

PREDIKSI PENERIMAAN KARYAWAN PT. BARNUM MENGGUNAKAN METODE NAIVE BAYES

Ibrah Maula Setya Aji¹, Muhammad Fikri Syahid²
Jurusan Teknik Informatika Fakultas Sains dan Teknologi
Universitas Islam Negeri Syarif Hidayatullah Jakarta

¹ibrah.aji18@mhs.uinjkt.ac.id

²fikri.syahid18@mhs.uinjkt.ac.id

Abstrak

Perusahaan pada umumnya merekrut pegawai yang baik dan memiliki kemampuan sesuai bidangnya untuk dapat bekerja secara maksimal dan tidak membuang-buang sumber daya maupun waktu dan tenaga. Perekrutan yang dilakukan terkadang tidak sesuai dengan apa yang dibutuhkan oleh perusahaan. Maka dari itu dibutuhkan sebuah metode untuk dapat mengidentifikasi calon karyawan. Penggunaan metode Naive Bayes untuk mengklasifikasi dan memprediksi calon karyawan yang berpotensi untuk masuk ke dalam perusahaan PT. BARNUM dengan melewati proses-proses dalam data mining seperti transformasi data sebelum menggunakan metode naive bayes. Dari data yang digunakan, keberhasilan prediksi dengan menggunakan metode Naive Bayes sebesar 86%.

Kata kunci: Calon karyawan, klasifikasi, prediksi, Naive Bayes, Data Mining

I. PENDAHULUAN

Pegawai merupakan sumber daya yang paling penting dalam suatu perusahaan dan pegawai yang baik dan memenuhi standar perusahaan hanya dapat diperoleh melalui proses penerimaan pegawai yang efektif. Pegawai yang baik dan berkualitas tentu saja memiliki dampak yang positif terhadap perusahaan tempat mereka bekerja. Oleh karena itu perusahaan harus menyeleksi orang-orang yang ingin masuk dan bekerja ke dalam perusahaan itu dengan baik.

Penerimaan calon pegawai baru merupakan sebuah tahap dimana sebuah perusahaan melakukan rekrutmen terhadap orang-orang yang melamar ke perusahaan tersebut dan menentukan apakah orang tersebut memenuhi kriteria dan kebutuhan unit kerja pada perusahaan tersebut. Alasan dilakukannya penerimaan calon pegawai baru adalah pengembangan unit usaha perusahaan PT. BARNUM yang mengakibatkan perlunya penambahan pegawai baru dan kebutuhan perusahaan tersebut untuk mengisi posisi kosong yang ditinggalkan oleh pegawai lama mereka.

Data mining merupakan sebuah proses penggalian informasi dari suatu data berukuran kecil maupun besar. Data mining yang bertujuan untuk mendapatkan pola-pola informasi tidak dilakukan begitu saja, data yang akan digunakan harus diolah terlebih dahulu, seperti pembersihan data, transformasi data, visualisasi data, baru

kemudian bisa didapat informasi dari data tersebut. Banyak metode yang digunakan dalam data mining, misalnya untuk klasifikasi: metode C4.5, C5.0, *Naive Bayes*. Untuk regresi: SVM (Support Vector Machine), Simple Linear Regression, Logistic Regression, dll.

Metode klasifikasi adalah pendekatan untuk menjalankan fungsi klasifikasi dalam data mining yaitu menggolongkan data. Teknik klasifikasi ini dapat pula digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Dan metode-metode yang telah disebutkan diatas seperti C4.5 dan *Naive Bayes* pada zaman modern ini sudah lebih mudah diimplementasikan. Bahasa-bahasa pemrograman sudah banyak menyediakan library metode klasifikasi dan tinggal menggunakannya saja, seperti pada bahasa R dan Python.

II. STUDI LITERATUR

Sebagai bahan referensi penulis menggunakan 3 jurnal :

1. Prediksi Ketepatan Masa Studi Mahasiswa dengan Algoritma Pohon Keputusan C45

Jurnal pertama yaitu penelitian mengenai prediksi ketepatan masa studi mahasiswa menggunakan algoritma C45. Jurnal ini dibuat oleh Sri Yunianita, Novi Setiani dan Sri Mulyati.

Penelitian pada jurnal ini menggunakan metode *10 fold cross-validation* sebagai penentu atribut yang digunakan juga metode filter dan wrapper pada subset fitur.

Pada bagian pembahasan didapati bahwa model yang dilatih dapat memprediksi data baru dengan akurasi 73.99% . Penulis dari jurnal ini menyarankan untuk menggunakan metode Naive Bayes pada penelitian selanjutnya. Dan karena inilah kami menggunakan metode Naive Bayes pada penelitian kali ini.

2. Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE

Pada jurnal kedua dengan topik prediksi penerimaan calon pegawai yang dibuat oleh Fandy Ferdian Harryanto dan Seng Hansun. Jurnal ini juga menggunakan metode C45 dengan *10 fold cross-validation* pada datasetnya.

Pada bagian hasil dan pembahasan didapati bahwa metode C45 dapat memprediksi calon pegawai baru dengan akurasi sebesar 71%.

Kami mengambil topik dari jurnal ini yaitu memprediksi calon karyawan yang akan diterima pada sebuah perusahaan.

Penulis jurnal ini berpesan untuk menggunakan dataset yang lebih besar, oleh karena itu kami memilih dataset yang lebih besar dari pada yang digunakan pada jurnal ini.

3. KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES

Penelitian pada jurnal ini menggunakan metode Naive Bayes untuk mengklasifikasikan masyarakat miskin. Penelitian ini menggunakan metode *split-validation* pada dataset nya dimana data latih sebanyak 90% dan data uji sebanyak 10%.

Hasil dari penelitian ini didapatkan bahwa algoritma Naive Bayes dapat mengklasifikasikan masyarakat miskin dengan akurasi sebesar 73%.

III. METODE PENELITIAN

Implementasi teknik data mining pada penelitian ini mengikuti tahapan dalam metode *Knowledge Discovery Process* yang diuraikan sebagai berikut:

A. Knowledge Discovery Goals

Tujuan dari penelitian ini adalah untuk menggali pengetahuan (*discovering knowledge*) mengenai pemodelan aturan untuk memprediksi penerimaan karyawan apakah tergolong dalam klasifikasi diterima atau tidak berdasarkan data yang dimiliki calon karyawan.

B. Data Integration

Data yang akan kami gunakan untuk penelitian ini kami dapatkan dari website yang bernama *kaggle*, dengan judul dataset itu sendiri yaitu Graduate Admissions 2. Dataset ini terdiri dari 9 variabel dan 500 baris data. Variabel-variabel ini terdiri dari Serial.No., GRE score, TOEFL score, University rating, SOP, LOR, CGPA, Research, dan Chance of admit.

C. Data Preprocessing

Untuk mengurangi variabel yang tidak relevan kami menggunakan metode seleksi fitur pada dataset. Serta merubah variabel terikat dari numerik menjadi nominal.

D. Data Mining

Selanjutnya kami akan melakukan proses data mining dengan algoritma naive bayes pada dataset yang telah kami lakukan seleksi fitur.

E. Knowledge Evaluation

Setelah kami melakukan data mining pada dataset akan kami lakukan evaluasi terhadap hasil dari data mining tersebut dengan menampilkan hasil diantaranya grafik dan akurasi terhadap prediksi yang kami inginkan.

IV. HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan pemrosesan data sampai mendapatkan hasil akurasi prediksi dari dataset yang digunakan menggunakan metode Naive Bayes. Penelitian ini menggunakan dataset

yang didapatkan dari Kaggle yang berjudul *graduate admissions* dengan jumlah 9 variabel dan terdiri dari 500 baris data. Variabel-variabel ini terdiri dari Serial.No., GRE score, TOEFL score, University rating, SOP, LOR, CGPA, Research, dan Chance of admit.

Pemrosesan data yang akan dilakukan secara berturut-turut adalah

A. Seleksi Fitur

Serial.No.	GRE.Score	TOEFL.Score	University.Rating	SOP	LOR	CGPA	Research	Chance.of.Admit
1	337	118	4	4.5	4.5	9.65	1	0.92
2	324	107	4	4.0	4.5	8.87	1	0.76
3	316	104	3	3.0	3.5	8.00	1	0.72
4	322	110	3	3.5	2.5	8.67	1	0.80
5	314	103	2	2.0	3.0	8.21	0	0.65
6	330	115	5	4.5	3.0	9.34	1	0.90

Gambar 1. Tabel Data Bagian Head

Pemrosesan data yang pertama dilakukan adalah dengan menggunakan uji korelasi untuk melihat korelasi antar variabel, berapakah besar korelasi antar variabel yang satu dengan yang lain dalam datanya. Dan variabel bebas dalam dataset ini adalah variabel chance of admit.

Serial.No.	GRE.Score	TOEFL.Score	University.Rating	SOP	LOR	CGPA	Research	Chance.of.Admit
Serial.No.	1.0000	0.0996	0.1426	0.0554	0.1442	0.0042	0.0751	0.0053
Research	0.0053	0.5785	0.4745	0.4354	0.4091	0.3762	0.5093	1.0000
LOR	0.0042	0.5144	0.5234	0.6023	0.6627	1.0000	0.6396	0.3762
SOP	0.1442	0.6207	0.6447	0.7294	1.0000	0.6627	0.7174	0.4091
University.Rating	0.0554	0.6434	0.6455	1.0000	0.7294	0.6023	0.7033	0.4354
TOEFL.Score	0.1426	0.8239	1.0000	0.6455	0.6447	0.5234	0.8095	0.4745
GRE.Score	0.0996	1.0000	0.8239	0.6434	0.6207	0.5144	0.8293	0.5785
CGPA	0.0751	0.8293	0.8095	0.7033	0.7174	0.6396	1.0000	0.5093
Chance.of.Admit	0.0017	0.8222	0.7936	0.7037	0.7028	0.6436	0.8888	0.5657

Gambar 2. Hasil Uji Korelasi

Dari hasil uji korelasi, didapat bahwa korelasi variabel bebas dan terikatnya. Variabel serial No, research, LOR, dan SOP merupakan 4 variabel yang tidak terlalu berpengaruh terhadap variabel chance of admit, sehingga 4 variabel tersebut akan kita eliminasi.

	GRE.Score	TOEFL.Score	University.Rating	CGPA	Chance.of.Admit
1	337	118	4	9.65	0.92
2	324	107	4	8.87	0.76
3	316	104	3	8.00	0.72
4	322	110	3	8.67	0.80
5	314	103	2	8.21	0.65
6	330	115	5	9.34	0.90

Gambar 3. Tabel Data Head Setelah Eliminasi

B. Transformasi Data

Langkah selanjutnya adalah merubah data value dari variabel chance of admit yang berbentuk data numerik ke dalam bentuk data ordinal berbentuk data YES/NO dengan kondisi apabila nilainya kurang dari 0.7, maka mengembalikan nilai NO, sedangkan apabila nilainya lebih dari sama dengan 0.7, maka mengembalikan nilai YES.

	GRE.Score	TOEFL.Score	University.Rating	CGPA	Chance.of.Admit
1	337	118	4	9.65	YES
2	324	107	4	8.87	YES
3	316	104	3	8.00	YES
4	322	110	3	8.67	YES
5	314	103	2	8.21	NO
6	330	115	5	9.34	YES

Gambar 4. Tabel Data Setelah Transformasi

C. Pembagian Data

Setelah dilakukan transformasi data, selanjutnya kita akan membagi data menjadi data latih dan data uji dengan data latih sebesar 80% dan data uji sebesar 20%.

	GRE.Score	TOEFL.Score	University.Rating	CGPA	Chance.of.Admit
1	337	118	4	9.65	YES
2	324	107	4	8.87	YES
3	316	104	3	8.00	YES
4	322	110	3	8.67	YES
6	330	115	5	9.34	YES
7	321	109	3	8.20	YES
8	308	101	2	7.90	NO
9	302	102	1	8.00	NO
10	323	108	3	8.60	NO
11	325	106	3	8.40	NO
12	327	111	4	9.00	YES
13	328	112	4	9.10	YES
15	311	104	3	8.20	NO

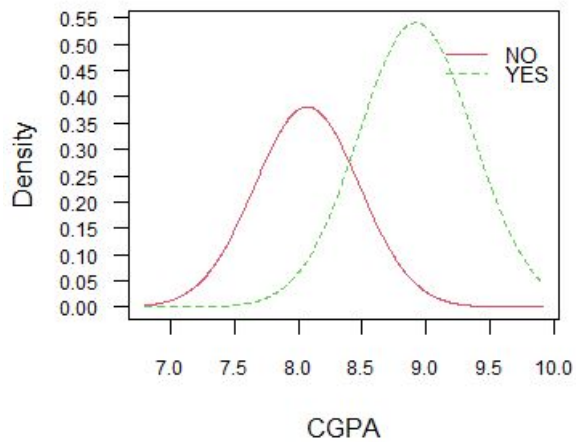
Gambar 5. Data Latih

	GRE.Score	TOEFL.Score	University.Rating	CGPA	Chance.of.Admit
5	314	103	2	8.21	NO
14	307	109	3	8.00	NO
16	314	105	3	8.30	NO
26	340	120	5	9.60	YES
28	298	98	2	7.50	NO
29	295	93	1	7.20	NO
39	304	105	1	7.50	NO
40	307	108	2	7.70	NO
60	311	104	2	8.30	NO
61	309	100	2	8.10	NO
72	336	112	5	9.76	YES
81	312	105	3	8.02	NO
86	319	103	4	8.66	YES

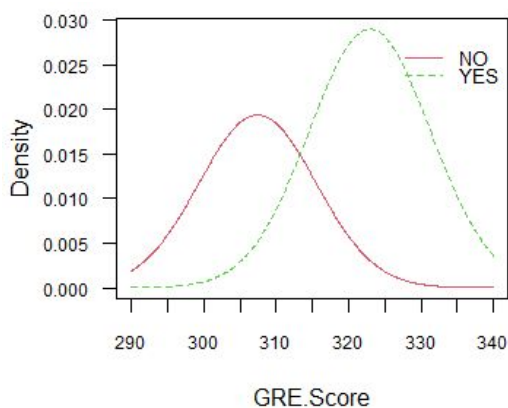
Gambar 6. Data Uji

D. Pembuatan Model

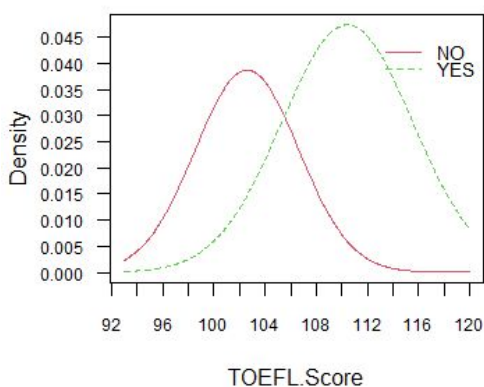
Selanjutnya kita membuat model naive bayes dari data latihnya menggunakan fungsi Naive Bayes yang telah disediakan di aplikasi RStudio yaitu, naive_bayes (variabel bebas terhadap variabel terikatnya, data = data latih). Kemudian menggunakan fungsi plot untuk memvisualisasikan datanya.



Gambar 7. Hasil Visualisasi variabel GPA terhadap variabel terikat



Gambar 8. Hasil Visualisasi variabel GRE Score terhadap variabel terikat



Gambar 9. Hasil Visualisasi variabel TOEFL Score terhadap variabel terikat



Gambar 10. Hasil Visualisasi variabel University Rating terhadap variabel terikat

Dari hasil pemrosesan data dengan metode naive bayes, didapat akurasi sebesar 0,8554 dari data latih dan akurasi sebesar 0,8587 dari data uji.

```
Accuracy : 0.8554
95% CI : (0.8175, 0.8881)
No Information Rate : 0.6078
P-Value [Acc > NIR] : < 2.2e-16
```

Gambar 11. Hasil Akurasi data latih

```
Accuracy : 0.8587
95% CI : (0.7705, 0.9226)
No Information Rate : 0.5652
P-Value [Acc > NIR] : 1.486e-09
```

Gambar 12. Hasil Akurasi data uji

Jadi dari hasil pengujian diatas, didapat tingkat keakuratan prediksi penerimaan karyawan baru di PT. BARNUM berdasarkan dataset yang ada adalah sebesar 86%.

V. KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari hasil penelitian diatas adalah bahwa dengan menggunakan metode naive bayes untuk memprediksi penerimaan karyawan baru di PT. BARNUM dengan dataset yang ada, didapat tingkat akurasi sebesar 86%. Diharapkan dengan menerapkan metode naive bayes, pihak-pihak dari universitas, SMK, dan instansi lain dapat membantu lulusan siswa/mahasiswanya agar dapat mendapatkan pekerjaan secara cepat dan tepat.

Saran yang diberikan penulis adalah mencoba memprediksi kembali dengan metode klasifikasi yang berbeda agar dapat menentukan metode mana yang lebih baik hasilnya. Dan mencoba menerapkan 10 fold cross validation daripada split validation.

REFERENSI

[1] Sri Yunianita, Novi Setiani , Sri Mulyati, “Prediksi Ketepatan Masa Studi Mahasiswa dengan Algoritma Pohon Keputusan C45”, Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Yogyakarta, 2018.

[2] Haditsah Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes”, Universitas Ichsan, Gorontalo, 2018.

[3] Fandy Ferdian Harryanto, Seng Hansun, “Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE”, Universitas Multimedia Nusantara; Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten-15811 Indonesia, 2017.