



## DETEKSI LOWONGAN KERJA FIKTIF MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN TF-IDF PADA PLATFORM REKRUTMEN KERJA ONLINE DI INDONESIA

Syarif Aminul Khoiri<sup>1)</sup>, Abdus Samad<sup>2)</sup>, Muhammad Fauzen Adiman<sup>3)</sup>

<sup>1</sup> Teknik Informatika, Universitas Ibrahimy

<sup>2</sup> Teknik Informatika, Universitas Ibrahimy

<sup>3</sup> Ilmu Komputer, Universitas Ibrahimy

email: <sup>1</sup> syarifak@ibrahimy.ac.id, <sup>2</sup> saintek\_somad@gmail.com, <sup>3</sup> fadhim16@gmail.com,

### ARTICLE INFO

#### Article History:

Received : 30 November 2026

Accepted : 10 Mei 2026

Published : 2 Juni 2026

#### Keywords:

Deteksi Lowongan Kerja,  
Lowongan Kerja Fiktif,  
Naïve Bayes,  
TF-IDF,  
Machine Learning,  
Rekrutmen Online,  
Text Mining.

#### IEEE style in citing this article:

S. A. Khoiri, A. Samad, M. F. Adiman, "Deteksi Lowongan Kerja Fiktif Menggunakan Algoritma Naïve Bayes Dan TF-IDF Pada Platform Rekrutmen Kerja Online di Indonesia", *Jurnal.Ilmiah.Informatika*, vol. 11, no. 1, pp. 11-xx, Bulan. Tahun.

### ABSTRACT

Perkembangan teknologi informasi membuat proses rekrutmen kerja semakin mudah dilakukan secara online. Namun, kemudahan ini juga membawa masalah baru, yaitu semakin banyaknya lowongan kerja fiktif yang beredar di berbagai platform rekrutmen di Indonesia. Lowongan seperti ini sering digunakan untuk penipuan, pencurian data pribadi, atau meminta sejumlah uang kepada pelamar. Banyak pencari kerja yang akhirnya menjadi korban karena sulit membedakan mana lowongan yang benar dan mana yang palsu. Berdasarkan masalah tersebut, penelitian ini dilakukan untuk mencoba membuat sebuah model deteksi lowongan kerja fiktif dengan memanfaatkan algoritma Naïve Bayes dan metode pembobotan teks TF-IDF.

Penelitian ini menggunakan data teks dari contoh-contoh lowongan kerja yang dikumpulkan dari beberapa platform online. Setiap data kemudian diberi label sebagai "asli" atau "fiktif" berdasarkan ciri-ciri umum yang biasanya muncul pada kasus penipuan. Setelah proses pembersihan data dan ekstraksi fitur menggunakan TF-IDF, data dilatih menggunakan algoritma Naïve Bayes untuk melihat seberapa baik model dapat mengenali pola pada lowongan fiktif. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa kombinasi Naïve Bayes dan TF-IDF mampu memberikan performa yang cukup baik dalam mendeteksi lowongan kerja fiktif.

Model yang dibangun dapat mengidentifikasi kata-kata atau pola tertentu yang sering muncul pada lowongan palsu, seperti tawaran gaji tidak wajar, proses seleksi instan, atau adanya permintaan biaya. Temuan ini diharapkan dapat membantu pencari kerja agar lebih waspada saat melamar pekerjaan secara online. Selain itu, penelitian ini juga dapat menjadi dasar untuk pengembangan sistem deteksi otomatis yang bisa digunakan oleh platform rekrutmen kerja di Indonesia. Penelitian ini masih memiliki keterbatasan dalam jumlah

---

*data, namun dapat dikembangkan lebih lanjut dengan dataset yang lebih besar dan lebih beragam.*

---

© 2026 Jurnal Ilmiah Informatika (Scientific Informatics Journal) with CC BY NC licence

## 1. PENDAHULUAN

Fenomena lowongan kerja fiktif semakin sering menjadi pembahasan di berbagai media karena banyaknya korban yang tertipu ketika mencari pekerjaan secara daring. Kemudahan akses internet dan perkembangan platform rekrutmen online membuat masyarakat dapat dengan cepat memperoleh informasi lowongan. Namun, kondisi ini sekaligus membuka peluang bagi pihak-pihak tidak bertanggung jawab untuk menyebarkan iklan lowongan palsu yang terlihat meyakinkan. Banyak kasus yang menunjukkan bahwa pencari kerja kerap diminta membayar biaya administrasi, menyerahkan data pribadi, bahkan mengikuti proses rekrutmen palsu yang berujung pada kerugian material maupun psikologis. Penelitian sebelumnya juga menjelaskan bahwa maraknya penipuan lowongan kerja terjadi karena tingginya persaingan kerja dan kurangnya literasi digital di kalangan pencari kerja. [1]

Selain faktor literasi digital, penyebaran lowongan palsu juga dipengaruhi oleh penggunaan media sosial dan aplikasi pesan instan yang semakin sulit dikontrol. Banyak iklan yang “terlihat resmi” padahal tidak memiliki kejelasan perusahaan, alamat, maupun prosedur perekrutan yang benar. Pelaku sering memanfaatkan situasi ekonomi masyarakat dan tingginya kebutuhan pekerjaan untuk memberikan tawaran kerja yang terdengar sangat menggiurkan. Beberapa penelitian di Indonesia mencatat bahwa modus penipuan yang paling umum adalah mengatasnamakan perusahaan besar,

meminta biaya pendaftaran, serta menyediakan informasi yang tidak sesuai dengan standar rekrutmen resmi. [2]

Walaupun pemerintah dan berbagai lembaga telah menyediakan sarana pelaporan, kenyataannya proses identifikasi lowongan palsu masih bergantung pada laporan manual dari korban atau masyarakat. Cara ini tentu tidak cukup cepat dan tidak dapat mengimbangi jumlah iklan yang terus bermunculan setiap hari. Di sisi lain, penegakan hukum terhadap pelaku penipuan lowongan daring juga menghadapi kendala seperti bukti digital yang mudah dihapus dan identitas pelaku yang sulit dilacak. Penelitian hukum di Indonesia menyebutkan bahwa aparat sering kesulitan membuktikan unsur penipuan karena tidak semua kasus dilaporkan dan minimnya pemahaman masyarakat tentang penipuan digital. [3]

Melihat situasi tersebut, teknologi dapat digunakan sebagai salah satu cara untuk membantu proses deteksi lebih awal terhadap iklan lowongan fiktif. Teknik analisis teks seperti *text mining* dapat mempelajari pola bahasa yang biasa muncul pada lowongan palsu dan membedakannya dari lowongan asli. Beberapa penelitian telah mencoba menggunakan metode berbasis machine learning untuk mengklasifikasikan iklan kerja, dan hasilnya cukup menjanjikan. Misalnya, studi mengenai deteksi lowongan palsu menggunakan model deep learning menunjukkan bahwa teks lowongan memang memiliki ciri tertentu yang dapat dipelajari oleh sistem otomatis. [4]

Namun, penelitian serupa dalam konteks Indonesia masih sangat terbatas, terutama yang membahas penggunaan metode seperti Naive Bayes dan teknik representasi teks TF-IDF. Padahal, metode tersebut dikenal cukup efektif untuk klasifikasi teks dan tidak membutuhkan sumber daya komputasi yang besar. Selain itu, kebanyakan penelitian yang ada lebih fokus pada studi hukum atau analisis kasus, bukan pada pengembangan sistem pendeteksian otomatis. Penelitian tentang pemalsuan lowongan kerja di Indonesia juga pada umumnya menyoroti aspek dampak dan modus, tetapi belum banyak yang mencoba membuat model deteksi berbasis data teks.[5] [6]

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk membangun model deteksi lowongan kerja fiktif menggunakan algoritma Naive Bayes dengan pendekatan TF-IDF pada teks lowongan kerja yang ditemukan di platform rekrutmen online di Indonesia. Dengan menggunakan metode yang relatif sederhana namun terbukti efektif dalam banyak penelitian klasifikasi teks, penelitian ini diharapkan dapat memberikan solusi awal untuk membantu pencari kerja mengenali lowongan yang mencurigakan. Selain itu, penelitian ini diharapkan dapat menjadi kontribusi penting bagi pengembangan teknologi deteksi penipuan digital yang lebih maju di masa mendatang.

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Penelitian ini menggunakan desain penelitian kuantitatif dan fokusnya adalah mengolah teks lowongan kerja dan mencoba membuat model yang bisa membedakan mana lowongan asli dan mana yang fiktif. Metode yang dipakai adalah pengolahan teks dengan TF-IDF dan klasifikasi menggunakan algoritma

Naïve Bayes karena keduanya cukup mudah digunakan untuk data teks berbahasa Indonesia. [7], [8]

Data dikumpulkan secara manual dari beberapa platform rekrutmen online seperti Jobstreet, Kalibrr, Glints, serta beberapa contoh lowongan kerja yang beredar di media sosial. Data yang terkumpul kemudian diperiksa satu per satu dan diberi label “asli” atau “fiktif” berdasarkan ciri-ciri umum yang ditemukan pada kasus lowongan palsu. Setelah itu, data disiapkan untuk tahap pemrosesan dan pelatihan model.

### 2.2 Pelabelan Data

Pada tahap ini, setiap lowongan kerja yang sudah dikumpulkan diberi label sesuai kategorinya, yaitu “asli” atau “fiktif”. Proses pelabelan dilakukan secara manual dengan melihat isi deskripsi lowongan dan mencocokkannya dengan ciri-ciri umum yang biasanya muncul pada kasus lowongan palsu. Beberapa tanda yang digunakan antara lain adanya permintaan uang, tawaran gaji yang tidak wajar, identitas perusahaan yang tidak jelas, atau proses seleksi yang terlalu cepat. [9]

Pelabelan dilakukan secara hati-hati agar data yang digunakan untuk melatih model benar-benar sesuai dengan kondisinya. Lowongan yang meragukan juga diperiksa melalui sumber tambahan seperti situs resmi perusahaan atau laporan dari pengguna internet. Setelah semua data selesai diberi label, barulah data tersebut dipakai untuk tahap pemrosesan dan pelatihan model.

### 2.3 Preprocessing Teks

Setelah data diberi label, langkah berikutnya adalah melakukan preprocessing atau pembersihan teks. Tahap ini penting supaya isi lowongan

kerja bisa dibaca oleh sistem secara lebih rapi dan seragam. Beberapa langkah yang dilakukan antara lain mengubah seluruh huruf menjadi huruf kecil, menghapus tanda baca, angka, dan simbol yang tidak diperlukan, serta membersihkan kata-kata yang tidak punya makna penting dalam proses analisis.[10]

Selain itu, teks juga diproses dengan tahap tokenisasi untuk memecah kalimat menjadi kata-kata. Setelah itu dilakukan penghapusan *stopword* atau kata-kata umum seperti “dan”, “yang”, “di”, dan lain-lain. Tahap terakhir adalah *stemming* atau mengubah kata menjadi bentuk dasarnya, misalnya “mengirimkan” menjadi “ kirim”. Semua langkah ini dilakukan supaya teks yang masuk ke sistem lebih sederhana dan mudah diolah pada tahap berikutnya.

2.3.1 Cleaning

Pada tahap cleaning, teks lowongan kerja dibersihkan dari bagian-bagian yang tidak diperlukan. Beberapa lowongan memiliki banyak simbol, angka, tanda baca berlebihan, atau sisipan seperti emoticon dan link yang tidak ada hubungan dengan isi lowongan. Semua elemen yang tidak penting ini dihapus supaya teks lebih rapi dan mudah dibaca oleh sistem. Selain itu, beberapa lowongan juga mengandung karakter HTML atau format penulisan aneh yang harus dibersihkan terlebih dahulu. Tujuan dari cleaning ini adalah agar teks yang digunakan benar-benar hanya berisi kalimat yang relevan dan siap masuk ke proses pengolahan berikutnya. [11]

Tabel 1. Contoh Cleaning

Sebelum	Sesudah
“LOWONGAN KERJA RESMI!!! Gaji 7-15jt/bln!□ Tanpa pengalaman!! Daftar	“lowongan kerja resmi gaji 7 15 jt bln tanpa pengalaman daftar lewat link

lewat link ini: ini”  
<https://bit.ly/karircepat123>

Pada contoh di atas, teks asli mengandung banyak simbol, huruf kapital berlebihan, serta link yang tidak diperlukan untuk analisis. Pada tahap cleaning, semua elemen tersebut dihapus. Hasilnya menjadi teks yang lebih bersih dan lebih mudah diproses oleh sistem selanjutnya.

2.3.2 Case Folding

Pada tahap case folding, semua huruf pada teks lowongan kerja diubah menjadi huruf kecil. Tujuan langkah ini yaitu supaya sistem tidak menganggap huruf besar dan huruf kecil sebagai bentuk kata yang berbeda. Misalnya kata “Kerja”, “KERJA”, dan “kerja” semuanya dianggap sama. Dengan cara ini, proses pengolahan data menjadi lebih rapi dan konsisten. [12]

Tabel 2. Contoh Case Folding

Sebelum	Sesudah
“LOWONGAN resmi tanpa Pengalaman”	“lowongan kerja resmi tanpa pengalaman”

Adapun pada contoh di atas, teks asli masih memiliki campuran huruf besar dan huruf kecil. Setelah dilakukan case folding, seluruh huruf diubah menjadi huruf kecil agar bentuk katanya seragam dan tidak dianggap berbeda oleh sistem.

2.3.3 Tokenizing

Pada tahap tokenizing, teks yang sudah dibersihkan kemudian dipisahkan menjadi potongan-potongan kata. Proses ini dilakukan supaya sistem bisa mengenali setiap kata secara terpisah saat melakukan perhitungan dan analisis. Misalnya satu kalimat panjang akan dipecah menjadi beberapa kata yang berdiri sendiri. Dengan cara ini, data teks

menjadi lebih mudah diproses pada tahap selanjutnya. [13]

Tabel 3. Contoh Tokenizing

Sebelum	Sesudah
"lowongan kerja resmi tanpa pengalaman"	["lowongan", "kerja", "resmi", "tanpa", "pengalaman"]

### 2.3.4 Stopword Removal

Pada tahap stopword removal, beberapa kata umum yang terlalu sering muncul dalam teks dihapus dari data. Kata-kata seperti "dan", "yang", "di", "ke", dan sejenisnya biasanya tidak terlalu berpengaruh terhadap makna atau kategori lowongan kerja, sehingga lebih baik dibuang saja. Dengan menghapus stopword, sistem bisa lebih fokus pada kata-kata penting yang benar-benar menentukan apakah sebuah lowongan itu asli atau fiktif. [14]

Tabel 4. Contoh Stopword Removal

Sebelum	Sesudah
["lowongan", "kerja", "resmi", "tanpa", "pengalaman", "di", "wilayah", "jakarta"]	["lowongan", "kerja", "resmi", "tanpa", "pengalaman", "wilayah", "jakarta"]

Pada contoh di atas, kata "di" dihapus karena termasuk stopword dan tidak memberikan informasi penting untuk proses klasifikasi. Dengan cara ini, daftar kata menjadi lebih ringkas dan hanya berisi kata-kata yang dianggap penting.

### 2.3.5 Stemming

Pada tahap stemming, setiap kata diubah menjadi bentuk dasarnya. Tujuannya supaya kata-kata yang sebenarnya memiliki arti sama tidak dianggap berbeda hanya karena bentuknya berubah. Misalnya kata "mengirimkan",

"mengirim", dan "dikirim" akan diubah menjadi "kirim". Dengan cara ini, jumlah kata yang harus diproses oleh sistem menjadi lebih sedikit dan lebih mudah untuk dianalisis. Stemming juga membantu model mengenali pola yang lebih jelas pada lowongan kerja asli maupun fiktif.[15]

Tabel 5. Contoh Hasil Stemming

Sebelum	Sesudah
["mengirimkan", "pencarian", "pekerjaan", "dibutuhkan", "pelamar"]	["kirim", "cari", "kerja", "butuh", "lamar"]

Pada contoh di atas, setiap kata diubah menjadi bentuk dasarnya agar lebih sederhana dan konsisten. Misalnya "mengirimkan" menjadi "kirim" dan "pencarian" menjadi "cari". Dengan langkah ini, teks menjadi lebih mudah diproses dan dianalisis oleh sistem.

### 2.4 Ekstraksi Fitur dengan TF-IDF

Pada tahap ini, setiap kata yang sudah melalui proses preprocessing diubah menjadi nilai angka menggunakan metode TF-IDF. Cara kerja TF-IDF yaitu menghitung seberapa sering sebuah kata muncul pada satu dokumen dan seberapa jarang kata tersebut muncul pada dokumen lain. Kata yang sering muncul di satu lowongan tetapi tidak terlalu sering muncul di lowongan lain biasanya dianggap lebih penting. Dengan TF-IDF, teks lowongan kerja yang awalnya hanya kumpulan kata berubah menjadi angka-angka yang bisa dibaca oleh algoritma Naïve Bayes. Nilai-nilai ini nantinya digunakan sebagai fitur utama dalam proses klasifikasi lowongan kerja asli dan fiktif.[16], [17]

TF dihitung dari seberapa sering kata muncul dalam dokumen, sedangkan IDF

melihat apakah kata tersebut muncul juga di banyak dokumen lain atau tidak. Rumus yang digunakan dapat dilihat pada persamaan berikut:

$$TF = \frac{\text{Jumlah Kemunculan Data}}{\text{Total Kata dalam Dokumen}}$$

$$IDF = \log\left(\frac{N}{df}\right)$$

$$TF-IDF = TF \times IDF$$

Nilai TF-IDF yang kecil biasanya menunjukkan kata yang umum, sedangkan kata yang jarang muncul dan dianggap lebih unik akan punya nilai lebih besar. Nilai-nilai ini nantinya dipakai sebagai fitur utama untuk membantu model Naïve Bayes dalam membedakan lowongan kerja asli dan fiktif.

Tabel 6. Perhitungan TF-IDF pada beberapa kata

Kata	TF (Frekuensi di Dokumen)	IDF (Kebalikan Frekuensi di Semua Dokumen)	Nilai TF-IDF
“gaji”	3	0,78	2.34
“resmi”	1	1.20	1.20
“pengalaman”	1	0.95	1.90
“deposit”	2	2.10	2.10

Contoh di atas menunjukkan bagaimana beberapa kata pada teks lowongan kerja diubah menjadi angka menggunakan metode TF-IDF. Kata “deposit” memiliki nilai IDF yang tinggi karena jarang muncul di dokumen lain, sehingga nilainya menjadi cukup besar. Ini penting karena kata semacam “deposit” sering muncul pada lowongan fiktif, sehingga fitur seperti ini dapat membantu model mengenali pola penipuan.

### 2.5 Algoritma Naïve Bayes

Pada penelitian ini, algoritma yang digunakan untuk melakukan klasifikasi adalah Naïve Bayes. Algoritma ini dipilih karena cara kerjanya cukup sederhana dan banyak digunakan untuk mengolah data teks. Secara umum, Naïve Bayes menghitung peluang suatu dokumen termasuk ke dalam kelas tertentu berdasarkan kata-kata yang muncul di

dalam dokumen tersebut. Disebut naïve karena algoritma ini menganggap setiap kata berdiri sendiri dan tidak saling mempengaruhi, walaupun pada kenyataannya tidak selalu begitu. [18], [19]

Rumus dasar naïve bayes sebagai berikut

$$P(C|X) = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Keterangan :

**P(C|X)** = peluang suatu dokumen X masuk ke kelas C

**P(X|C)** = peluang kata-kata dalam dokumen muncul pada kelas tersebut

**P(C)** = peluang awal dari kelas

**P(X)** = peluang dokumen secara keseluruhan

Dalam proses klasifikasi lowongan kerja, algoritma ini akan menghitung kata-kata yang ada pada lowongan dan

membandingkannya dengan pola yang sudah dipelajari dari data latih. Hasil akhirnya adalah prediksi apakah lowongan tersebut termasuk kategori “asli” atau “fiktif”. Selain mudah digunakan, Naïve Bayes juga cukup cepat dalam memproses data yang jumlahnya banyak, sehingga cocok untuk penelitian yang menggunakan teks sebagai bahan utama.

## 2.6 Pembagian Data

Setelah data selesai diproses dan siap digunakan, langkah selanjutnya adalah membagi data menjadi dua bagian, yaitu data latih dan data uji. Pembagian ini dilakukan supaya model bisa belajar dari sebagian data, lalu diuji pada data lain yang belum pernah dilihat sebelumnya. Pada penelitian ini, pembagian dilakukan dengan perbandingan 80% untuk data latih dan 20% untuk data uji. Cara ini cukup umum digunakan, terutama untuk penelitian yang menggunakan jumlah data terbatas.

Setelah dibagi, data latih digunakan untuk melatih model Naïve Bayes dengan fitur TF-IDF yang sudah dihitung sebelumnya. Pada tahap pelatihan ini, model akan mempelajari pola-pola kata yang sering muncul pada lowongan kerja asli dan lowongan kerja fiktif. Setelah model selesai dilatih, data uji digunakan untuk melihat seberapa baik model dapat memprediksi kategori lowongan kerja yang belum pernah digunakan saat pelatihan. Hasil pengujian ini nantinya digunakan sebagai dasar untuk mengukur akurasi dan performa model secara keseluruhan

## 3 HASIL DAN PEMBAHASAN

### 3.1 Hasil Preprocessing

Pada tahap preprocessing, data lowongan kerja yang sudah dikumpulkan kemudian dibersihkan agar lebih rapi dan siap digunakan pada proses selanjutnya.

Beberapa contoh hasil preprocessing dapat dilihat pada tabel berikut. Contoh ini menunjukkan bagaimana teks yang awalnya masih bercampur dengan tanda baca, huruf besar, emoticon, dan kata-kata yang tidak penting diubah menjadi teks yang lebih bersih. Tahap ini penting supaya model bisa membaca isi lowongan dengan lebih jelas dan tidak bingung dengan bagian-bagian yang tidak diperlukan.

Tabel 7. Hasil Preprocessing Teks

Sebelum	Sesudah
“LOWONGAN KERJA RESMI!!! Gaji 7-15jt/bln!! □ Tanpa pengalaman!! Daftar lewat link ini: https://bit.ly/karircep at123 ”	“lowongan kerja resmi gaji 7 15 jt bln tanpa pengalaman daftar lewat link ini”
“KERJA SANTAI DIBAYAR HARIAN!!! Cuma input data 1-2 jam per hari. Gaji sampai 10 juta per bulan. Untuk aktivasi akun silakan transfer 300rb.”	“kerja santai dibayar harian cuma input data 1 2 jam per hari gaji sampai 10 juta per bulan untuk aktivasi akun silakan transfer 300rb”

Dari tabel tersebut terlihat bahwa teks yang awalnya banyak mengandung simbol atau variasi penulisan, seperti huruf besar dan emoticon, sudah berhasil dirapikan melalui proses cleaning dan case folding. Setelah itu teks juga sudah dipisahkan menjadi kata-kata yang lebih sederhana melalui tokenizing, dan kata-kata umum yang tidak terlalu penting juga sudah dibuang pada tahap stopword removal. Tahap preprocessing ini membuat teks lowongan kerja menjadi lebih mudah diproses oleh metode TF-IDF dan algoritma Naïve Bayes pada langkah berikutnya.

### 3.2 Hasil Ekstraksi Fitur TF-IDF

Setelah melalui tahap preprocessing, setiap teks lowongan kerja kemudian diubah menjadi bentuk angka menggunakan metode TF-IDF. Dari proses ini, setiap kata di dalam dokumen memiliki nilai tertentu yang menunjukkan seberapa penting kata tersebut di dalam kumpulan data. Kata yang sering muncul di satu jenis lowongan, tetapi jarang muncul di lowongan lain, biasanya akan memiliki nilai TF-IDF yang lebih tinggi dan dianggap lebih berpengaruh dalam proses klasifikasi.

Secara umum, dari hasil perhitungan TF-IDF terlihat bahwa beberapa kata yang sering muncul pada lowongan kerja fiktif memiliki nilai yang cukup tinggi, seperti kata "deposit", "biaya", dan "transfer". Sementara itu, pada lowongan kerja asli, kata-kata seperti "resmi", "bpjs", dan "kontrak" cenderung lebih menonjol. Contoh sebagian hasil perhitungan TF-IDF ditunjukkan pada tabel berikut.

Tabel 8. Hasil Nilai TF-IDF Beberapa Kata pada Dataset

Kata	Kategori	Nilai TF-IDF
deposit	Fiktif	2,10
biaya	Fiktif	1,95
transfer	Fiktif	1,88
resmi	Asli	1,20
bpjs	Asli	1,15
Kontrak	Asli	1,05

Dari tabel tersebut terlihat bahwa kata-kata yang berhubungan dengan permintaan uang atau setoran awal cenderung memiliki nilai TF-IDF yang tinggi pada lowongan fiktif. Sebaliknya, kata-kata yang berkaitan dengan hak kerja dan fasilitas, seperti "bpjs" dan "kontrak", lebih sering muncul pada lowongan asli. Pola perbedaan ini membantu model

Naïve Bayes dalam membedakan kedua jenis lowongan kerja pada tahap klasifikasi.

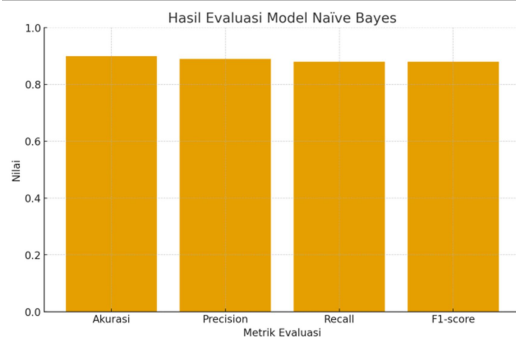
### 3.3 Hasil Pengujian

Setelah model Naïve Bayes dilatih menggunakan data latih, langkah berikutnya adalah menguji model dengan data uji. Pada penelitian ini, data dibagi menjadi 80% untuk data latih dan 20% untuk data uji. Dari proses pengujian tersebut, diperoleh beberapa nilai evaluasi yang digunakan untuk melihat seberapa baik model dalam membedakan lowongan kerja asli dan lowongan kerja fiktif.

Tabel 9. Hasil evaluasi model Naïve bayes

Metrik	Nilai
Akurasi	0,90
Precision	0,89
Recall	0,88
F1-Score	0,88

Nilai akurasi 0,90 menunjukkan bahwa sebagian besar data uji dapat diprediksi dengan benar oleh model. Precision yang mendekati 0,89 berarti bahwa dari semua lowongan yang diprediksi sebagai fiktif, sebagian besar memang benar-benar fiktif. Sementara itu, recall sebesar 0,88 menunjukkan bahwa model cukup baik dalam menangkap sebagian besar lowongan fiktif yang ada di data uji. Nilai F1-score yang seimbang juga menunjukkan bahwa model tidak hanya akurat, tetapi juga cukup stabil antara ketepatan dan kemampuan dalam mendeteksi lowongan fiktif.



Gambar 1. Grafik Hasil Evaluasi Model Naïve Bayes

Gambar diatas menunjukkan perbandingan nilai akurasi, precision, recall, dan F1-score dari model Naïve Bayes yang digunakan dalam penelitian ini. Terlihat bahwa semua metrik memiliki nilai yang cukup tinggi dan saling berdekatan, yaitu akurasi sebesar 0,90, precision sebesar 0,89, recall sebesar 0,88, dan F1-score sebesar 0,88. Hal ini menunjukkan bahwa model tidak hanya mampu melakukan prediksi dengan cukup tepat, tetapi juga cukup seimbang dalam mendeteksi lowongan kerja fiktif tanpa terlalu banyak salah mengenali lowongan kerja asli.

### 3.4 Confusion Matrix

Selain melihat nilai akurasi, precision, recall, dan F1-score, hasil pengujian model juga dapat dilihat melalui confusion matrix. Confusion matrix digunakan untuk mengetahui berapa banyak data yang diprediksi dengan benar dan berapa yang masih salah untuk masing-masing kelas. Dalam penelitian ini, kelas yang digunakan adalah "asli" dan "fiktif". Contoh hasil confusion matrix dari model Naïve Bayes ditunjukkan pada Tabel 10.

Tabel 10. Confusion Matrix

	Prediksi Negatif	Prediksi Fiktif
Lowongan Asli	9	1

Lowongan Asli	9	1
Lowongan Fiktif	1	0

Dari Tabel 10 terlihat bahwa model berhasil memprediksi dengan benar 9 data lowongan kerja asli dan 9 data lowongan kerja fiktif. Masih ada 1 data asli yang salah diprediksi sebagai fiktif, dan 1 data fiktif yang salah diprediksi sebagai asli. Secara keseluruhan, ada 18 data yang benar dan 2 data yang salah, sehingga hasil ini sejalan dengan nilai akurasi yang diperoleh sebelumnya yang berada di kisaran 0,90. Confusion matrix ini menunjukkan bahwa model cukup seimbang dalam mengenali kedua jenis lowongan, meskipun masih ada beberapa kesalahan yang perlu diperhatikan.

### 3.5 Pembahasan Hasil

Berdasarkan hasil pengujian yang sudah dilakukan, model Naïve Bayes dengan fitur TF-IDF sebenarnya sudah cukup baik dalam membedakan lowongan kerja asli dan lowongan kerja fiktif. Nilai akurasi yang diperoleh berada di kisaran 0,90, yang artinya sebagian besar data uji berhasil diprediksi dengan benar. Nilai precision, recall, dan F1-score yang juga mendekati angka 0,90 menunjukkan bahwa performa model tidak hanya baik dari satu sisi saja, tetapi cukup seimbang antara ketepatan dan kemampuan dalam mendeteksi lowongan fiktif.

Jika dilihat dari confusion matrix, model mampu mengklasifikasikan dengan benar sebagian besar lowongan kerja, baik yang asli maupun yang fiktif. Terdapat 9 data asli yang terdeteksi dengan benar sebagai asli dan 9 data fiktif yang terdeteksi sebagai fiktif. Namun, masih ada 1 data asli yang salah diprediksi sebagai fiktif dan 1 data fiktif yang salah diprediksi sebagai asli. Kesalahan ini menunjukkan bahwa masih ada beberapa lowongan

yang bahasanya mirip, sehingga membuat model agak ragu dalam membedakannya. Misalnya, bisa saja ada lowongan asli yang menggunakan kata-kata promosi yang cukup berlebihan, atau lowongan fiktif yang ditulis dengan gaya yang terlihat meyakinkan.

Hasil TF-IDF sebelumnya juga mendukung pola ini. Kata-kata seperti "deposit", "biaya", dan "transfer" cenderung memiliki nilai TF-IDF yang tinggi pada lowongan fiktif karena sering muncul di lowongan yang meminta pembayaran di awal. Sementara itu, kata-kata seperti "resmi", "bpjs", dan "kontrak" lebih sering muncul pada lowongan asli yang biasanya disertai informasi fasilitas dan status kerja yang jelas. Perbedaan pola kata inilah yang dimanfaatkan oleh Naïve Bayes untuk menghitung peluang suatu lowongan termasuk ke dalam kelas asli atau fiktif.

Walaupun hasil yang diperoleh sudah cukup baik untuk percobaan awal, penelitian ini masih memiliki beberapa keterbatasan. Salah satunya adalah jumlah data yang digunakan belum terlalu banyak, sehingga variasi bentuk penulisan lowongan kerja yang masuk ke dalam model juga masih terbatas. Selain itu, model hanya melihat teks dari isi lowongan tanpa mempertimbangkan faktor lain, seperti reputasi perusahaan, alamat kantor, atau informasi tambahan di luar teks. Ke depan, penelitian ini masih bisa dikembangkan lagi, misalnya dengan menambah jumlah data, menambah fitur lain di luar teks, atau membandingkan Naïve Bayes dengan algoritma lain untuk melihat apakah ada metode yang bisa memberikan hasil yang lebih baik.

#### 4 KESIMPULAN

Dari penelitian yang saya lakukan, secara umum bisa disimpulkan bahwa metode TF-IDF dan algoritma Naïve Bayes bisa dipakai untuk membantu mendeteksi lowongan kerja fiktif di platform rekrutmen online. Setelah teks lowongan dibersihkan dulu lewat proses preprocessing, data jadi lebih rapi dan mudah diolah oleh sistem.

Hasil perhitungan TF-IDF juga menunjukkan bahwa ada beberapa kata yang sering muncul di lowongan fiktif, misalnya "deposit", "biaya", atau "transfer". Sementara itu, di lowongan yang cenderung asli lebih sering muncul kata seperti "resmi", "bpjs", atau "kontrak". Pola kata-kata seperti ini kemudian dimanfaatkan oleh Naïve Bayes untuk menebak apakah sebuah lowongan termasuk asli atau fiktif.

Dari pengujian model, didapatkan nilai akurasi yang cukup baik, sekitar 0,90, dengan nilai precision, recall, dan F1-score yang juga tidak terlalu jauh. Artinya, model yang dibangun sudah lumayan mampu membedakan dua jenis lowongan tersebut, walaupun masih ada beberapa data yang salah diprediksi.

Penelitian ini tentu saja masih punya kekurangan. Jumlah data yang digunakan masih sedikit dan hanya mengambil informasi dari teks lowongan saja. Ke depan, penelitian seperti ini bisa dikembangkan lagi dengan menambah jumlah data, menambah fitur lain seperti informasi perusahaan, atau mencoba algoritma lain untuk dibandingkan dengan Naïve Bayes, supaya hasilnya bisa jadi lebih baik lagi.

## REFERENSI

- [1] F. Akbar dan M. Margaret, "Jaringan Kejahatan Penipuan dalam Perekrutan Tenaga Kerja PT.X di Jakarta Selatan," *J. Kriminol. ANOMIE*, vol. 2, no. 3, hlm. 153–167, Des 2020.
- [2] Y. Nadifa, "Tinjauan Kriminologi Terhadap Penipuan Rekrutmen Tenaga Kerja Melalui Media Online Di Indonesia," *J. Pencerah Bangsa*, vol. 4, no. 1, hlm. 54–59, 2024.
- [3] L. Chrisendo M.S., "Analisis Yuridis Penegakan Hukum Terhadap Tindak Pidana Penipuan dengan Modus Lowongan Kerja Melalui Media Sosial Berdasarkan Hukum Positif di Indonesia," *COMSERVA J. Penelit. Dan Pengabd. Masy.*, vol. 2, no. 12, hlm. 3195–3218, Apr 2023, doi: 10.59141/comserva.v2i12.707.
- [4] herwanto dan D. P. Budiyanasyah, "PREDIKSI REAL OR FAKE JOB POSTING MENGGUNAKAN METODE LONG SHORT-TERM MEMORY," *Innotech J. Ilmu Komput. Sist. Inf. Dan Teknol. Inf.*, vol. 2, no. 1, Jan 2025, Diakses: 16 November 2025. [Daring]. Tersedia pada: <https://ejournal.cyber-univ.ac.id/index.php/innotech/article/view/84>
- [5] - IVAN ASRI PRATAMA, "PERLINDUNGAN HUKUM TERHADAP PENIPUAN LOWONGAN KERJA ONLINE DITINJAU OLEH UNDANG-UNDANG NOMOR 19 TAHUN 2016 TENTANG INFORMASI DAN TRANSAKSI ELEKTRONIK," skripsi, UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU, 2025. Diakses: 16 November 2025. [Daring]. Tersedia pada: <https://repository.uin-suska.ac.id/90892/>
- [6] E. D. Silambi, "PEMUTUSAN HUBUNGAN KERJA DITINJAU DARI SEGI HUKUM ( STUDI KASUS PT.MEDCO LESTARI PAPUA)," *J. ILMU Ekon. Sos.*, vol. 5, no. 2, hlm. 507–516, Okt 2014, doi: 10.35724/jies.v5i2.70.
- [7] Syarif Aminul Khoiri dan Abdul Wahid, "Analisis Kinerja Algoritma Machine Learning dalam Prediksi Harga Cryptocurrency," *JUSTINDO J. Sist. Dan Teknol. Inf. Indones.*, vol. 9, no. 2, hlm. 133–141, Agu 2024, doi: 10.32528/justindo.v9i2.1965.
- [8] F. K. Atmoko dan J. Siwalankerto, "PEMAKAIAN REKRUTMEN ONLINE DI INDONESIA PADA PERUSAHAAN-PERUSAHAAN TERDAFTAR DI BURSA EFEK INDONESIA".
- [9] V. A. Saputra dan S. A. Arnomo, "PENERAPAN METODE MACHINE LEARNING DALAM MENGIDENTIFIKASI BERITA HOAKS," *Comput. Based Inf. Syst. J.*, vol. 12, no. 1, hlm. 112–121, Mar 2024, doi: 10.33884/cbis.v12i1.8442.
- [10] R. Kosasih dan A. Alberto, "Analisis Sentimen Produk Permainan menggunakan Metode TF-IDF dan Algoritma K-Nearest Neighbor," *InfoTekJar J. Nas. Inform. Dan Teknol. Jar.*, vol. 6, no.

- 1, hlm. 134–139, Sep 2021, doi: 10.30743/infotekjar.v6i1.3893.
- [11] R. Rinandyaswara, Y. A. Sari, dan M. T. Furqon, “Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi),” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 9, no. 4, hlm. 717, Agu 2022, doi: 10.25126/jtiik.2022934707.
- [12] R. T. Wahyuni dan D. Prastiyanto, “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi,” vol. 9, no. 1, 2017.
- [13] V. Amrizal, “PENERAPAN METODE TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DAN COSINE SIMILARITY PADA SISTEM TEMU KEMBALI INFORMASI UNTUK MENGETAHUI SYARAH HADITS BERBASIS WEB (STUDI KASUS: HADITS SHAHIH BUKHARI-MUSLIM),” *J. Tek. Inform.*, vol. 11, no. 2, hlm. 149–164, Nov 2018, doi: 10.15408/jti.v11i2.8623.
- [14] R. Rinandyaswara, Y. A. Sari, dan M. T. Furqon, “Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi),” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 9, no. 4, hlm. 717–724, Agu 2022, doi: 10.25126/jtiik.2022934707.
- [15] D. Wahyudi, T. Susyanto, dan D. Nugroho, “IMPLEMENTASI DAN ANALISIS ALGORITMA STEMMING NAZIEF & ADRIANI DAN PORTER PADA DOKUMEN BERBAHASA INDONESIA,” *J. Ilm. SINUS*, vol. 15, no. 2, Jul 2017, doi: 10.30646/sinus.v15i2.305.
- [16] A. Pranata, R. Rudiman, dan N. A. Verdikha, “Klasifikasi Teks Quick Count Pemilihan Presiden 2024 pada Twitter menggunakan Metode TF-IDF dan Naive Bayes,” *J. Inform. Terpadu*, vol. 10, no. 2, hlm. 93–100, Okt 2024, doi: 10.54914/jit.v10i2.1279.
- [17] E. E. A. K. Saputra, dan A. H. N. Kushariantoko, “Analisis Performansi Naïve Bayes pada Klasifikasi Plagiarisme Dokumen Berdasarkan Pembobotan Teks Menggunakan Algoritma TF-IDF,” *EXPERT J. Manaj. Sist. Inf. Dan Teknol.*, vol. 14, no. 2, hlm. 115, Des 2024, doi: 10.36448/expert.v14i2.3952.
- [18] A. D. Herlambang dan S. H. Wijoyo, “Algoritma Naive Bayes untuk Klasifikasi Sumber Belajar Berbasis Teks pada Mata Pelajaran Produktif di SMK Rumpun Teknologi Informasi dan Komunikasi,” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 6, no. 4, hlm. 430, Jul 2019, doi: 10.25126/jtiik.2019641323.
- [19] R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, dan F. A. Zami, “Analisis Perbandingan Algoritma Naive Bayes Classifier

---

dan Support Vector Machine  
untuk Klasifikasi Hoax pada

Berita Online Indonesia,” vol. 13,  
no. 2, 2022.