Relevansi Pekerjaan," *Jurnal Peendidikan dan Ekonomi*, pp. 554-565, 2018.
- [2] M. Astuti, "Faktor-faktor yang Mempengaruhi Lama Menganggur Bagi Pekerja di Industri Perkebunan Kelapa Sawit Ogan Komering Ulu," *Jurnal Ilmiah STIE MDP*, pp. 130-149, 2013.

- [3] R. T. Putranto and M. Mashuri, “Analisis Statistika Tentang Faktor-Faktor yang Mempengaruhi Waktu Tunggu Kerja Fresh Graduate di Jurusan Statistika Institut Teknologi Sepuluh Nopember (ITS),” *Jurnal Sains dan Seni ITS* , pp. 324-328, 2012.
- [4] F. Handayani and F. S. Pribadi, “Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110,” *Jurnal Teknik Elektro*, pp. 19-24, 2015.
- [5] S. and A. A. Muin, “Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi),” *Jurnal Ilmiah Ilmu Komputer* , pp. 22-26, 2016.
- [6] D. Berrar, “Bayes' Theorem and Naive Bayes Classifier,” *Encyclopedia of Bioinformatics and Computational Biology*, pp. 403-412, 2018.
- [7] J. Wu, Z. Cai and X. Zhu, “Self-Adaptive Probability Estimation for Naive Bayes Classification,” in *Proceedings of International Joint Conference on Neural Networks*, Texas, 2013.
- [8] G. H. John and P. Langley, “Estimating Countinous Distribution in Bayesian Classifiers,” in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, 1995.
- [9] S. Raschka, “arXiv,” 14 Februari 2017. [Online]. Available: <https://arxiv.org/abs/1410.5329>.
- [10] L. R. Putri, “Klasifikasi Sentimen Ulasan Buku Berbahasa Inggris Menggunakan Information Gain dan Naive Bayes,” in *e-Proceeding of Engineering*, 2017.
- [11] J. Han, M. Kamber and J. Pei, *Data Mining : Concept and Techniques 3rt Edition*, Elsevier: Morgan Kaufmann, 2006.
- [12] R. M. Gray, *Entropy and Information Theory Second Edition*, New York: Springer, 2010.