

PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENDETEKSI PENIPUAN LOWONGAN PEKERJAAN

Muhammad Hairu Dzikri*¹, Iwan Rizal Setiawan², Didik Indrayana³

¹²³Program Studi Teknik Informatika, Universitas Muhammadiyah Sukabumi, Sukabumi, Indonesia
e-mail: ^{1*} mhairudzikri018@ummi.ac.id, ² myfrank@gmail.com, ³ didik.ind@ummi.ac.id

Banyak orang kini mencari pekerjaan melalui media sosial. Namun, tidak semua lowongan pekerjaan di media sosial itu benar. Banyak orang yang tidak bertanggung jawab memposting lowongan pekerjaan untuk tujuan penipuan atau keuntungan pribadi. Akibatnya pencari kerja tidak dapat memperoleh pekerjaan yang layak dan tertipu oleh orang-orang tertentu. Ada berbagai jenis penipuan di internet. Perusahaan sering kali memalsukan informasi sumber daya manusia dan mempublikasikannya seolah-olah itu adalah informasi perusahaan, mempostingnya di facebook, email, situs web, Google, dll. Oleh karena itu, penelitian ini fokus pada pendeteksian penipuan pada iklan lowongan kerja dengan menggunakan teknik algoritma Naive Bayes. Dan hasil dari penelitian ini menghasilkan data bukan penipuan sebesar 106 data, dan penipuan sebesar 94 data, total akurasi keseluruhan sebesar 89%, Untuk label bukan penipuan menghasilkan precision sebesar 83%, recall sebesar 100%, dan f1-score sebesar 91%. Untuk label penipuan menghasilkan precision sebesar 100%, recall sebesar 76%, dan f1-score sebesar 86%.

Kata Kunci— *Lowongan pekerjaan, Penipuan, Naive Bayes, NLP, Python*

I. PENDAHULUAN

Banyak orang yang mencari suatu pekerjaan melalui media sosial. Tetapi, tidak semua lowongan pekerjaan di media sosial itu jujur, Banyak oknum-oknum yang tidak bertanggung jawab memposting lowongan pekerjaan dengan tujuan penipuan atau demi keuntungan pribadi. Akibatnya para pencari kerja tidak bisa mendapatkan pekerjaan yang benar dan akhirnya ditipu oleh orang-orang tertentu. Berbagai jenis penipuan terjadi di Internet. Terdapat contoh perusahaan yang memalsukan informasi rekrutmen dan mempostingnya seolah-olah itu adalah informasi perusahaan. Postingan pekerjaan online sering kali diposting di Facebook, Email, Website dan Google. bertujuan untuk memberikan peluang bagi pencari kerja di era digital. Namun, selain itu, banyak informasi palsu dan menyesatkan mengenai rekrutmen, terutama informasi palsu dan penipuan yang memanipulasi situs informasi rekrutmen untuk tujuan keuntungan. Hal ini sering menjadi permasalahan bagi para pencari kerja. [1] Penipu menawarkan kesempatan

kerja yang menguntungkan kepada pencari kerja dan menuntut kompensasi alternatif calon pencari kerja didorong untuk berinvestasi. Ini adalah masalah yang berbahaya [2]

Saat ini masih belum ada program untuk mendeteksi kalimat penipuan lowongan pekerjaan dengan menggunakan frasa penipuan dan bukan penipuan, Dan tujuan penelitian ini untuk mendeteksi suatu penipuan dalam kalimat lowongan pekerjaan berdasarkan frasa penipuan atau bukan penipuan menggunakan *software Jupyter Notebook* dengan algoritma *Naive Bayes*. Maka dari itu dalam penelitian ini berfokus untuk merancang program klasifikasi untuk mengklasifikasikan kalimat lowongan pekerjaan itu apakah termasuk kedalam label penipuan atau bukan penipuan. dengan teknologi pemrosesan bahasa alami (*NLP*) yang menggunakan bahasa pemrograman *python*.

Terdapat perbandingan penelitian *Naive Bayes* dengan penelitian lain. Perbandingan dengan *SVM* yang digunakan untuk mengklasifikasikan spam SMS Indonesia dengan hasil yang akurat. Algoritma *Naive Bayes* lebih unggul dalam hal recall sebesar 94% [3] *Naive Bayes* menggunakan teknik *SMOTE* menunjukkan performa terbaik dengan akurasi 71,66%, disusul *Naive Bayes* tanpa *SMOTE* (76,03%) dan *KNN* dengan *SMOTE* (80,47%). *ANN* tanpa *SMOTE* menunjukkan akurasi tertinggi (83,02%), sedangkan *Naive Bayes* menggunakan *SMOTE* menunjukkan keseimbangan yang lebih baik antara presisi, presisi, dan recall. Kami menunjukkan bahwa teknik *SMOTE* dapat meningkatkan kinerja *Naive Bayes* dengan meningkatkan presisi dan perolehan serta mengatasi ketidakseimbangan kelas dalam kumpulan data. [4] *Naive Bayes*, Pohon Keputusan, dan *ANN*. Perbandingan data latih dan data uji adalah 80: 20. Nilai akurasi yang diperoleh untuk *Naive Bayes* sebesar 85,56%, nilai akurasi yang diperoleh untuk *Decision Tree* sebesar 72,78%, dan *KNN* untuk aplikasi Twitter. [5] Kesimpulan dan pembahasan studi perbandingan algoritma *Naive Bayes* dan algoritma *K-NN* untuk mengklasifikasikan teman MI AL-Muslim dari 186 data siswa yang terdiri dari 150 data latih dan 36 data uji digunakan 36 algoritma klasifikasi. Pada data uji, metode *K-Nearest Neighbors* masing-masing mencapai 91,67%. Berdasarkan skor akurasi kedua algoritma, akurasi klasifikasi tergolong baik, dengan algoritma *Naive Bayes* mengungguli algoritma *K-Nearest Neighbors* dalam mengklasifikasikan peer. [6] Metode

klasifikasi lain yang disebut metode *Naive Bayes* digunakan untuk pengolahan data. Berdasarkan perbandingan kedua algoritma, akurasi algoritma *Naive Bayes* lebih baik dibandingkan dengan C4.5, selisihnya 11,77%. Akurasi algoritma C4.5 menjadi 58,82 dengan menggunakan uji validitas cross validation. Pada uji validitas dengan validasi silang, rata-rata skor akurasi menggunakan algoritma *Naive Bayes* sebesar 70,59. Dengan menggunakan uji validitas cross validation diperoleh rata-rata skor akurasi kelas kurang memuaskan sebesar 0%, akurasi kelas kurang memuaskan sebesar 76,92%, akurasi kelas kurang memuaskan 0% sebesar 33,33%. Akurasi kelas yang sangat kurang memuaskan adalah: 100. Tidak puas dengan tingkat panggilan balik" adalah 0,91%, tingkat panggilan balik kelas adalah 25%, tingkat panggilan balik kelas adalah 25%, tidak puas 100%. [7]

Dari semua hasil perbandingan dengan penelitian lain maka penelitian dengan menggunakan *Naive Bayes* ini menghasilkan hasil yang lebih akurat jika dibandingkan penelitian yang lain dengan hasil penelitian total akurasi keseluruhan sebesar 89%, Untuk label bukan penipuan menghasilkan precision 83%, recall 100%, dan f1-score s 91%. Untuk label penipuan menghasilkan precision 100%, recall 76%, dan f1-score 86%. Dari total hasil tersebut hasilnya cukup tinggi jika dibandingkan dengan penelitian lain maka penelitian ini hasilnya lebih tinggi, lebih akurat, dan algoritma *Naive Bayes* sangat cocok untuk penelitian ini.

II. LANDASAN TEORI

A. Data Mining

Data mining merupakan teknik penambangan data dalam jumlah yang besar dari database besar yang dapat dipahami, berguna, dan digunakan untuk mengambil keputusan bisnis yang penting. Serangkaian teknik yang bertujuan untuk menemukan pola di dalam data yang sudah dikumpulkan. [8]

B. Naive Bayes

Naive Bayes merupakan metode klasifikasi berbasis probabilitas sederhana yang dimaksudkan untuk digunakan dengan asumsi bahwa variabel penjelas bersifat independen. Algoritma ini berfokus pada memperkirakan probabilitas pembelajaran. Keuntungan algoritma *Naive Bayes* adalah memiliki tingkat kesalahan yang rendah ketika kumpulan datanya besar, dan meningkatkan akurasi dan kecepatan ketika diterapkan pada kumpulan data yang lebih besar. *Naive Bayes* dirumuskan pada persamaan. [9]

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

$$P(\text{Cspam}|X) = \frac{P(X|\text{Cspam}) \cdot P(\text{Cspam})}{P(X)} \quad (2)$$

Kemungkinan bahwa kalimat bukan penipuan di fitur-fitur X dalam $P(\text{Cham}|X)$

Kemungkinan bahwa kalimat adalah penipuan di fitur-fitur X dalam $P(\text{Cspam}|X)$

Kemungkinan dari fitur X dalam $P(X|\text{Cham})$ mengingat kalimat bukan penipuan

Kemungkinan dari fitur X adalah kalimat penipuan,

$P(\text{Cham})$ adalah kemungkinan kalimat bukan penipuan.

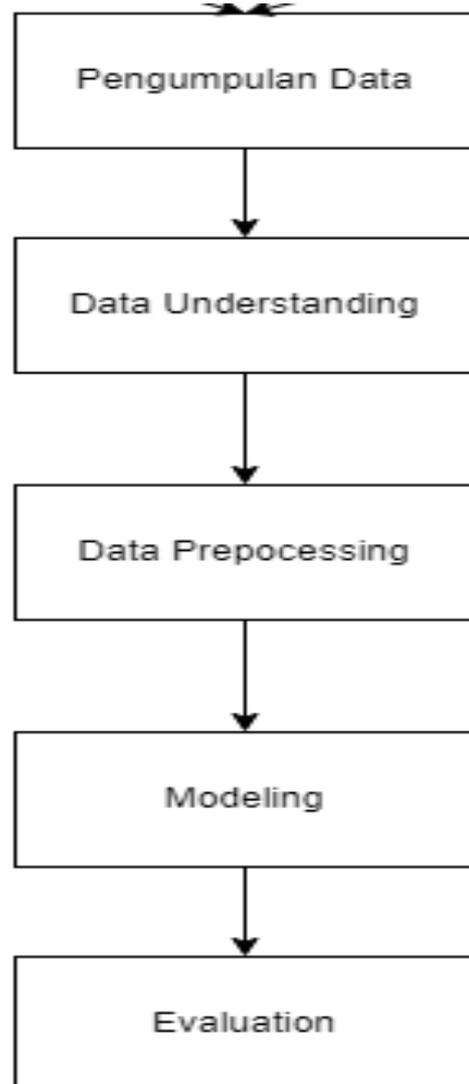
$P(\text{Cspam})$ adalah kemungkinan kalimat adalah penipuan, X adalah vektor fitur yang mewakili pesan,

Dalam praktiknya memperkirakan kemungkinan ini menggunakan data latih yang terdiri dari kalimat yang ditandai sebagai (penipuan dan bukan penipuan) dan membangun model Bayesian sederhana yang digunakan untuk mengklasifikasikan kalimat baru. Ciri-cirinya yaitu kemunculan kata-kata dalam kalimat, kemunculan pola-pola tertentu, dan sebagainya. Atribut lain yang cocok untuk membedakan pernyataan yang salah dan menyesatkan.

C. Crisp DM

CRISP-DM adalah metode penambangan data yang dikembangkan bersama oleh DaimlerChrysler, SPSS, dan NCR. Seperti namanya, ini adalah proses netral yang dapat digunakan dengan alat dari bidang bisnis apa pun. Anda dapat menggunakannya. Sebagai sebuah metodologi, *CRISP-DM* menjelaskan fase-fase proyek, tugas-tugas yang termasuk dalam setiap fase, dan deskripsi hubungan antara tugas-tugas ini, memberikan gambaran lengkap tentang siklus hidup data mining jika dilihat sebagai model proses [10].

III. METODE PENELITIAN



Gambar 1 Tahapan Penelitian

Pada gambar 1 ini merupakan tahapan penelitian berupa.

1. Pengumpulan Data

Mengumpulkan kalimat penipuan lowongan pekerjaan di google, facebook, dan website setelah itu data tersebut akan dikumpulkan ke dalam klasifikasi di jupyter notebook.

2. Data Understanding

Pada fase ini, melakukan pemahaman data seperti Menganalisis dan menyelidiki data teks yang dikumpulkan dan menganalisis serta menyelidiki kalimat lowongan pekerjaan palsu.

3. Data Preprocessing

Selama fase ini, data dipilih dan atribut dikelompokkan untuk digunakan dalam fase pemodelan. Menambahkan atribut unik ke data yang dikumpulkan dan melakukan pembersihan data. Mengklasifikasikan teks dengan menghilangkan atribut yang tidak penting atau mengubah huruf menjadi huruf kecil.

4. Modeling

Pada fase ini menggunakan data pelatihan yang telah diproses sebelumnya untuk melatih model Naive Bayes untuk kalimat yang diduga penipuan. Model mengklasifikasikan kalimat menjadi kalimat penipuan dan non-penipuan Untuk melatih model ini dengan menggunakan metode yang mengklasifikasikan kalimat ke dalam kategori penipuan dan bukan penipuan.

5. Evaluation

Model harus dievaluasi untuk memastikan bahwa pada pengujian model ini dapat menghasilkan suatu nilai akurasi, precision, recall, dan f1-score.

IV. HASIL DAN PEMBAHASAN

A Pengumpulan Dataset

Pada pengumpulan dataset ini, terdapat 200 kalimat lowongan pekerjaan yang didapat dari website, google, dan facebook.

Tabel 1. Kalimat lowongan pekerjaan

| No | Kalimat Lowongan Pekerjaan |
|----|---|
| 1 | Tersedia lowongan pekerjaan hanya kerja dirumah saja bisa mendapatkan penghasilan 1hari 500rb |
| 2 | PT Terkemuka kami dengan hormat mengundang spesialis untuk bergabung dengan perusahaan kami |
| 3 | Dibutuhkan lowongan pekerjaan di bagian staff dengan syarat mengirimkan foto ijazah, ktp, dan kk ke alamat email ini. |
| 4 | Lowongan pekerjaan bagian programmer syarat -Ijazah tidak diperlukan -Administrasi sebesar 10 juta rupiah |

| | |
|----|---|
| 5 | Pendaftaran ini terdapat biaya pendaftaran, untuk biayanya bisa di transfer |
| 6 | Kami perusahaan yang resmi dengan syarat ijazah minimal S1, pengalaman kerja 2tahun. |
| 7 | Dibuka lowongan pekerjaan hanya bergabung dengan kami maka kami akan menjamin kesuksesan dengan cepat |
| 8 | Tersedia lowongan pekerjaan untuk bagian teller bank, - Ijazah minimal S1, - Pengalaman 1tahun |
| 9 | Lowongan pekerjaan untuk bagian programmer, dengan syarat -Menguasai bahasa pemrograman Java -Ijazah minimal S1, -Pengalaman minimal 1tahun, |
| 10 | Dibuka lowongan pekerjaan staff dengan syarat, - ijazah minimal yaitu SMA/SMK - Administrasi sebesar 15juta - bisa masuk kerja tanpa wawancara |

Pada tabel 1 ini merupakan beberapa kalimat lowongan pekerjaan yang diambil dari dataset yang sudah dikumpulkan dari website, google, dan facebook.

B Data Understanding

Pada fase ini mempelajari tentang kalimat yang dikumpulkan, dan memahami atribut-atribut yang dimasukan dalam pengumpulan data dan fungsi dari atribut data itu sendiri. Grafik dibuat untuk no text, Isi text, dan kategori text.

C Penggunaan Atribut

Pada atribut ini menampilkan penggunaan atribut yang digunakan

Tabel 2. Penggunaan Atribut

| No | Atribut | Keterangan |
|----|---------------|--|
| 1 | No text | Nomor urut |
| 2 | Isi text | Yang berisi kalimat lowongan pekerjaan |
| 3 | Kategori text | Klasifikasi data kalimat dengan kategori penipuan atau bukan |

Pada tabel 2 ini merupakan tabel penggunaan atribut berupa nomor text sebagai nomor urut, isi text yang berisi kalimat lowongan pekerjaan, dan kategori text untuk mengklasifikasikan kalimat sebagai penipuan atau bukan penipuan. Dengan adanya penggunaan atribut ini maka hasil pengujian akan lebih lengkap dan rapi.

D. Eksplorasi Data

Eksplorasi data dilakukan di Jupyter Notebook untuk mendapatkan informasi lengkap tentang setiap atribut data Informasi ini mencakup pengetahuan tentang nama kolom, jumlah konten data, jumlah nilai yang tidak kosong,

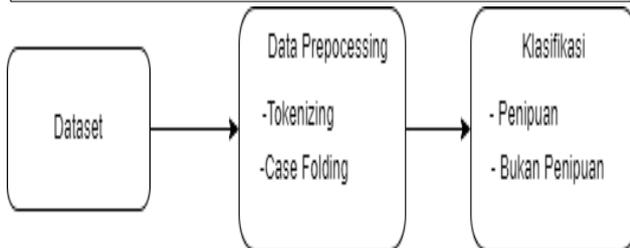


Gambar 2 Eksplorasi Data

Pada gambar 2 ini merupakan proses untuk memasukan dataset dari excel yang berisi dataset kalimat lowongan pekerjaan

E. Data Preprocessing

Selama fase ini, data dipilih dan atribut dikelompokkan untuk digunakan dalam fase pemodelan. Tambahkan atribut unik ke kalimat yang dikumpulkan dan lakukan pembersihan data. Misalnya, dapat mengubah teks menjadi huruf kecil semua, melakukan tokenisasi dan mengkategorikan kalimat.



Gambar 3 Data preprocessing

Pada gambar 3 ini proses dari data preprocessing yang berawal dari membaca dataset lalu dataset tersebut akan melalui fase tokenizing (pemisahan kata), dan case

folding (membuat semua text huruf kecil), lalu setelah itu akan melalui proses klasifikasi penipuan atau bukan penipuan

F. Tokenizing

Pada tahap tokenizing ini dengan menggunakan jupyter notebook, berikut hasil tokenizing dari kalimat lowongan pekerjaan



Gambar 4 Tokenizing



Pada gambar 4 ini adalah proses tokenizing fungsinya untuk memisahkan semua kalimat lowongan pekerjaan menjadi beberapa kata.

G. Case Folding

Fungsinya untuk merubah semua kalimat menjadi huruf kecil semua tanpa adanya huruf besar

Tabel 3 Case folding

| No | Kalimat lowongan pekerjaan | Hasil case folding |
|----|--|--|
| 1 | Tersedia lowongan pekerjaan hanya kerja dirumah saja | Tersedia lowongan pekerjaan hanya kerja dirumah saja |

| | | |
|---|--|--|
| | bisa mendapatkan penghasilan 1hari 500rb | bisa mendapatkan penghasilan 1hari 500rb |
| 2 | Pt terkemuka kami dengan hormat mengundang spesialis untuk bergabung dengan perusahaan kami | Pt terkemuka kami dengan hormat mengundang spesialis untuk bergabung dengan perusahaan kami |
| 3 | Dibutuhkan lowongan pekerjaan di bagian staff dengan syarat mengirimkan foto ijazah,ktp, dan kk ke alamat email ini. | Dibutuhkan lowongan pekerjaan di bagian staff dengan syarat mengirimkan foto ijazah,ktp, dan kk ke alamat email ini. |
| 4 | Lowongan pekerjaan bagian programmer syarat -ijazah tidak diperlukan -administrasi sebesar 10 juta rupiah | Lowongan pekerjaan bagian programmer syarat -ijazah tidak diperlukan -administrasi sebesar 10 juta rupiah |
| 5 | Pendaftaran ini terdapat biaya pendaftaran, untuk biayanya bisa di transfer | Pendaftaran ini terdapat biaya pendaftaran untuk biayanya bisa di transfer |
| 6 | Kami perusahaan yang resmi dengan syarat ijazah minimal s1, pengalaman kerja 2tahun. | Kami perusahaan yang resmi dengan syarat ijazah minimal s1, -pengalaman kerja 2tahun. |
| 7 | Dibuka lowongan pekerjaan hanya bergabung dengan kami maka kami akan menjamin kesuksesan dengan cepat | Dibuka lowongan pekerjaan hanya bergabung dengan kami maka kami akan menjamin kesuksesan dengan cepat |
| 8 | Tersedia lowongan pekerjaan untuk bagian teller bank, - ijazah minimal s1, - pengalaman 1tahun | Tersedia lowongan pekerjaan untuk bagian teller bank, - ijazah minimal s1, - pengalaman 1tahun |
| 9 | Lowongan pekerjaan untuk | Lowongan pekerjaan untuk |

| | | |
|----|--|--|
| | bagian programmer, dengan syarat -menguasai bahasa pemrograman java -ijazah minimal s1, -pengalaman minimal 1tahun, | bagian programmer, dengan syarat -menguasai bahasa pemrograman java -ijazah minimal s1, -pengalaman minimal 1tahun, |
| 10 | Dibuka lowongan pekerjaan staff dengan syarat, - ijazah minimal yaitu sma/smk - administrasi sebesar 15juta - bisa masuk kerja tanpa wawancara | Dibuka lowongan pekerjaan staff dengan syarat, - ijazah minimal yaitu sma/smk - administrasi sebesar 15juta - bisa masuk kerja tanpa wawancara |

Pada tabel 3 ini merupakan proses untuk menjadikan semua kalimat menjadi huruf kecil semua tanpa adanya huruf besar.

H Pengujian *Modeling*

Model mengklasifikasikan kalimat lowongan pekerjaan sebagai penipuan atau bukan penipuan berdasarkan fitur yang diekstraksi. Berikut hasil dari uji data kalimat penipuan.

Tabel 4 Hasil pengujian

| No | Kalimat Lowongan Pekerjaan | Kategori |
|----|--|----------------|
| 1 | Tersedia lowongan pekerjaan hanya kerja dirumah saja bisa mendapatkan penghasilan 1hari 500rb | PENIPUAN |
| 2 | Pt terkemuka kami dengan hormat mengundang spesialis untuk bergabung dengan perusahaan kami | BUKAN PENIPUAN |
| 3 | Dibutuhkan lowongan pekerjaan di bagian staff dengan syarat mengirimkan foto ijazah,ktp, dan kk ke alamat email ini. | PENIPUAN |
| 4 | Lowongan pekerjaan bagian programmer syarat -ijazah tidak diperlukan -administrasi | PENIPUAN |

| | | |
|----|--|----------------|
| | sebesar 10 juta rupiah | |
| 5 | Pendaftaran ini terdapat biaya pendaftaran untuk biayanya bisa di transfer | PENIPUAN |
| 6 | Kami perusahaan yang resmi dengan syarat -ijazah minimal s1, - pengalaman kerja 2tahun. | BUKAN PENIPUAN |
| 7 | Dibuka lowongan pekerjaan hanya bergabung dengan kami maka kami akan menjamin kesuksesan dengan cepat | PENIPUAN |
| 8 | Tersedia lowongan pekerjaan untuk bagian teller bank, - ijazah minimal s1, - pengalaman 1tahun | BUKAN PENIPUAN |
| 9 | Lowongan pekerjaan untuk bagian programmer, dengan syarat -menguasai bahasa pemrograman java -ijazah minimal s1, -pengalaman minimal 1tahun, | BUKAN PENIPUAN |
| 10 | Dibuka lowongan pekerjaan staff dengan syarat, - ijazah minimal yaitu sma/smk - administrasi sebesar 15juta - bisa masuk kerja tanpa wawancara | PENIPUAN |

Pada tabel 4 hasil pengujian tersebut jupyter notebook dapat mengklasifikasikan suatu kalimat lowongan pekerjaan tersebut masuk kedalam label penipuan atau bukan penipuan berdasarkan frasa.

Berikut ditampilkan data frasa dalam bentuk *wordcloud*.



Gambar 7. Frasa Penipuan

Pada gambar 7 ini berisi semua frasa kata kata untuk mendeteksi suatu kalimat dalam kategori penipuan yang sudah dikumpulkan dan dijadikan sebuah gambar *wordcloud*



Gambar 8. Frasa bukan penipuan

Pada gambar 8 ini berisi semua frasa kata kata untuk mendeteksi suatu kalimat dalam kategori bukan penipuan yang sudah dikumpulkan dan dijadikan sebuah gambar *wordcloud*

I. Pembahasan

Dari hasil semua pengujian yang sudah dilakukan dengan menggunakan *jupyter notebook* berikut hasil dari evaluasi pengujian tersebut

Tabel 5 *Confusion matrix*

| Kategori | Prediksi salah | Prediksi benar |
|----------------|-------------------------|-------------------------|
| Bukan Penipuan | False Positive (TP) = 0 | True Positive (TP) = 86 |

| | | |
|----------|--------------------------|-------------------------|
| Penipuan | False Negative (FN) = 18 | True Negative (TN) = 56 |
|----------|--------------------------|-------------------------|

Pada tabel 5 ini terdapat hasil sebagai berikut.

1. Positif salah (FP): Total data (bukan penipuan) yang salah diklasifikasikan = 0
2. Negatif benar (TN): Total data diklasifikasikan dengan benar (bukan penipuan) = 86
3. Negatif palsu (FN): Total data (penipuan) diklasifikasikan dengan salah= 18
4. Positif benar (TP): Total data diklasifikasikan dengan benar (penipuan) = 56

Tabel 6 Skor evaluasi

| Kategori | Precision | Recall | F1 Score | Support |
|----------------|-----------|--------|----------|---------|
| Bukan Penipuan | 0.83 | 1.00 | 0.91 | 86 |
| Penipuan | 1.00 | 0.76 | 0.86 | 74 |
| Accuracy | | | 0.89 | 160 |
| Macro avg | 0.91 | 0.88 | 0.88 | 160 |
| Weighted avg | 0.91 | 0.89 | 0.89 | 160 |

Pada tabel 6 ini dari hasil pengujian dengan range 80:20 data uji 80% dan data latih 20% kesimpulan sebagai berikut.

1. Kategori "BUKAN PENIPUAN":
 - Hasil Precision adalah 83% (model memprediksi 83% dari kasus bukan penipuan dengan benar)
 - Hasil Recall adalah 100% (model berhasil mengidentifikasi 100% dari semua kasus bukan penipuan)
 - Hasil F1-score adalah 91% keseimbangan antara hasil *precision* dan *recall*
 - Hasil support jumlah dalam kategori bukan penipu total 86
2. Kategori "PENIPUAN":
 - Hasil Precision adalah 100% (model memprediksi 100% dari kasus bukan penipuan dengan benar)
 - Hasil Recall adalah 76% (model berhasil mengidentifikasi 76% dari semua kasus penipuan)
 - Hasil F1-score adalah 86% keseimbangan antara hasil *precision* dan *recall*
 - Hasil support jumlah dalam kategori penipuan total 74
3. Total akurasi keseluruhan sebesar 89%
4. Macro avg:

-Hasil Precision adalah 100% (model memprediksi 100% dari kasus bukan penipuan dengan benar)

-Hasil Recall adalah 76% (model berhasil mengidentifikasi 76% dari semua kasus penipuan)

-Hasil F1-score adalah 86% keseimbangan antara hasil *precision* dan *recall*

5. Weighted avg:

-Hasil Precision adalah 100% (model memprediksi 100% dari kasus bukan penipuan dengan benar)

-Hasil Recall adalah 76% (model berhasil mengidentifikasi 76% dari semua kasus penipuan)

-Hasil F1-score adalah 86% keseimbangan antara hasil *precision* dan *recall*

6. Total data uji keseluruhan yaitu 160 data

Tabel 7. Total hasil data

| Bukan Penipuan | Penipuan |
|----------------|----------|
| 106 Data | 94 Data |

Pada tabel 7 ini menampilkan total hasil data uji dan latih untuk kategori bukan penipuan total 106 data, dan kategori penipuan total 94 data.

V KESIMPULAN DAN SARAN

Kesimpulan hasil penelitian dengan menggunakan algoritma klasifikasi *Naïve Bayes* untuk mendeteksi penipuan lowongan pekerjaan dapat disimpulkan bahwa dalam melakukan pendeteksian ini dengan menggunakan aplikasi *jupyter notebook* bahasa pemrograman *Python* mendeteksi kalimat berdasarkan kata kata frasa penipuan dan bukan penipuan yang sudah dimasukan kedalam kode program di *jupyter notebook* sehingga jika kalimat tersebut terdapat kata kata frasa penipuan maka kalimat tersebut akan otomatis masuk ke label penipuan. Dan hasil dari penelitian ini menghasilkan data bukan penipuan sebesar 106 data, dan penipuan sebesar 94 data, total akurasi keseluruhan sebesar 89%, Untuk label bukan penipuan menghasilkan *precision* 83%, *recall* 100%, dan *f1-score* 91%. Untuk label penipuan menghasilkan *precision* 100%, *recall* 76%, dan *f1-score* sebesar 86%. Dari total hasil tersebut hasilnya cukup tinggi jika dibandingkan dengan penelitian lain maka penelitian ini hasilnya lebih tinggi, lebih akurat, dan algoritma *Naïve Bayes* sangat cocok untuk penelitian ini.

Saran untuk penelitian ini adalah dengan sedikitnya jumlah dataset yang hanya 200 dataset saja sehingga untuk penelitian selanjutnya diharapkan dapat menggunakan dataset kalimat lowongan pekerjaan dengan lebih banyak lagi. Dan hasil akurasi, *precision*, *recall*, dan *f1-score* yang lebih besar lagi.

DAFTAR PUSTAKA

- [1] M. M. Fajar, A. R. Putri, and K. F. H. Holle, "Perbandingan Metode Klasifikasi Data Mining Untuk Deteksi Keaslian Lowongan Pekerjaan di Medsos," *J. Ilm. Inform.*, vol. 7, no. 1, pp. 41–48, 2022, doi: 10.35316/jimi.v7i1.41-48.
- [2] H. Sabita, F. Fitria, and R. Herwanto, "Analisa Dan Prediksi Iklan Lowongan Kerja Palsu Dengan Metode Natural Language Programing Dan Machine Learning," *J. Inform.*, vol. 21, no. 1, pp. 14–22, 2021, doi: 10.30873/ji.v21i1.2865.
- [3] U. B. Jaya, P. Studi, T. Informatika, F. I. Komputer, and U. B. Jaya, "PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)," vol. 3, no. 2, pp. 178–194, 2019.
- [4] M. R. Hunafa and A. Hermawan, "Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Imbalance Class Dataset Penyakit Diabetes," *Media Online*, vol. 4, no. 3, pp. 1551–1561, 2023, doi: 10.30865/klik.v4i3.1486.
- [5] M. Iqbal, A. D. Wiranata, R. Suwito, R. F. Ananda, U. Muhammadiyah, and P. Hamka, "Perbandingan Algoritma Naïve Bayes , KNN , dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter," vol. 4, no. 3, pp. 1799–1807, 2023, doi: 10.30865/klik.v4i3.1402.
- [6] I. Artikel, "STUDI KOMPARASI ALGORITMA NAÏVE BAYES DAN K-NN UNTUK KLASIFIKASI PENERIMAAN BEASISWA DI MI AL – ISLAMIYAH," vol. 2, no. 4, pp. 172–180, 2023.
- [7] F. Narti, "JTIM : Jurnal Teknologi Informasi dan Multimedia Perbandingan Algoritma C4 . 5 dan Naive Bayes dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," vol. 4, no. 1, 2022.
- [8] I. Ahmad, S. Samsugi, and Y. Irawan, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/107>
- [9] Rayuwati, Husna Gemasih, and Irma Nizar, "IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [10] R. Novita, A. N. Khomarudin, R. Aulia, J. Jamaluddin, A. Yuditihwa, and A. Ayuri, "Penerapan Algoritma K-Means dan Analisisnya untuk Menentukan Kebijakan Strategis Penyelesaian Studi Mahasiswa," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 22, no. 2, p. 401, 2023, doi: 10.53513/jis.v22i2.8461.