P-ISSN : 2549-7480
E-ISSN : 2549-6301

JURNAL ILMIAH INFORMATIKA
E-ISSN : 2549-6301
ISSN : 2549-7480
Faculty of Sains and Technology, Ibrahimy University
S4-Accredited : 105/E/KPT/2022
Journal page is available to : https://journal.ibrahimy.ac.id/index.php/JIMI

# PERBANDINGAN METODE KLASIFIKASI DATA MINING UNTUK DETEKSI KEASLIAN LOWONGAN PEKERJAAN DI MEDSOS

**Mohammad Malik Fajar [1)], Annisa Rizkiana Putri [2)], Khadijah Fahmi Hayati Holle [3)]**

[1] *Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim*
[2] *Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim*
[3] *Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim*

*email: [1] 18650033@student.uin-malang.ac.id, [2] 18650048@student.uin-malang.ac.id,*
*[3] khadijah.holle@uin-malang.ac.id*

**ARTICLE INFO**

**Article History:**
Recieved : 6 January 2022
Accepted : 20 June 2022
Published : 30 June 2022

**Keywords:**
Classification
Naïve Bayes
KNN
Decision Tree
Accuracy Value

**IEEE style in citing this article:**
M. M. Fajar, A. R. Putri, K. F. H. Holle, "Perbandingan Metode Klasifikasi Data Mining Untuk Deteksi Keaslian Lowongan Pekerjaan di Medsos", *Jurnal.ilmiah.informatika*, vol. 7, no. 1, pp. 41-48, Jun. 2022.

**Corresponding Author:**
Mohammad Malik Fajar
Universitas Islam Negeri Maulana Malik Ibrahim Malang

**ABSTRACT**

The COVID-19 pandemic has resulted in more and more people losing their jobs. Due to layoffs or bankrupt companies. This has resulted in many people looking for job vacancies. Job vacancies are circulating on social media but there are real and fake ones. Irresponsible people create job vacancies on social media with fraudulent purposes or for personal gain. So, a comparison of data mining classification methods was made for the detection of authenticity of job vacancies on social media. The method used is naive bayes, KNN, and decision tree. In order to find out which method has the highest accuracy value and can be used to classify the authenticity of job vacancies, and fraud on social media can be prevented. Based on this research, the method that has the highest accuracy value is the KNN method. The accuracy value is 94.93%, while the Decision Tree model has an accuracy value of 91.57% and the Naive Bayes model has an accuracy of 84.35%. The KNN method is the best method for classifying the authenticity of job vacancies.




DOI : https://doi.org/10.35316/jimi.v7i1.41-48

P-ISSN : 2549-7480
M. M. Fajar dkk/ JIMI 7 (1) pp. 41-48 E-ISSN : 2549-6301

## 1. PENDAHULUAN

Indonesia menempati posisi keenam terbesar di dunia dan keempat di Asia dalam tindak kejahatan di internet, meski tidak disebutkan secara rinci kejahatan macam apa saja yang terjadi di Indonesia maupun Warga Negara Indonesia yang terlibat dalam kejahatan tersebut. Hal ini sebagai peringatan bagi semua untuk mewaspadai kejahatan yang telah, sedang, dan akan muncul dari pengguna teknologi informasi. Banyak orang yang mencari pekerjaan melalui media sosial di zaman yang serba modern. Namun tidak semua lowongan pekerjaan yang ada di media sosial benar. Banyak orang yang tidak bertanggung jawab memposting lowongan pekerjaan dengan tujuan untuk penipuan atau kepentingan pribadi. Hal ini menyebabkan pencari kerja tidak mendapatkan pekerjaan yang layak namun ditipu oleh oknum tertentu.

Banyak modus penipuan yang terjadi di internet, salah satu modusnya adalah dengan cara memalsukan informasi lowongan dari perusahaan dan mempublikasikannya yang datanya seakan-akan milik perusahaan tersebut. Contoh kasus yang terjadi, tahun 2007 dan 2009 Perusahaan *Go Public* yang sering muncul dalam facebook, email dan google tentang Informasi lowongan kerja online. Bertujuan memberikan peluang kepada pencari kerja di era digital ini. Namun disamping itu sering sekali informasi lowongan kerja tersebut tidak benar atau palsu, hal tersebut menjadi masalah hukum yaitu tentang pemalsuan dan penipuan situs informasi lowongan kerja dengan motif untuk mencari keuntungan dengan cara memanipulasi situs informasi lowongan kerja tersebut untuk dipublikasikan, kemudian pencari kerja dalam melakukan pelamaran juga dilakukan secara online, hal ini seringkali meresahkan para pencari kerja dan perusahaan lainnya, karena nama baik perusahaan akan tercemar

Pada penelitian ini, penulis akan membandingkan model klasifikasi untuk menentukan suatu postingan lowongan pekerjaan pada media sosial merupakan postingan lowongan pekerjaan asli atau palsu. Dimana dalam pembuatan model-model klasifikasi ini memanfaatkan dataset yang didapatkan dari situs kaggle.com, dan pembuatan model ini dibuat melalui aplikasi rapidminer. Kemudian model-model tersebut akan dibandingkan performanya untuk mencari model klasifikasi yang terbaik dan memiliki nilai akurasi tertinggi. Dalam penelitian ini model klasifikasi yang digunakan atau yang akan dibandingkan adalah Naive Bayes, KNN (K-Nearest Neighbor), dan Decision Tree.

Berdasarkan penelitian yang dilakukan untuk membangun model klasifikasi peserta didik untuk mengelompokkan peserta program pendidikannya. Pada percobaannya, didapatkan hasil bahwasanya Decision Tree C 4.5 memberikan hasil terbaik dengan akurasi 96.73 % [1]. Berdasarkan penelitian terdahulu yang melakukan yang membandingkan model klasifikasi untuk prediksi penyakit kulit. Model klasifikasi yang akan dibandingkan antara lain Decision Tree, Naive Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM). Pada percobaannya, didapatkan hasil bahwasanya Naive Bayes dan SVM merupakan algoritma klasifikasi yang lebih baik dibandingkan C4.5 dan KNN untuk prediksi penyakit kulit [2], [3]. Penelitian selanjutnya dilakukan oleh Riyan Eko Putri, yang membandingkan metode naive bayes dan KNN dalam analisis data status kerja di Kabupaten Demak tahun 2012. Hasil dari penelitiannya diperoleh nilai laju error untuk metode naive bayes sebesar 0.0591 dan metode KNN sebesar 0.0394. Sehingga

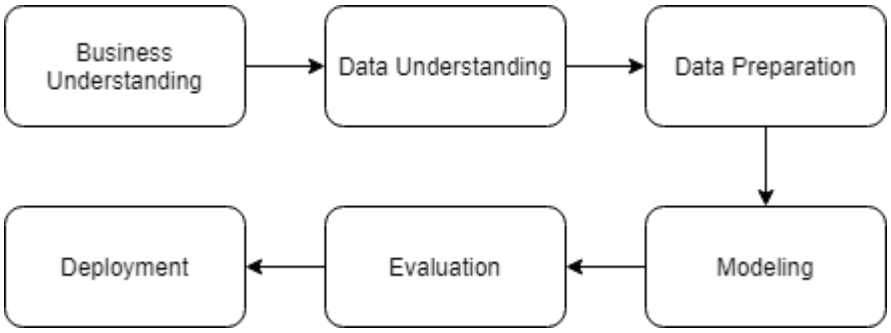

DOI : https://doi.org/10.35316/jimi.v7i1.41-48

P-ISSN: 2549-7480
M. M. Fajar dkk/ JIMI 7 (1) pp. 41-48 E-ISSN: 2549-6301

metode KNN lebih baik dibandingkan dengan metode naive bayes dalam hal mengklasifikasikan status kerja di Kabupaten Demak tahun 2012 [4]. Penelitian berikutnya dilakukan untuk melakukan klasifikasi masyarakat miskin di wilayah pemerintahan Kecamatan Tibawa Kab. Gorontalo menggunakan metode naive bayes. Berdasarkan hasil pengujian confussion matrix dengan teknik split validasi dan menggunakan metode naive bayes menghasilkan tingkat akurasi sebesar 73% dengan nilai Precision sebesar 92% dan Recall sebesar 86% [5].

## 2. METODE PENELITIAN

Metode penelitian yang digunakan adalah metode CRISP-DM. CRIPS-DM merupakan salah satu standar proses untuk data mining, dimana terdapat 6 proses yakni pada gambar 1.




Gambar 1. Metode Penelitian

Business understanding adalah tahapan pertama dalam metode CRISP-DM. Dibutuhkan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan [6]. Menentukan tujuan dari penelitian berdasarkan situasi bisnis yang ada. Sedangkan data understanding adalah proses pengumpulan dan mempelajari suatu dataset, sehingga berdasarkan data tersebut dan teridentifikasi masalah dan menghasilkan summary data yang bertujuan untuk memastikan apakah data sudah seperti yang diharapkan atau adanya penyimpangan yang perlu ditangani dalam tahap berikutnya. Yaitu tahap preparation. Pada tahap ini dilakukan pembersihan data yang akan digunakan dalam modeling. Melakukan perubahan pada variabel jika diperlukan. Melakukan data transformation sehingga data siap untuk modeling [7]. Dalam tahap modeling diterapkan algoritma data mining, dan memilih tools data mining untuk mendapatkan hasil. Kemudian dilakukan tahap evaluasi terhadap model yaitu interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada tahap sebelumnya [6]. Tahap terakhir deployment yaitu penyusunan laporan dari informasi yang telah didapatkan dari hasil evaluasi pada tahapan sebelumnya

## 3. HASIL DAN PEMBAHASAN

### *3.1 Business Understanding*

Problem yang diangkat dalam penelitian yaitu banyaknya pengangguran yang disebabkan oleh pandemi covid-19. Ada yang tempat kerjanya mengalami kebangkrutan atau menjadi pegawai yang ter PHK. Hal ini menyebabkan banyak orang berusaha mencari pekerjaan baik secara langsung maupun lewat media sosial. Postingan lowongan pekerjaan di media sosial jumlahnya tidak terbatas. Namun hal ini juga merugikan sebagian pencari kerja karena lowongan pekerjaan tersebut sebenarnya tidak ada atau palsu. Orang-orang yang tidak bertanggung


DOI : https://doi.org/10.35316/jimi.v7i1.41-48

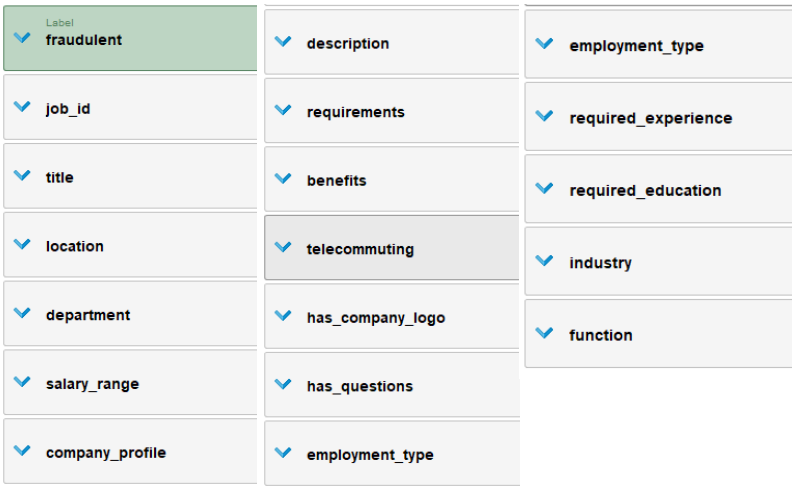
P-ISSN : 2549-7480
M. M. Fajar dkk/ JIMI 7 (1) pp. 41-48 E-ISSN : 2549-6301

jawab, membuat lowongan pekerjaan palsu untuk kepentingan pribadi dan penipuan.

Tujuan dari penelitian ini yaitu melakukan perbandingan antara metode klasifikasi decision tree, KNN dan naive bayes. Hasil yang terbaik dari 3 metode ini akan digunakan untuk mengklasifikasi lowongan pekerjaan yang ada di media sosial termasuk kedalam lowongan pekerjaan asli atau palsu.

### 3.2 Data *Understanding*

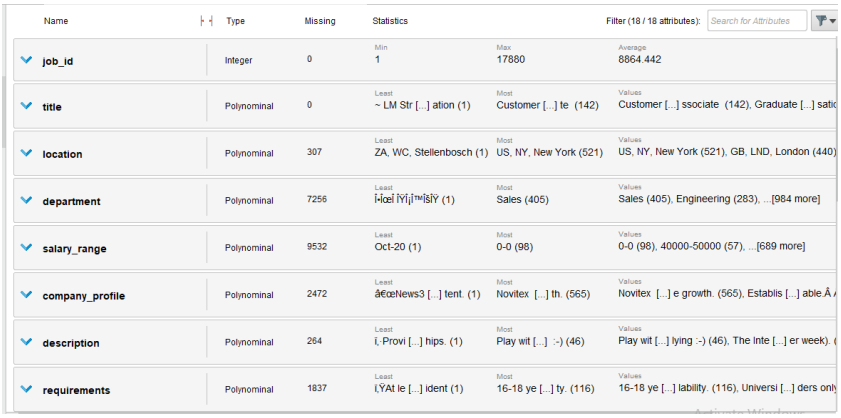
Penelitian ini menggunakan dataset yang bersumber dari kaggle.com. Dataset yang digunakan penulis berjudul *fake job postings*. Berisikan data lowongan pekerjaan yang benar dan palsu. Format dari dataset yaitu csv. Atribut yang ada dalam dataset yaitu *job_id, title, location, department, salary_range, company_profile, description, requirements, benefits, telecommuting, has_questions, employment_type, required_experience, required_education, industry, function dan fraudulent.*




Gambar 2. Atribut yang digunakan

### 3.3 *Data Preparation*

Data yang digunakan mencapai 17880 data dengan atribut sebanyak 18 atribut. Kemudian tipe dari data ini adalah *polynomial* dan juga *binomial*. Missing data ditemukan dalam dataset yang digunakan sebanyak 49468 data seperti pada gambar 3.

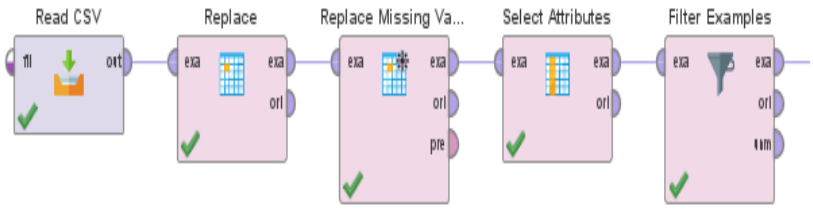



Gambar 3. Missing Data Dalam Dataset

Pada tahap data *preparation* langkah pertama yaitu menggunakan *Read CSV* untuk membaca dataset. Kemudian sambungkan dengan *Replace* dan *Replace Missing Value* untuk mengganti data-data pada atribut *company_profile* dengan *value* 0


DOI : https://doi.org/10.35316/jimi.v7i1.41-48

untuk *missing value* dan *value* 1 untuk yang bukan *missing value*. Kemudian langkah selanjutnya disambungkan dengan *Select Attributes* yang digunakan untuk memilih atribut mana saja yang akan digunakan atau yang akan dihapus. Dan yang terakhir yakni *Filter Examples* yang digunakan untuk memilih atau menyaring data yang tidak ada *missing value* nya. Tahap-tahap tersebut digambarkan pada gambar 4.




Gambar 4. Data *Preparation*

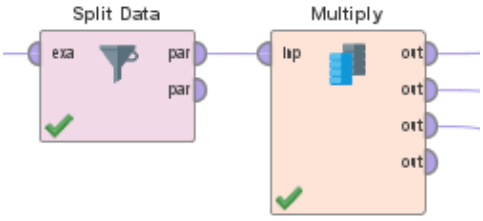
Setelah dijalankan tahap-tahap tersebut, maka missing value pada dataset akan hilang dan siap untuk diproses. Hasil data preparation seperti pada gambar 5.

| | | | | | |
|---|---|---|---|---|---|
| fraudulent | Binominal | 0 | Negative: 0 | Positive: 1 | Values: 0 (1190), 1 (102) |
| company_profile | Polynominal | 0 | Least: 0 (301) | Most: 1 (991) | Values: 1 (991), 0 (301) |
| job_id | Polynominal | 0 | Least: 9999 (0) | Most: 10007 (1) | Values: 10007 (1), 10014 (1), ... |
| title | Polynominal | 0 | Least: ~ LM Str [...] ation (0) | Most: Customer [...] tive (28) | Values: Customer [...] sentative |
| location | Polynominal | 0 | Least: ZM, , (0) | Most: GB, LND, London (108) | Values: GB, LND, London (108), |
| salary_range | Polynominal | 0 | Least: Oct-20 (0) | Most: 0-0 (50) | Values: 0-0 (50), 40000-50000 |
| requirements | Polynominal | 0 | Least: ï¿½At le [...] ident (0) | Most: Experien [...] seÂ (13) | Values: Experien [...] d mouseÂ |

Gambar 5. Hasil Data *Preparation*

### *3.4 Modeling*

Tahapan berikutnya yaitu *modeling*. Setelah data siap digunakan, maka dilakukan *split* data yang bertujuan membagi dataset menjadi data *training* dan data *testing*. Selanjutnya sambungkan dengan *multiply*, seperti gambar 6.

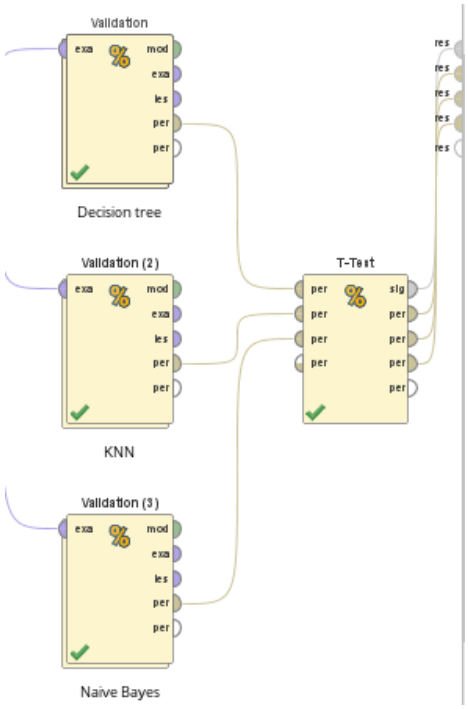



Gambar 6. *Split* Data dan *Multiply*

Kemudian Model algoritma yang digunakan decision tree, KNN dan naive bayes. Digunakan *cross validation nominal* untuk menggunakan model-model tersebut karena data berupa nominal. Kemudian masing-masing *cross validation* dari model disambungkan ke T-test, seperti gambar 7.

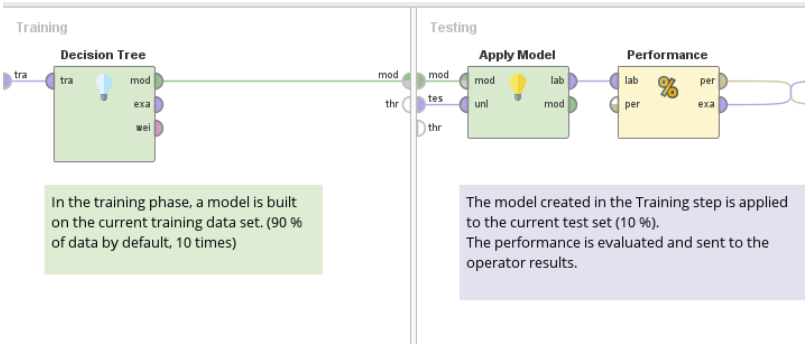

DOI : https://doi.org/10.35316/jimi.v7i1.41-48

P-ISSN : 2549-7480
M. M. Fajar dkk/ JIMI 7 (1) pp. 41-48 E-ISSN : 2549-6301

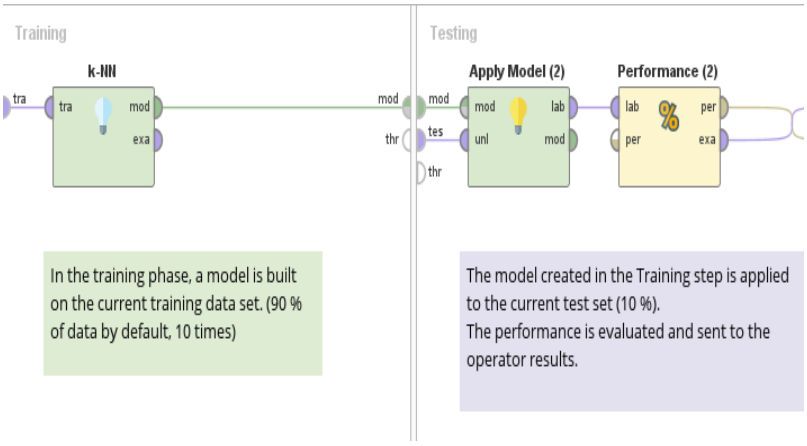



Gambar 7. Gambar *Modeling*

Di dalam *cross validation nominal* tersebut berisi pemodelan-pemodelan dari masing-masing algoritma. Gambar dari pemodelan algoritma Decision Tree, KNN, dan Naive Bayes didalam *cross validation nominal* digambarkan pada gambar 8, gambar 9, dan gambar 10.

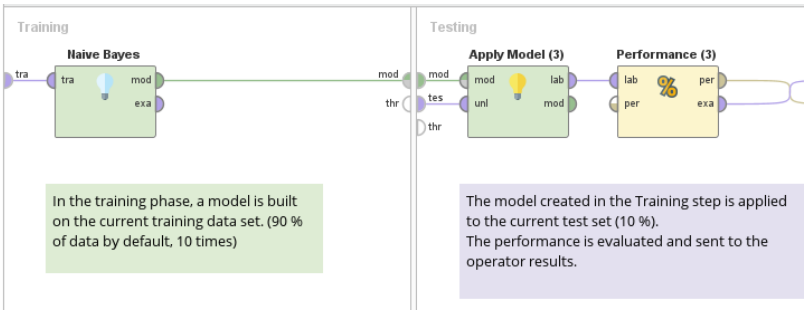



Gambar 8. Gambar Model *Decision Tree*




Gambar 9. Gambar Model KNN


DOI : https://doi.org/10.35316/jimi.v7i1.41-48



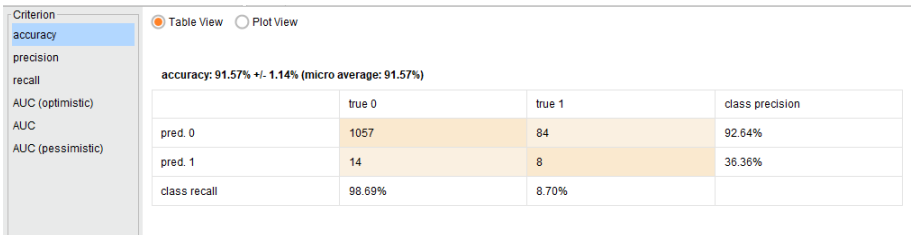

Gambar 10. Gambar Model N*aive Bayes*

### *3.5 Evaluation*

Dalam tahap *evaluation* dilihat dari hasil *accuracy* yang dihasilkan tiap-tiap model. Hasil *accuracy* dari tiap model didapatkan bahwa model yang memiliki *accuracy* paling tinggi adalah model KNN dengan nilai *accuracy* sebesar 94.93%, sedangkan untuk model Decision Tree memiliki nilai *accuracy* sebesar 91.57% dan untuk model naive bayes memiliki *accuracy* sebesar 84.35%. Untuk hasil T-test dapat dilihat di gambar 11, sedangkan untuk detail nilai *accuracy* tiap model dapat dilihat di gambar 12, gambar 13, dan gambar 14.

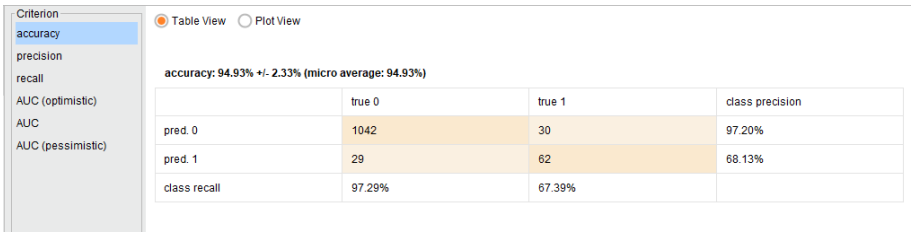
| A | B | C | D |
|---|---|---|---|
| | 0.916 +/- 0.011 | 0.949 +/- 0.023 | 0.843 +/- 0.036 |
| 0.916 +/- 0.011 | | 0.001 | 0.000 |
| 0.949 +/- 0.023 | | | 0.000 |
| 0.843 +/- 0.036 | | | |

Gambar 11. Gambar Hasil T-test



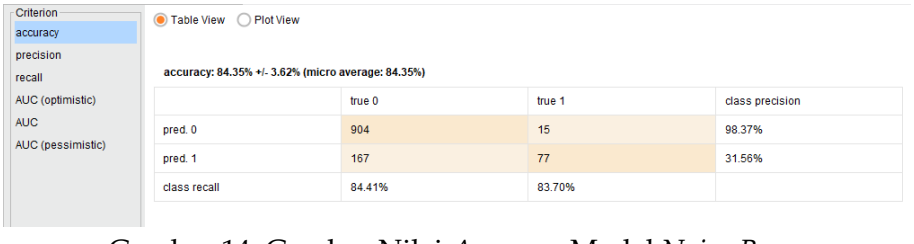

| | true 0 | true 1 | class precision |
|---|---|---|---|
| pred. 0 | 1057 | 84 | 92.64% |
| pred. 1 | 14 | 8 | 36.36% |
| class recall | 98.69% | 8.70% | |

Gambar 12. Gambar Nilai *Accuracy* Model *Decision Tree*




| | true 0 | true 1 | class precision |
|---|---|---|---|
| pred. 0 | 1042 | 30 | 97.20% |
| pred. 1 | 29 | 62 | 68.13% |
| class recall | 97.29% | 67.39% | |

Gambar 13. Gambar Nilai *Accuracy* Model KNN




| | true 0 | true 1 | class precision |
|---|---|---|---|
| pred. 0 | 904 | 15 | 98.37% |
| pred. 1 | 167 | 77 | 31.56% |
| class recall | 84.41% | 83.70% | |

Gambar 14. Gambar Nilai *Accuracy* Model *Naive Bayes*


DOI : https://doi.org/10.35316/jimi.v7i1.41-48

P-ISSN : 2549-7480
M. M. Fajar dkk/ JIMI 7 (1) pp. 41-48 E-ISSN : 2549-6301

### *3.6 Deployment*

Model yang telah digunakan untuk mendeteksi keaslian postingan pekerjaan di medsos sudah bisa dijalankan, namun nilai accuracy nya perlu ada peningkatan lagi agar benar-benar bisa dan lebih akurat untuk mendeteksi keaslian postingan pekerjaan di medsos. Untuk meningkatkan accuracy nya bisa dilakukan dengan menambahkan jumlah data yang digunakan untuk proses mining, menambahkan varian data, semakin data itu bervariasi semakin banyak menghasilkan pengetahuan, dan mencoba model lain tentunya juga dengan menambahkan jumlah dana dan menambahkan varian data.

## 4. KESIMPULAN

Penelitian mengenai perbandingan metode klasifikasi data mining untuk deteksi keaslian lowongan pekerjaan di media sosial menggunakan 3 metode. Metode decision tree, KNN, dan naive bayes. Model yang menghasilkan nilai accuracy tertinggi adalah model KNN dengan nilai sebesar 94.93%, sedangkan untuk model Decision Tree memiliki nilai *accuracy* sebesar 91.57% dan untuk model naive bayes memiliki *accuracy* sebesar 84.35%. Berdasarkan nilai accuracy yang didapatkan, model yang paling baik digunakan untuk klasifikasi deteksi keaslian lowongan pekerjaan di media sosial yaitu model KNN.

## 5. REFERENSI

[1] I. Sutoyo, "Perbandingan 5 Algoritma Data Mining Untuk Klasifikasi Data Peserta Didik," *Simnasiptek 2017*, 2017.

[2] D. Prajarini, S. Tinggi, S. Rupa, D. Desain, and V. Indonesia, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit," *Informatics Journal*, vol. 1, no. 3, 2016.

[3] N. I. Wibowo, T. A. Maulana, H. Muhammad, and N. A. Rakhmawati, "Perbandingan Algoritma Klasifikasi Sentimen Twitter Terhadap Insiden Kebocoran Data Tokopedia," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 6, no. 2, 2021, doi: 10.14421/jiska.2021.6.2.120-129.

[4] R. E. Putri, Suparti, and R. Rahmawati, "Perbandingan Metode Klasifikasi Naãve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012," *Jurnal Gaussian*, vol. 3, no. 4, 2014.

[5] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *ILKOM Jurnal Ilmiah*, vol. 10, no. 2, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.

[6] G. Fiastantyo, "Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naive Bayes dan Algoritma C4.5 untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," *Semantic Journal*, 2014.

[7] A. W. Indra Purnama, Ragil Saputra, "Implementasi Data Mining Menggunakan Crisp-Dm Pada Sistem Informasi Eksekutif Dinas Kelautan Dan Perikanan Provinsi Jawa Tengah," *Annual Review of Information Science and Technology*, vol. 36, 2017.


DOI : https://doi.org/10.35316/jimi.v7i1.41-48