SEMESTA TEKNIKA
Vol. 21, No. 2, 189-197, November 2018
DOI: 10.18196/st.212225

Perkiraan Masa Tunggu Alumni Mendapatkan Pekerjaan Menggunakan Metode Prediksi Data Mining Dengan Algoritma Naive Bayes *Classifier*

(The Estimated Wait Times Alumni are Given Positions Using The Data Prediction Method Used With A Bayes-Naïve Algorithm Classifier)

ASRONI, NADIYAH MAHARTY ALI, SLAMET RIYADI

ABSTRACT

Student and Alumni data Universitas Muhammadiyah Yogyakarta is very common, and one of these is the alumni data obtained from work after the completion of undergraduate studies. Former students are given jobs caused or influenced by a range of factors. This research aims to have the grace period Classification or old alumni gain positions by triggering a process of data extraction and using the Bayes naïve classification algorithm. The algorithms used later succeeded in predicting sooner or later to get a job, the predictive results alumni can be used to make decisions to improve the quality of a university. Research on the support system using several parameters, i.e., gender, faculty, GPA, year of graduation, and job status. The data used are as much as 435, including seven years of 2011-2014 volume. The results of this study have the accuracy level of former students having the grace period come to 71% and of the calculated results of the predictions of the former students obtaining a job at Universitas Muhammadiyah Yogyakarta of the year 2011-2014 the Ensure that the work is carried out more quickly with the status of the slow to deliver the work.

Keywords: Forecasting the grace period getting the job, data mining, Naïve Bayes Classifier, RapidMiner

PENDAHULUAN

Sebuah Perguruan Tinggi (PT) pasti memiliki data alumni seperti nama mahasiswa, program studi yang diambil, jenis kelamin, serta data mahasiswa yang sudah lulus dan yang sudah mendapatkan pekerjaan. Ketika data tidak di klasifikasikan dalam kelompok atau aturan tertentu akan dapat menimbulkan masalah lain yaitu memperlambat proses pencarian data. Data yang sudah dibuat pengklasifikasian dapat memberikan informasi yang banyak apabila di analisis lebih dalam.

Alumni sangat berperan penting dalam peningkatan kualitas yang telah dicapai oleh sebuah PT, karena semakin cepat alumni mendapatkan pekerjaan itu artinya sistem belajar mengajar pada sebuah universitas sudah cukup baik. Dari sinilah penelitian ini memerlukan data mining, Dengan menggunakan teknik data mining dapat memprediksi lama alumni mendapatkan pekerjaan setelah menyelesaikan program studi S1. Menggunakan Metode Prediksi Data Mining dengan *Algoritma Naive Bayes Classifier.*

Rumusan masalah pada penelitian adalah data tidak diklasifikasi atau dikelompokkan sesuai dengan keterangan sehingga tidak dapat digunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya, artinya itu akan dapat menimbulkan masalah lain yaitu memprediksi masa tenggang waktu alumni Universitas Muhammadiyah Yogyakarta (UMY) mendapatkan pekerjaan setelah menyelesaikan studi S1.

Tujuan yang ingin dicapai pada penelitian ini adalah memprediksi perkiraan masa tenggang atau lama waktu alumni UMY. mendapatkan pekerjaan setelah menyelesaikan studi S1. Pengetahuan terkait prediksi dari lama alumni mendapatkan pekerjaan dapat diambil manfaat bagi calon alumni atau mahasiswa yang masih

aktif kuliah untuk lebih rajin untuk kuliah agar mendapatkan Indeks Prestasi Kuliah (IPK) diatas rata-rata, dan bisa lulus tepat waktu (Sabna dan Muhardi 2016). Selain itu, memanfaatkan jurusan yang sudah dijalani agar mendapatkan pekerjaan yang sesuai dengan jurusannya. Menfaat untuk universitas yaitu Pengelola Data Alumni UMY tidak perlu lagi menggunakan cara manual dalam memprediksi lama alumni mendapatkan pekerjaan.

Data Mining

Data mining adalah sebuah *knowledge discovery in database* (KDD) (Fithri dan Darmanto 2014). Data mining merupakan salah satu cara yang digunakan untuk mendapatkan pengetahuan baru dengan memanfaatkan jumlah data yang sangat besar. Beberapa teknik telah dikembangkan dan dimplementasikan untuk mengekstrak pengetahuan yang mungkin berguna untuk pengambilan keputusan. Teknik-teknik yang digunakan untuk pengekstrakan pengetahuan dalam data mining adalah pengenalan pola, clustering, memperlambat proses pencarian data yang dibutuhkan. Maka penulis akan menghitung dan asosiasi, prediksi dan klasifikasi (Defiyanti dan Jajuli 2015).

Ada beberapa teknik yang dimiliki *data mining* berdasarkan tugas yang dilakukan, yaitu (Ridwan, Suyono, dan Sarosa 2013) :

a. Deskripsi: Para peneliti biasanya mencoba untuk mendeskripsikan pola dan tren yang tersembunyi dalam data.
b. Prediksi: Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukan sesuatu yanng belum terjadi (mungkin terjadi di masa depan). perlu adanya Perkiraan Masa Tunggu Alumni Mendapatkan Pekerjaan.
c. Estimasi: Estimasi mirip dengan klasifikasi, kecuali variable tujuan yang lebih kearah numerik dari pada kategori.
d. Klasifikasi: Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam 3 kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
e. Clustering: Clustering lebih ke arah pengelomokan record, pengamatan, atau kasus dalam kelas yang memili kemiripan.
f. Asosiasi: Tugas asosiasi adalah mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

Algoritma Naive Bayes

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probalitas keanggotaan suatu *class.Bayesian classification* didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network. Bayesian classfication* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar. (Jananto 2013) *Naive Bayes* memiliki persamaan seperti berikut (Yuda, 2014)

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

Keterangan:

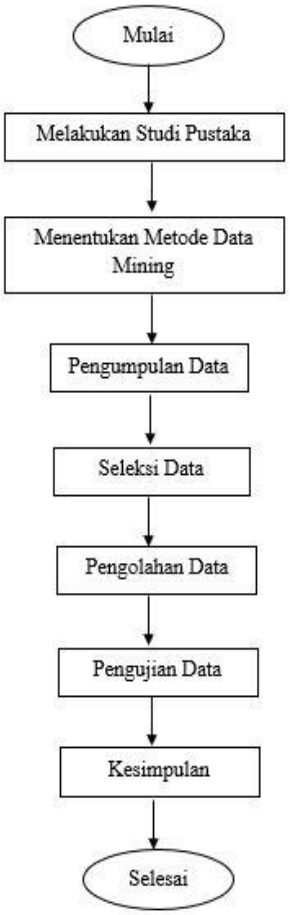
X	:	Data dengan class yang belum diketahui
H	:	Hipotesis data *X* merupakan suatu class spesifik
P(H|X)	:	Probanilitas hipotesis H berdasarkan kondisi *X*
P(H)	:	Probabilitas hipotesis *H*
P(X|H)	:	Probabilitas *X* berdasarkan kondisi hipotesis *H*
P(X)	:	Probalitas *X*

METODE PENELITIAN

Penerapan metodologi penelitian dilakukan dengan melakukan studi pustaka, menentukan metode data mining, pengumpulan data, seleksi data, pengolahan data, pengujian data, dan membuat kesimpulan seperti pada Gambar 1 (Saleh 2015).

1. Melakukan Studi Pustaka

Pertama hal yang dilakukan adalah melakukan studi pustaka yang berkaitan dengan penelitian-penelitian sebelumnya tentang penggunaan algortima *Naive Bayes*, serta untuk menguatkan penelitian ini berdasarkan teori yang digunakan. Dalam kasus ini peneliti ingin melakukan analisis pada masa tunggu alumni medapatkan pekerjaan di Universitas Muhammadiyah Yogyakarta dengan menggunakan algoritma *Naive Bayes* sekaligus untuk mengetahui pola dan cara kerja klasifikasi dalam *data mining* tersebut.



GAMBAR 1. Alur Metode Penilitian

2. Menentukan Metode Data Mining

Setelah melakukan studi pustaka tahap selanjutnya adalah menentukan metode yang sesuai untuk digunakan dalam teknik klasifikasi, setelah melakukan pengamatan dan observasi peneliti memilih menggunakan algoritma *Naive Bayes*. Karena algoritma *Naive Bayes* dapat lakukan transformasi atau mengubah nilai atribut data ke dalam bentuk data yang sesuai agar data dapat diproses menggunakan algoritma *Naive Bayes*. Sehingga akan diperoleh dataset utuh yang digunakan untuk proses ke tahapan selanjutnya.

3. Melakukan Tahapan Metode Klasifikasi

Berikut tahapan metode klasifikasi menggunakan algoritma *Naive Bayes* yang akan dilakukan peneliti untuk memperoleh hasil penelitian antara lain:

4. Melakukan Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan data dan ini merupakan tahapan yang penting karena dapat berpengaruh terhadap hasil penelitian, sehingga dalam mengumpulkan data harus dilakukan dengan benar. Peneliti mendapatkan data ini dari *database server* BSI (Biro Sistem Informasi) Universitas Muhammadiyah Yogyakarta.

5. Melakukan Seleksi Data

Tahap ini dilakukan seleksi terhadap data *database excel*. Karena data yang diperoleh tidak semuanya digunakan, dipilih sesuai dengan atribut atau variabel yang dibutuhkan dalam penelitian dengan melakukan seleksi data sehingga menjadi *dataset*. Sebagai contoh pada penelitian ini artibut yang dipilih adalah fakultas, tahun lulus, ipk, angkatan, dan jenis pekerjaan. Pada tahap ini akan menghilangkan data yang *null*, data tidak valid, dan data yang ganda. karena data yang kosong ataupun tidak valid akan berpengaruh terhadap hasil yang diperoleh. Kemudian data akan diolah menggunakan *Microsoft Office Excel* 2016.

6. Melakukan Pengolahan Data

Setelah semua data yang diperlukan telah dipilih, maka tahap penelitian selanjutnya adalah pengolahan data. Pada tahap ini akan

dilakukan transformasi atau mengubah nilai atribut data ke dalam bentuk data yang sesuai agar data dapat diproses menggunakan algoritma *Naive Bayes*. Sehingga akan diperoleh dataset utuh yang digunakan untuk proses ke tahapan selanjutnya.

7. Melakukan Pengujian Data

Pada tahap pengujian hasil akan dilakukan pengujian data baik secara manual dengan algoritma *Naive Bayes* dan menggunakan *softwar*.

8. Membuat Kesimpulan Penelitian

Berdasarkan hasil pengujian maka dapat ditarik kesimpulan yang mengacu pada rumusan masalah dan tujuan penelitian. Dan saran yang digunakan untuk mengembangkan penelitian selanjutnya serta dimasukkan untuk meningkatkan kualitas penelitian.

HASIL DAN PEMBAHASAN

Pada penelitian ini, yang menjadi obyek penelitian adalah alumni UMY dari tahun 2011-2014, dengan 7 fakultas. Langkah atau tahapan pada penelitian ini dilakukan sebagai berikut:

1. Studi literatur, yaitu proses awal dimana penulis mengumpulkan informasi yang berkaitan serta diperlukan pada penelitian ini.
2. Pengumpulan Data, yaitu pada proses ini penulis mendapatkan data awal sebanyak 689 data, dengan banyak attibut pada file excel yang didapatkan.
3. Seleksi Data, yaitu proses dimana penyeleksian dari banyaknya atribut yang ada, maka penulis menggunakan 4 atribut sebagai class biasa, dan 1 class utama sebagai *label* yaitu *gender*, *fakultas*, *ipk*, *tahun lulus*, *& tanggal mulai kerja.* Karena informasi yang terkandung didalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan *indicator* penelitian seperti pada Tabel 1 (Lumenta and Jacobus 2017).
4. Pembersihan data, yaitu proses dimana penghapusan dan menghilangkan data yang masih bernilai *null* (nol) Setelah penulis melakukan pemberihan data, maka data yang digunkan untuk penelitian ini sebanyak 435 data. Dan akan diolah menajdi 2 proses yaitu data *training* dan data *testing*.
5. Tranformasi Data (Tabel 2) dan Inisialisasi Data (Tabel 3 dan Tabel 4), yaitu tahap mengubah data menjadi bentuk yang sesuai untuk proses penelitian ini. Pengubahan dilakukan pada awalnya *id_prodi* diubah menjadi *fakultas*, *tanggal_mulai_kerja* diubah menjadi *status_mulai_kerja*, dan pada atribut IPK dan *Status_Mulai_Kerja value*-nya di inisialisasikan.

TABEL 1. Penyeleksian Data

Jenis_kelamin	Id_prodi	tahun_lulus	ipk	Tanggal_mulai_kerja
L	32	2003	NULL	NULL
L	11	2009	3.14	NULL
L	52	2003	2.04	NULL
L	104	NULL	NULL	NULL
L	61	2011	3.06	2013-11-01
L	11	2011	2.34	NULL

TABEL 2. Transformasi Data

Jenis_kelamin	Fakultas	ipk	tahun	Tanggal_mulai_kerja
L	HUKUM	3.00	2011	2015-11-01
L	TEKNIK	3.48	2011	2012-04-14
L	KEDOKTERAN	3.04	2011	2015-12-01
P	KEDOKTERAN	3.46	2011	2015-10-01
L	KEDOKTERAN	3.04	2011	2013-11-01
L	KEDOKTERAN	2.99	2011	2015-12-15

TABEL 3. Inisialisasi IPK

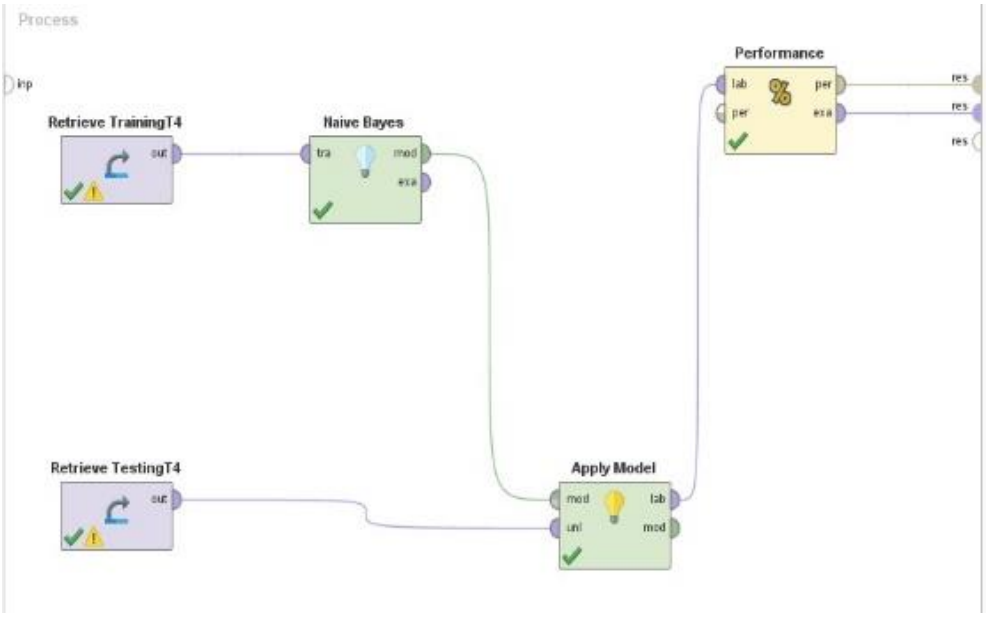
Nilai IPK	Inisialisasi
IPK<3	KURANG DARI 3
IPK 3>3.50	DARI 3 SAMPAI 3.50
IPK >3.50	LEBIH DARI 3.50

TABEL 4. Inisialisasi Status Mulai Kerja

Status Mulai Kerja	Inisialisasi
Lulus – 2 Tahun	Cepat
Lebih dari 2 Tahun	Lambat

6. Implementasi merupakan sebuah proses menentukan informasi dari data yang digunakan yaitu metode *prediksi* dengan algoritma *naive bayes* dengan *software RapidMiner*. Prosesnya yaitu *import file* data *training* dan data *testing* pada *software* yang digunakan lalu disambungkan dengan tools *operators* pada *RapidMiner* seperti Gambar 2.

Setelah dirun maka akan muncul hasil prediksi pada *view result*. Terlihat seperti pada Tabel 5 dibawah. Pada gambar di atas memberikan informasi tentang alumni yang cepat atau lambat mendapatkan pekerjaan. Pada Tabel 6 adalah tampilan tingkat *accuracy*, class prediction serta ketepatan prediksi, dan class recall (Murtopo 2016). Berdasarkan Gambar 3 bahwa alumni UMY pada tahun 2011 dan 2012 lebih lambat mendapatkan pekerjaan dibanding alumni 2013 dan 2014.

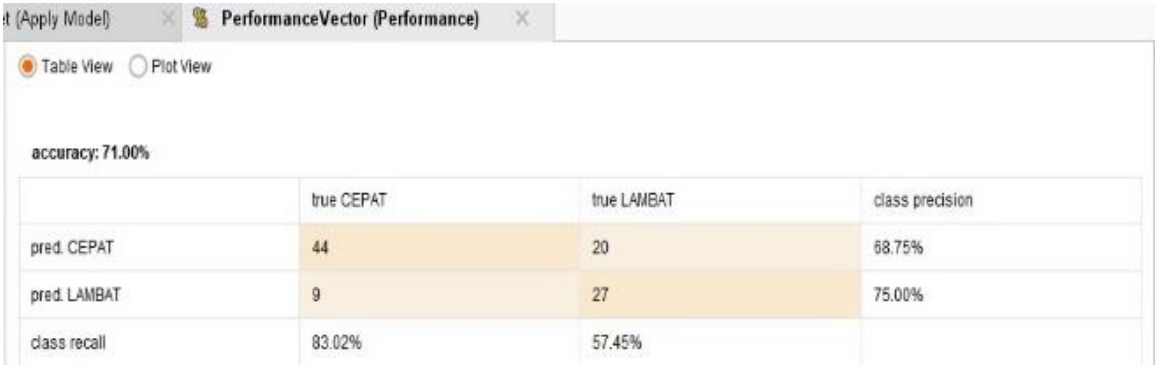


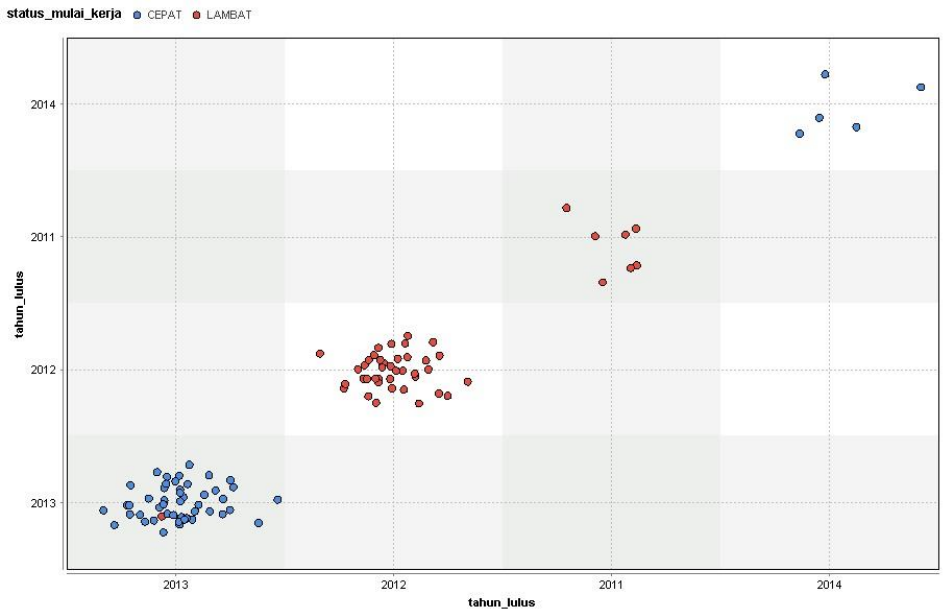
GAMBAR 2. Konfigurasi Operator Performance dengan RapidMiner

TABEL 5. Hasil Perhitungan (Example Set)

Row No.	status_mulia_kerja	prediction	Confidence (Cepat)	Confidence (Lambat)	Jenis_kelamin	fakultas	ipk	Tahun_lulus
1	CEPAT	CEPAT	0.757	0.243	P	ISIPOL	DARI 3 SAMPAI 3.50	2013
2	CEPAT	CEPAT	0.750	0.240	P	EKONOMI	DARI 3 SAMPAI 3.50	2013
3	CEPAT	CEPAT	0.625	0.175	L	PAI	LEBIH DARI 3.50	2013
4	LAMBAT	CEPAT	0.562	0.438	L	ISIPOL	DARI 3 SAMPAI 3.50	2012
5	LAMBAT	LAMBAT	0.060	0.940	P	KEDOKTERAN	DARI 3 SAMPAI 3.50	2011
6	LAMBAT	LAMBAT	0.546	0.454	L	TEKNIK	DARI 3 SAMPAI 3.50	2012

TABEL 6. Hasil Perhitungan Accuracy (*Performance Vector*)





GAMBAR 3. Grafik *Scatter*

TABEL 7. Perhitungan Model Naive Bayes Manual

P/tahun_lulus	CEPAT	LAMBAT
2011	3%	5%
2012	37%	66%
2013	54%	29%
2014	6%	1%
	100%	100%

Dari hasil diatas, dapat dilihat bahwa tingkat keakurasian dari data yang dipakai sebesar 71% dan *class prediction CEPAT* sebesar 68.75% *LAMBAT* sebesar 75.00%. Sedangkan untuk *class recall* sendiri untuk kelas *CEPAT* sebesar 83.02% dan untuk kelas *LAMBAT* sebesar 57.45%. Secara Umum precision, recall, dan accuracy dapat dirumuskan sebagai berikut (Hastuti 2012):

- Untuk Kelas CEPAT

$$\text{precition} = \frac{44}{44+20} = \frac{44}{64} = 0{,}6875 = 68{,}75\%$$

$$\text{recall} = \frac{44}{44+9} = \frac{44}{53} = 0{,}83018868 = 83{,}02\%$$

- Untuk Kelas LAMBAT

$$\text{precition} = \frac{27}{27+9} = \frac{27}{36} = 0{,}75 = 75\%$$

$$\text{recall} = \frac{27}{27+20} = \frac{27}{47} = 0{,}57446809 = 57{,}45\%$$

- Untuk Tingkat Accuracy

$$accuracy = \frac{44+27}{44+27+20+9} = \frac{71}{100} = 0{,}71 = 71\%$$

Berikut adalah hasil dari pengujian yang telah dilakukan pada RapidMiner dan Algoritma *naive bayes*, sebagai berikut:

1. Pada penelitian ini penulis menggunakan 435 data sebagai data *training* dan 100 data sebagi data *testing* sebagai pengujian manual di excel maupun pada pengujian menggunakan *software* RapidMiner.
2. Pengujian manual pada excel (Tabel 7) menggunakan model *naive bayes* untuk alumni tahun 2011 lebih banyak yang lambat mendapatkan pekerjaan, dengan perbandingan cepat 3% - lambat 5%, pada tahun 2012 lebih banyak yang lambat mendapatkan pekerjaan, dengan perbandingan cepat 37% - lambat 66%, pada tahun 2013 lebih banyak yang cepat mendapatkan pekerjaan, dengan perbandingan cepat 54% - lambat 29%, dan pada tahun 2014 lebih banyak yang cepat mendapatkan pekerjaan, dengan perbandingan cepat 6% - lambat 1%.
3. Untuk bagian *Class Prediction* menggunakan 100 data pada data *testing*. Terdapat data yang cocok antara class *cepat* dan class *predicted cepat* yaitu sebanyak 44 data dan untuk data lambat yang cocok antara class *lambat* dan class *predicted lambat* sebanyak 27 data. Sedangkan untuk prediction yang tidak tepat sebanyak 29 data.
4. Tingkat *accuracy* pada perhitungan manual di excel dan pengujian menggunakan *software* RapidMiner sama yaitu sebesar 71%. Untuk kelas CEPAT pada *RapidMiner* predictionnya sebesar 68,75%, dan *recall* pada kelas CEPAT sebesar 83,02%. Sedangkan prediction untuk kelas LAMBAT sebesar 75% dan *recall* pada kelas LAMBAT sebesar 57,45%.
5. Faktor yang memengaruhi hasil akhir *CEPAT* dan *LAMBAT* pada *software RapidMiner* alumni mendapatkan pekerjaan karena dari perhitungan manual model naive bayes pada excel lebih tinggi persen nya pada class *cepat*, yang terdiri dari atribut jenis kelamin, fakultas, ipk, tahun lulun, dan status mulai kerja.

KESIMPULAN

Setelah melakukan pengujian dan analisis penulis mendapatkan kesimpulan sebagai berikut:

1. Algoritma *Naive Bayes* dapat digunakan untuk memprediksi masa tenggang atau lama alumni UMY mendapatkan pekerjaan.
2. Algoritma *Naive Bayes* dalam memprediksi masa tenggang atau lama alumni mendapatkan pekerjaan memiliki tingkat *accuracy* dari *performanceVector* yaitu 71%, *class precision* yaitu CEPAT 68%, LAMBAT 75%, dan untuk *class recall* yaitu CEPAT 83,02% sedangkan LAMBAT 57,45%.
3. Faktor yang memengaruhi hasil akhir *CEPAT* dan *LAMBAT* pada *software RapidMiner* alumni mendapatkan pekerjaan karena dari perhitungan manual model naive bayes pada excel lebih tinggi persennya pada class *cepat*, yang terdiri dari atribut jenis kelamin, fakultas, ipk, tahun lulun, dan status mulai kerja.
4. Informasi yang didapatkan dalam penelitian ini adalah bahwa alumni UMY diprediksi mendapatkan pekerjaan lebih cepat setelah menyelesaikan studi S1.

UCAPAN TERIMA KASIH

Penulis dapat menuliskan ucapan terima kasih kepada CDC UMY, BSI UMY dan LP3M UMY yang telah membantu dalam penelitian yang dipublikasikan dalam jurnal ini.

DAFTAR PUSTAKA

Defiyanti, Sofi, and Mohamad Jajuli. (2015). "Integrasi Metode Klasifikasi Dan Clustering Dalam Data Mining." Konferensi Nasional Informatika (KNIF), 39–44.

Fithri, Diana Laily, and Eko Darmanto. (2014). "Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Nave Bayes." In . Muria Kudus University.

Hastuti, Khafiizh. (2012). "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif." Semantik 2 (1).

Jananto, Arief. (2013). "Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa." Teknologi Informasi DINAMIK 18 (1): 9–16.

Lumenta, Arie SM, and Agustinus Jacobus. (2017). "Prediksi Masa Studi Mahasiswa Dengan Menggunakan Algoritma Naïve Bayes." Jurnal Teknik Informatika Universitas Sam Ratulangi 11 (1).

Murtopo, Aang Alim. (2016). "Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma Naïve Bayes." CSRID (Computer Science Research and Its Development Journal) 7 (3): 145–54.

Ridwan, Mujib, Hadi Suyono, and M Sarosa. (2013). "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier." Eeccis 7 (1): 59–64. https://doi.org/10.1038/hdy.2009.180.

Sabna, Eka, and Muhardi Muhardi. (2016). "Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, Dan Hasil Belajar." Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi 2 (2): 41–44.

Saleh, Alfa. (2015). "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga." Creative Information Technology Journal 2 (3): 207–17.

Yuda, Nugroho Septian. (2014). "Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro.(Studi Kasus: Fakultas Ilmu Komputer Angkatan 2009)." Skripsi, Fakultas Ilmu Komputer.

PENULIS:

Asroni

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Yogyakarta, Bantul, D.I. Yogyakarta

Email: asroni@umy.ac.id

Nadiyah Maharty Ali

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Yogyakarta, Bantul, D.I. Yogyakarta

Email: mahartyali30@ft.umy.ac.id

Slamet Riyadi

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Yogyakarta, Bantul, D.I. Yogyakarta

Email: riyadi@umy.ac.id