

ALGORITMA *NAÏVE BAIYES* UNTUK PREDIKSI PROFESI BERDASARKAN *SKILL JOB SEEKER*

Dewi Hastuti¹, Ayu Sabrina Syair², Asih Setiyorini³, WD. Rizqanun Karyima Bolu⁴,
Yuwanda Purnamasari⁵, Natalis Ransi⁶

^{1,2,3,4,5,6}Jurusan Teknik Informatika, Universitas Halu Oleo – Kendari – Sulawesi Tenggara

*¹dewi.hastuti97@yahoo.co.id, ²ayu.sabrina777@gmail.com, ³asetiyorini@gmail.com,

⁴rizqanunbolu@gmail.com, ⁵yuwandapurnamasari@gmail.com, ⁶natalis.ransi@uho.ac.id

Abstrak

Banyak hal yang dilakukan oleh organisasi ataupun perusahaan dalam mencari sumber daya manusia yang berkualifikasi dan berkompeten. Cara yang digunakan adalah memanfaatkan teknologi untuk memberikan informasi lowongan kerja dan melakukan seleksi calon pelamar. Namun, kedua hal tersebut masih sangat menyulitkan penyedia kerja dalam menyaring calon pelamar. Sehingga untuk mengatasi hal tersebut salah satu cara yang dapat digunakan adalah menerapkan teknik klasifikasi data mining menggunakan metode *Naïve Bayes* dalam memprediksikan profesi berdasarkan keahlian pencari kerja. Pada penelitian ini menggunakan 657 yang dipilih secara random dan diuji dengan menggunakan teknik *ten cross validation*. Hasil penelitian menunjukkan bahwa keakuratan akurasi mencapai persentase 97.26 %, atau terdapat 639 *data testing* terklasifikasi dengan benar.

Kata kunci— *Data Mining*, *Naïve Bayes*, *Job Seeker*.

1. PENDAHULUAN

Sumber daya manusia merupakan salah satu faktor yang sangat penting dari sebuah organisasi ataupun perusahaan. Hal ini disebabkan karena sumber daya manusia mempengaruhi efisien, efektivitas, kesuksesan perusahaan, juga merupakan pengeluaran pokok perusahaan dalam menjalankan bisnis [1]. Sampai saat ini sudah banyak penyedia kerja yang telah memanfaatkan teknologi untuk memberikan informasi lowongan kerja kepada pencari kerja.

Salah satu cara yang dapat dilakukan adalah seleksi calon pelamar. Seleksi adalah usaha yang sering dilakukan perusahaan untuk memperoleh karyawan yang berkualifikasi dan berkompeten yang akan menjabat serta mengerjakan semua pekerjaan pada perusahaan. Selain itu, beberapa perusahaan telah memanfaatkan bursa lowongan kerja yang ada di internet. Tetapi, sistem informasi lowongan kerja yang sudah ada masih menggunakan pencarian berdasarkan masing-masing atribut permintaan dan belum ada pengolahan data profil dari penyedia kerja dan pencari kerja [2]. Tetapi, pada kenyataannya

kedua hal itu masih menyulitkan penyedia kerja dalam menyaring calon pelamar yang berkualifikasi dan berkompeten, sehingga perlu dikembangkan sebuah metode yang dapat mempermudah dan menambah keakuratan dalam proses pemilihannya.

Metode data mining merupakan salah satu solusi yang dapat kita lakukan untuk memprediksi profesi berdasarkan skill pencari kerja. Salah satu metode data mining yang dapat digunakan adalah metode klasifikasi, seperti metode *Naïve Bayes* dan ID3.

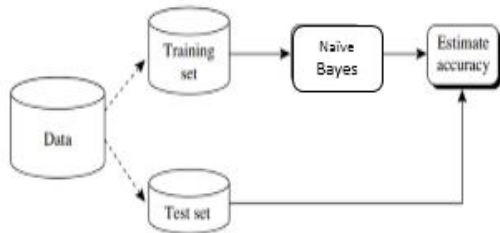
Pada penelitian sebelumnya metode *Naive Bayes* juga digunakan dalam memprediksi penyakit Dermatologi yang diabaikan tapi bahkan dapat menyebabkan kematian di mana metode *Naive Bayes* digunakan untuk mengenal pola data untuk mengungkap kemungkinan penyakit dermatologi [2]. Metode *Naive Bayes* juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi [3], misalnya algoritma ID3 yang dikenalkan oleh Ross Quinlan, menggunakan teori informasi untuk menentukan atribut yang paling informatif. Algoritma ini memiliki kelemahan dalam hal

ketidakstabilan dalam melakukanklasifikasi data apabila terjadi sedikit perubahan pada data training [4].

Berdasarkan pada penelitian sebelumnya, penelitian ini menggunakan Algoritma *Naïve Baiyes* untuk melakukan prediksi profesi berdasarkan *Skill Job Seeker*.

2. METODE PENELITIAN

Adapun penggambaran proses dari model pengelompokkan profesi berdasarkan keahlian pencari kerja digunakan seperti pada Gambar 1.



Gambar 1. Metode untuk mengestimasi keakuratan prediksi

Dalam paper ini data pelamar kerja atau calon pelamar yang digunakan diambil secara acak sedangkan data profesi diambil menurut syarat kemampuan lowongan kerja pada penelitian yang dilakukan oleh Chuck Litecky, Andrew Aken, Altaf Ahmad, and H. James Nelson [5], kemudian akan diuji coba dengan menggunakan *tool* weka.

Dalam paper ini juga menggunakan salah satu teknik memvalidasi yaitu *Cross validation*. *Cross Validation* merupakan teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Metode ini membagi data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Selanjutnya, setelah data diuji dilakukan proses silang dimana data pengujian lantas dijadikan data pelatihan ataupun sebaliknya, data pelatihan sebelumnya dijadikan kini menjadi data pengujian. Dalam paper ini kami menggunakan *10-Fold Cross-Validation* dari data 657 data yang disediakan.

2.1 Metode *Naïve Bayes*

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan

mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [3]. Definisi lain mengatakan *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [4].

Naïve Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu [5]. Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [6]. Adapun Persamaan (1) adalah persamaan teorema Bayes.

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

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class Spesifik ($H|X$): Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
- $P(H)$: Probabilitas hipotesis H (prior probability) $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
- $P(X)$: Probabilitas X

Secara garis besar model NBC (*Naïve Bayes Classifier*) ditunjukkan oleh Persamaan (2).

$$P(C|F_1, \dots, F_n) = \frac{P(C) \cdot P(F_1, \dots, F_n|C)}{P((F_1, \dots, F_n))} \quad (2)$$

Dimana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut

prior), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik – karakteristik sampel secara global (disebut juga evidence). Karena itu, Persamaan (2) dapat pula ditulis secara sederhana sebagai Persamaan (3).

$$\begin{aligned}
 & \text{Posterior} \\
 &= \frac{\text{Prior} \times \text{likelihood}}{\text{evidence}} \quad (3)
 \end{aligned}$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai – nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $P(C|F_1, \dots, F_n)$ menggunakan aturan perkalian seperti pada Persamaan (4).

$$\begin{aligned}
 P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\
 &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\
 &= P(C)|P(F_1|C) P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\
 &= P(C)|P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\
 &= P(C)|P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1})
 \end{aligned} \quad (4)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor – faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisis satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing – masing petunjuk (F_1, \dots, F_n) saling bebas (independen) satu sama lain. Dengan asumsi $P(F_1, \dots, F_n)$, maka berlaku suatu kesamaan yaitu :

$$\begin{aligned}
 P(P_i|F_j) &= \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \\
 & \text{Untuk } i \neq j, \text{ sehingga} \\
 & P(F_i|C, F_j) \\
 &= P(F_i|C) \quad (5)
 \end{aligned}$$

Dari Persamaan (5) dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi :

$$\begin{aligned}
 & P(C|F_1, \dots, F_n) \\
 &= P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots \\
 &= P(C) \prod_{i=1}^n P(F_i|C) \quad (6)
 \end{aligned}$$

3. HASIL DAN PEMBAHASAN

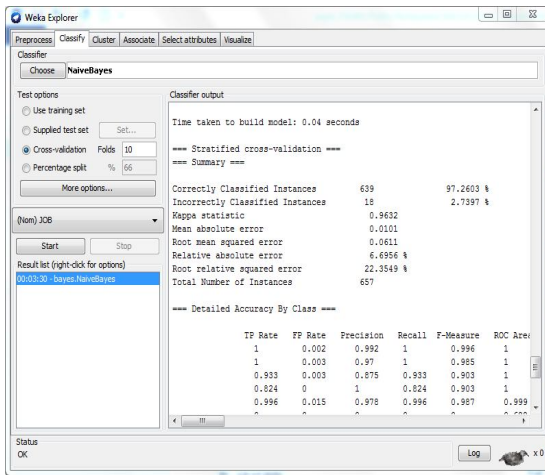
3.1 Penerapan Metode Naive Bayes

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam metode *Naive Bayes* data string yang bersifat konstan dibedakan dengan data numerik yang bersifat kontinyu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria dengan nilai data string maupun kriteria dengan nilai data numerik.

Tabel 1. Data Prediksi Berdasarkan Metode Naive Bayes di Tools Weka

The screenshot shows the Weka interface with a data table. The table has columns for various attributes (e.g., SMK, SD, SMP, SMA, etc.) and a predicted class column. The data rows show the results of the Naive Bayes classification for each instance.

Dari 657 data, semua diselesaikan dengan menggunakan tools weka sehingga dihasilkan hasil klasifikasi profesi berdasarkan keahlian pencari kerja seperti pada Gambar 2.



Gambar 2. Tingkat Keakuratan Data Menggunakan teknik Cross Validation di Tools Weka

4. ANALISIS

Berdasarkan Gambar 2 terlihat bahwa persentase untuk *Correctly Classified Instance* adalah sebesar 97.26% sementara persentase untuk *Incorrectly Classified Instance* adalah sebesar 2.74%. Di mana dari 657 data yang diuji, ada sebanyak 639 data yang berhasil diklasifikasikan dengan benar menggunakan metode *Naïve Bayes* dengan menggunakan teknik *cross validation* dan sebanyak 18 data tidak berhasil diklasifikasikan dengan benar.

5. KESIMPULAN

Berdasarkan penelitian tentang mengelompokkan profesi berdasarkan keahlian pencari kerja dapat ditarik beberapa kesimpulan sebagai berikut :

1. Metode *Naïve Bayes* memanfaatkan data *training* untuk menghasilkan probabilitas setiap kriteria untuk *class* yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk mengelompokkan profesi berdasarkan keahlian pencari kerja berdasarkan proses klasifikasi yang dilakukan oleh metode *Naïve Bayes* itu sendiri.
2. Berdasarkan data *testing* yang diuji dengan data *training*, metode *Naïve Bayes* berhasil mengklasifikasikan 639 data dengan benar. Sehingga metode *Naïve Bayes* berhasil memprediksi profesi berdasarkan keahlian pencari kerja dengan persentase keakuratan sebesar 97.26 %.

6. SARAN

Saran untuk penelitian selanjutnya adalah :

1. Penelitian tentang memprediksi profesi berdasarkan keahlian pencari kerja yaitu pengujian sebaiknya dilakukan dengan menggunakan metode lain seperti C4.5 untuk melihat metode mana yang lebih akurat dalam memprediksi profesi berdasarkan keahlian pencari kerja.
2. Sebaiknya data yang disediakan bukanlah data random yang ditetapkan sendiri melainkan data yang memiliki tingkat keakuratan yang lebih baik sehingga prediksi data yang dihasilkan lebih akurat lagi.

DAFTAR PUSTAKA

- [1] Rachman, Andy. 2008. Sistem Pendukung Keputusan Seleksi Sumber Daya Manusia Di Perusahaan, *Prosiding Seminar Nasional Teknoin 2008 Bidang Teknik Informatika*, ISBN : 978-979-3980-15-7, Hal 21-26.
- [2] Ichsan, Muhammad, Rahmad Syah, Muh. Iqbal. Tanpa Tahun. Pengembangan Model *Fuzzy Multiple Attribute Decision Making* (fmadm) Untuk Pengelolaan Lowongan Kerja. *TECHSI: Jurnal Penelitian Teknik Informatika*, Hal 1-19.
- [3] Bustami. Tanpa Tahun. Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI: Jurnal Penelitian Teknik Informatika*, Hal 127-146.
- [4] Andreas Prathama, Adry, Hapnes Toba. 2013. Sistem Pendukung Keputusan untuk Seleksi Administrasi Penerimaan Pegawai dengan Pohon Keputusan ID3 (Studi Kasus PT Jasamedika Saranatama). *Jurnal Informatika*, Vol. 9 No. 2, Hal 117 – 131.
- [5] Litecky, Chuck, Andrew Aken, Altaf Ahmad, dkk. Mining for Computing Jobs. 2010. IEEE Computer Society ISSN 0740-7459, Hal 78-85.
- [6] Manjusha K. K., Sankaranarayanan, K., Seena P., 2014, Prediction of Different Dermatological Conditions Using Naive Bayesian Classification, *International*

- Journal of Advanced Research in Computer Science and Software Engineering*, Vol 4, No 1, Hal 864-868.
- [7] Ting, S. L., Ip, W. H., Tsang, A. H.C., 2011, Is Naive Bayes a Good Classifier for Document Classification?, *International Journal of Software Engineering and Its Applications*, Vol. 5, No. 3, Hal 37-46.
- [8] Patil, T. R., Sherekar, M. S., 2013, Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, *International Journal of Computer Science and Applications*, Vol. 6, No. 2, Hal 256-261.
- [9] Bustami., 2013, Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, *TECHSI : Jurnal Penelitian Teknik Informatika*, Vol. 3, No.2, Hal. 127-146.
- [10] Saleh, Alfa.2015. Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga.*Citec Journal*, ISSN: 2354-5771, vol 2, No. 3, Hal 207-217
- [11] Pattekari, S. A., Parveen, A., 2012, Prediction System for Heart Disease Using Naive Bayes, *International Journal of Advanced Computer and Mathematical Sciences*, ISSN 2230-9624, Vol. 3, No 3, Hal 290-294
- [11] Ridwan, M., Suyono, H., Sarosa, M., 2013, Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *Jurnal EECCIS*, Vol 1, No. 7, Hal. 59-64
-