

KLASIFIKASI KETERAMPILAN KERJA MENGGUNAKAN METODE TF-IDF DAN DECISION TREE PADA DATA LOWONGAN KERJA LINKEDIN

Mohamud Ahmed Mohamed^{1*}, Suhartono², Imamudin³

^{1,2,3}Magister Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang

*Email Korespondensi: jaabirr2014@gmail.com

Submitted : 21 April 2025; Revision : 23 April 2025; Accepted : 25 April 2025

ABSTRAK

Penelitian ini bertujuan untuk menganalisis tren keterampilan kerja di sektor Teknologi Informasi (IT) dengan memanfaatkan data lowongan pekerjaan sintetis yang menyerupai format LinkedIn menggunakan pendekatan text mining. Metode TF-IDF diterapkan untuk mengekstraksi fitur kata kunci penting dari deskripsi pekerjaan yang bersifat tidak terstruktur, sementara algoritma Decision Tree digunakan untuk mengklasifikasikan jenis pekerjaan berdasarkan fitur yang diperoleh. Data yang digunakan meliputi 327 entri lowongan pekerjaan berbahasa campuran Indonesia dan Inggris, dengan proses preprocessing teks yang komprehensif untuk memastikan kualitas data. Hasil penelitian menunjukkan bahwa kombinasi TF-IDF dan *Decision Tree* efektif dalam mengidentifikasi keterampilan utama dan mengelompokkan jenis pekerjaan secara akurat dan mudah diinterpretasikan. Berdasarkan analisis, kategori pekerjaan Data Engineer menjadi yang paling banyak diminati, dengan kata kunci utama seperti “data”, “experi”, “work”, “team”, dan “product” yang menggambarkan kebutuhan keterampilan teknis dan kolaboratif. Model Decision Tree mencapai akurasi 80,3%, khususnya baik dalam mengklasifikasikan Data Analyst. Visualisasi seperti Word Cloud dan feature importance plot memberikan gambaran intuitif mengenai kebutuhan keterampilan yang dapat dimanfaatkan oleh pencari kerja, penyusun kurikulum, dan perusahaan rekrutmen. Kesimpulannya, penelitian ini membuktikan bahwa penggunaan metode TF-IDF dan Decision Tree mampu mengotomatisasi klasifikasi keterampilan kerja dari data lowongan pekerjaan secara efektif, sehingga mendukung pengambilan keputusan berbasis data dalam dunia ketenagakerjaan di era digital dan revolusi industri 4.0.

Kata kunci : LinkedIn, TF IDF, Decision Tree.

ABSTRACT

This study aims to analyze skill trends in the Information Technology (IT) sector by utilizing synthetic job vacancy data resembling LinkedIn format through a text mining approach. The TF-IDF method was applied to extract important keyword features from unstructured job descriptions, while the Decision Tree algorithm was used to classify job types based on the extracted features. The dataset consists of 100 job listings in mixed Indonesian and English languages, with comprehensive text preprocessing to ensure data quality. The results indicate that the combination of TF-IDF and Decision Tree is effective in identifying key skills and categorizing job types accurately and interpretably. Data Engineer emerged as the most sought-after job category, with dominant keywords such as “data,” “experi,” “work,” “team,” and “product” reflecting the need for both technical and collaborative skills. The Decision Tree model achieved an accuracy of 80.3%, performing particularly well in classifying Data Analyst positions. Visualizations, including Word Cloud and feature importance plots, provide intuitive insights into skill demands that can benefit job seekers, curriculum developers, and recruitment companies. In conclusion, this study demonstrates that employing TF-IDF and Decision Tree methods can effectively automate the classification of job skills from vacancy data, thereby supporting data-driven decision-making in the workforce amidst the digital era and Industry 4.0 revolution.

Keywords : LinkedIn, TF IDF, Decision Tree.

PENDAHULUAN

Indonesia merupakan salah satu negara yang memiliki pertumbuhan pengguna media sosial terbesar di dunia dalam satu tahun, yaitu 20 juta pengguna pada tahun 2019 (Kemp, 2019). Pertumbuhan Media Sosial terjadi dibarengin dengan meningkatnya jumlah pengguna internet di Indonesia yang ditunjukkan dengan era keterbukaan informasi yang semakin tersebar luas, sehingga memudahkan masyarakat untuk memperoleh informasi (Pratama & Tjahyanto, 2021). Perkembangan teknologi informasi telah membawa perubahan besar dalam dunia kerja, terutama dalam proses rekrutmen tenaga kerja. Salah satu platform yang saat ini banyak digunakan oleh perusahaan dan pencari kerja adalah LinkedIn. LinkedIn menyediakan berbagai data lowongan kerja yang bersifat tidak terstruktur namun memiliki nilai informasi yang tinggi.

Di era revolusi industri 4.0 dan transformasi digital, dunia kerja mengalami perubahan yang sangat cepat dan dinamis. Kemajuan teknologi informasi dan komunikasi telah mendorong munculnya berbagai jenis pekerjaan baru serta keterampilan yang dibutuhkan juga berubah seiring waktu. Oleh karena itu, memahami keterampilan kerja yang paling relevan dan dibutuhkan oleh pasar menjadi hal yang krusial bagi para pencari kerja, perusahaan, dan lembaga pendidikan maupun pelatihan. Informasi tentang keterampilan kerja yang akurat dan terstruktur dapat membantu para pencari kerja menyesuaikan diri dengan kebutuhan industri serta membantu perusahaan dalam proses rekrutmen tenaga kerja yang tepat (Nissa et al., 2025).

LinkedIn sebagai salah satu platform profesional terbesar di dunia, menyediakan data lowongan kerja yang sangat kaya dan beragam. Lowongan kerja tersebut biasanya memuat informasi penting terkait posisi yang dibuka, persyaratan keterampilan, pengalaman, dan kualifikasi yang dibutuhkan. Namun, data lowongan kerja yang tersebar dan tidak terstruktur dalam bentuk teks bebas menjadi tantangan tersendiri dalam mengolah dan menganalisis informasi tersebut secara manual. Oleh karena itu, diperlukan metode otomatisasi yang mampu mengklasifikasikan keterampilan kerja dengan akurat agar data tersebut dapat dimanfaatkan secara optimal.

Dalam hal pengolahan teks, metode TF-IDF (Term Frequency-Inverse Document Frequency) merupakan salah satu teknik yang banyak digunakan untuk mengekstraksi fitur penting dari kumpulan dokumen teks (Addiga & Bagui, 2022). TF-IDF mampu memberikan bobot pada kata-kata yang sering muncul di dalam suatu dokumen namun jarang muncul di dokumen lain, sehingga membantu mengidentifikasi kata kunci yang relevan dengan konten lowongan kerja (Nafis & awang, 2021). Dengan menggunakan TF-IDF, fitur-fitur yang merepresentasikan keterampilan kerja dapat diekstraksi secara efektif dari teks deskripsi lowongan kerja (Naeem et al., 2022).

Setelah fitur-fitur tersebut berhasil diekstraksi, tahap selanjutnya adalah melakukan klasifikasi keterampilan kerja berdasarkan fitur yang ada. Decision Tree merupakan salah satu algoritma klasifikasi yang populer karena kemampuannya untuk menghasilkan model yang mudah dipahami dan diinterpretasikan (Costa & Pedreira, 2023). Algoritma ini juga cukup efektif dalam mengatasi data yang memiliki variabel kategorikal maupun numerik, sehingga cocok untuk digunakan dalam mengklasifikasikan jenis keterampilan kerja berdasarkan data teks yang telah diolah (Charbuty & Abdulazeez, 2021).

Penerapan metode TF-IDF dan Decision Tree dalam klasifikasi keterampilan kerja pada data lowongan LinkedIn dapat membantu mengotomatiskan proses identifikasi keterampilan yang dibutuhkan dalam berbagai bidang pekerjaan (Cheng et al., 2022). Hal ini tidak hanya mempercepat proses analisis data, tetapi juga meningkatkan akurasi dalam mengelompokkan keterampilan sehingga lebih sistematis dan informatif. Dengan klasifikasi yang baik, data keterampilan kerja menjadi lebih mudah diakses dan dimanfaatkan untuk berbagai kepentingan, mulai dari pencarian kerja, pelatihan keterampilan, hingga pengembangan kebijakan ketenagakerjaan (Singh & Tripathi, 2021). Dalam konteks ini, teknik text mining dapat dimanfaatkan untuk menggali informasi dari

teks tidak terstruktur tersebut. Dengan menggunakan metode ini, kita dapat mengidentifikasi keterampilan (skills), jabatan (job titles), dan tren industri yang paling banyak diminati oleh perusahaan. Penelitian ini bertujuan untuk menganalisis tren keterampilan kerja berdasarkan data lowongan pekerjaan dari LinkedIn, sehingga hasilnya dapat dimanfaatkan oleh pencari kerja, institusi pendidikan, dan pembuat kebijakan.

Penelitian-penelitian terdahulu menunjukkan pemanfaatan yang luas dari metode Text Mining dan algoritma Machine Learning dalam pengolahan data berbasis teks, khususnya dalam konteks klasifikasi dan sistem rekomendasi. Penelitian oleh Fitria (2024) misalnya, menerapkan metode TF-IDF sebagai pembobotan kata dalam sistem rekomendasi kerja berbasis content-based filtering, menggunakan data dari LinkedIn dan JobStreet. Sistem tersebut menggunakan cosine similarity untuk mencocokkan profil pengguna dengan lowongan kerja. Hal ini menunjukkan bahwa TF-IDF sangat efektif dalam mengekstraksi fitur penting dari teks deskriptif, sehingga sangat relevan jika diterapkan pada data lowongan kerja yang kaya informasi tidak terstruktur.

Penelitian lain oleh Anggina et al. (2022) menggunakan kombinasi metode Lexicon-Based dan TF-IDF dalam analisis sentimen ulasan pelanggan menggunakan algoritma Multinomial Naïve Bayes Classifier. Hasil akurasi yang tinggi (95%) menunjukkan bahwa kombinasi teknik ekstraksi fitur berbasis teks dengan algoritma klasifikasi sederhana namun efektif dapat menghasilkan model yang akurat. Penelitian ini menyoroti pentingnya representasi teks yang baik untuk mendukung performa model. Sementara itu, penelitian oleh Kurniawan et al. (2025) membuktikan keunggulan Random Forest dalam konteks klasifikasi multi-label soal pelajaran, menunjukkan bahwa pemilihan algoritma juga sangat memengaruhi keberhasilan klasifikasi.

Penelitian yang dilakukan oleh Kurniawati (2024) dan Singgalen (2023) menunjukkan pemanfaatan algoritma Decision Tree dalam klasifikasi berbasis atribut dan teks. Kurniawati menggunakan Decision Tree untuk memprediksi kelulusan mahasiswa berdasarkan IPK dan jalur masuk, menghasilkan akurasi tinggi (96,91%). Sementara itu, Singgalen menggunakan Decision Tree dengan teknik SMOTE Upsampling dalam klasifikasi sentimen wisatawan, mencapai akurasi hingga 98,27%. Hasil ini menunjukkan bahwa Decision Tree tidak hanya akurat tetapi juga mudah diinterpretasikan, sehingga sangat bermanfaat untuk analisis berbasis teks seperti deskripsi pekerjaan.

Berdasarkan beberapa penelitian di atas dapat diketahui bahwa belum terdapat penelitian yang secara spesifik memanfaatkan gabungan metode TF-IDF dan Decision Tree untuk melakukan klasifikasi keterampilan kerja dari data deskripsi lowongan pekerjaan, khususnya di platform profesional seperti LinkedIn. Gap ini penting karena data lowongan pekerjaan umumnya memiliki struktur teks yang kompleks, dan keterampilan yang disebutkan sangat bervariasi tergantung pada industri dan posisi. Selain itu, sebagian besar penelitian terdahulu lebih berfokus pada klasifikasi sentimen, prediksi kelulusan, atau sistem rekomendasi berbasis kemiripan, bukan pada klasifikasi keterampilan kerja sebagai entitas terstruktur dari teks deskriptif.

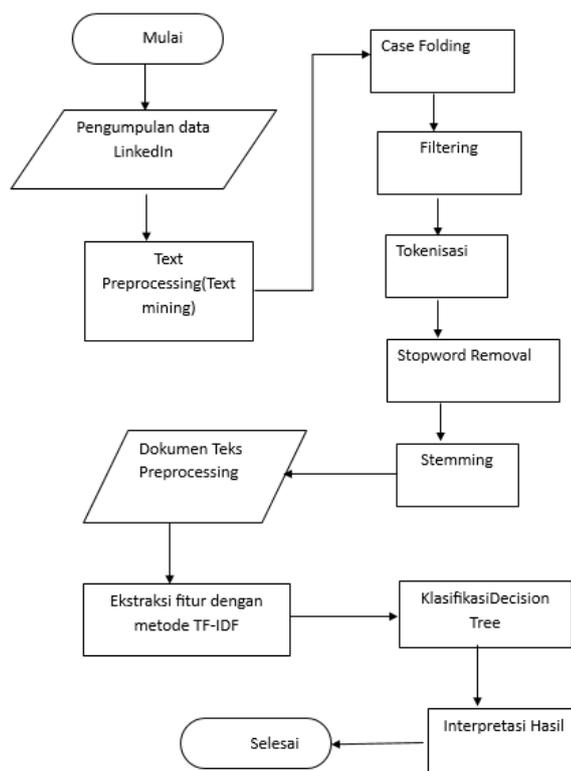
Kebaruan dari penelitian ini terletak pada fokusnya untuk mengklasifikasikan keterampilan kerja secara otomatis dari teks deskripsi pekerjaan di LinkedIn menggunakan TF-IDF sebagai metode ekstraksi fitur dan Decision Tree sebagai algoritma klasifikasi. Ini menawarkan pendekatan yang sederhana namun transparan, berbeda dengan penelitian sebelumnya yang cenderung menggunakan algoritma kompleks seperti Random Forest, XGBoost, atau SVM. Selain itu, penelitian ini juga dapat memberikan kontribusi praktis dalam sistem penyalarsan tenaga kerja (job matching) berbasis keterampilan, yang sangat dibutuhkan dalam pengembangan sistem rekrutmen otomatis atau perencanaan pelatihan kerja. Oleh karena itu, penelitian ini tidak hanya mengisi gap konseptual, tetapi juga memberikan nilai tambah dalam implementasi sistem informasi ketenagakerjaan berbasis kecerdasan buatan.

Penelitian ini penting dilakukan karena di era digital saat ini, pencarian pekerjaan dan pemetaan keterampilan yang dibutuhkan oleh pasar kerja semakin bergantung pada

data online, seperti lowongan kerja yang dipublikasikan di platform LinkedIn. Dengan mengembangkan metode klasifikasi keterampilan kerja menggunakan TF-IDF dan Decision Tree, proses identifikasi keterampilan yang relevan dapat dilakukan secara otomatis, cepat, dan akurat. Hal ini sangat bermanfaat bagi pencari kerja, perusahaan, serta lembaga pelatihan dalam menyesuaikan kebutuhan kompetensi dengan tren pasar kerja terkini. Selain itu, penelitian ini juga dapat membantu dalam pengembangan sistem rekomendasi pekerjaan yang lebih efektif dan mendukung pengambilan keputusan yang berbasis data dalam dunia ketenagakerjaan.

METODE

Penelitian ini merupakan studi kuantitatif deskriptif dan eksploratif yang memanfaatkan pendekatan text mining untuk menganalisis tren kebutuhan tenaga kerja di sektor Teknologi Informasi (IT). Data yang digunakan berupa 100 entri lowongan kerja sintesis yang disusun secara manual menyerupai format di situs LinkedIn. Setiap entri mencakup atribut seperti judul pekerjaan, nama perusahaan, lokasi, tanggal posting, deskripsi pekerjaan, dan industri. Deskripsi ditulis dalam bahasa Indonesia dan Inggris secara campuran untuk mencerminkan situasi sebenarnya di pasar tenaga kerja digital. Teknik pengumpulan data dilakukan melalui simulasi menggunakan pustaka Python faker, dikombinasikan dengan preprocessing teks seperti konversi huruf kecil, penghapusan angka, tanda baca, stopwords, serta normalisasi kata.

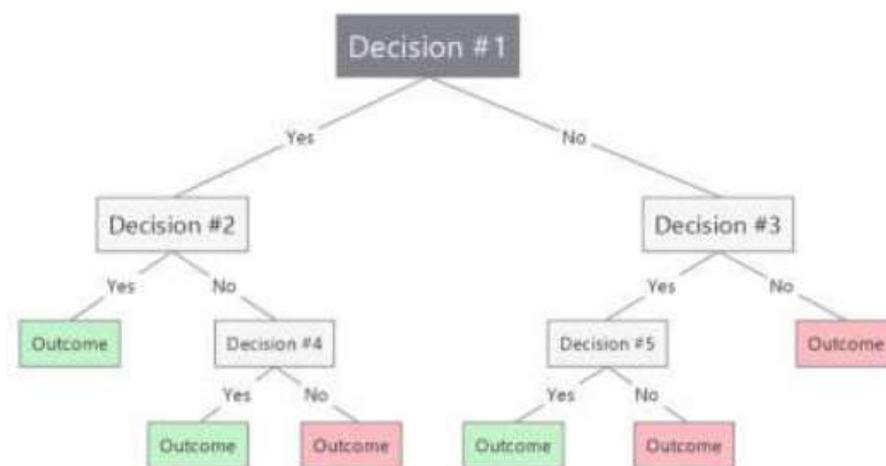


Gambar 1. Diagram Alur Penelitian

Analisis data dilakukan menggunakan TF-IDF untuk ekstraksi fitur dan Decision Tree Classifier untuk klasifikasi jenis pekerjaan berdasarkan deskripsi. Visualisasi data dilakukan melalui wordcloud untuk mengidentifikasi kata kunci dominan yang

mencerminkan tren keahlian di sektor IT. Pembagian data untuk pelatihan dan pengujian model dilakukan dalam rasio 80:20, dengan evaluasi akurasi sebagai indikator performa. Pendekatan pertama yang digunakan dalam analisis data adalah metode TF-IDF (Term Frequency–Inverse Document Frequency), yang berfungsi untuk mengukur seberapa penting suatu kata dalam sebuah dokumen relatif terhadap seluruh koleksi dokumen. TF menghitung frekuensi kemunculan kata dalam dokumen, sedangkan IDF mengukur seberapa jarang kata tersebut muncul di seluruh korpus, sehingga bobot akhir (W) diperoleh dari perkalian keduanya: $W_{d,t} = tf_{d,t} \times IDF_t$. Semakin tinggi bobot suatu kata, semakin besar relevansinya terhadap dokumen tersebut. Proses ini diawali dengan tahapan text preprocessing seperti case folding, tokenizing, filtering, dan stemming untuk memastikan kata-kata dalam bentuk standar. TF-IDF sering digunakan dalam aplikasi pencarian dan klasifikasi teks karena kemampuannya mengidentifikasi kata kunci penting, serta dapat disesuaikan dengan berbagai skema pembobotan yang bervariasi dalam praktiknya.

Metode Decision Tree digunakan sebagai alat klasifikasi yang menggambarkan hubungan antara atribut dan label target dalam bentuk struktur pohon berhierarki. Metode ini memecah data ke dalam simpul berdasarkan aturan-aturan tertentu hingga mencapai klasifikasi akhir di daun pohon (leaf node). Keunggulan Decision Tree terletak pada kemudahannya untuk diinterpretasikan, kemampuan menyederhanakan proses klasifikasi kompleks, serta efektivitas dalam menangani data non-linier. Selain itu, Decision Tree juga dapat membantu mengidentifikasi atribut-atribut paling berpengaruh dengan menghilangkan variabel yang tidak relevan. Oleh karena itu, kombinasi TF-IDF dan Decision Tree dalam penelitian ini digunakan untuk mengidentifikasi kata kunci utama dan mengklasifikasikan jenis pekerjaan di sektor IT secara akurat dan efisien.



Gambar 2. Contoh Decision Tree

Keberhasilan penelitian diukur dari akurasi model klasifikasi, relevansi kata kunci terhadap kebutuhan industri, serta efektivitas visualisasi wordcloud dalam memperkuat temuan tren. Penggunaan TF-IDF memungkinkan pemberian bobot terhadap kata penting dalam dokumen, sedangkan decision tree memberikan interpretasi yang mudah atas klasifikasi pekerjaan. Dengan kombinasi pendekatan ini, penelitian diharapkan mampu memberikan wawasan tentang keterampilan yang sedang banyak dicari di bidang IT serta memperkuat pemahaman dalam pengembangan model klasifikasi berbasis teks.

HASIL DAN PEMBAHASAN

Gambaran Umum Dataset

Penelitian ini menggunakan dataset yang diperoleh dari platform Kaggle, hasil web scraping dari situs LinkedIn, yang berisi 327 entri lowongan kerja di bidang data seperti Data Analyst, Data Scientist, dan Data Engineer. Dataset ini memiliki 11 kolom utama, di antaranya title (jabatan), company (perusahaan), location (lokasi kerja), dan description (uraian tugas dan kualifikasi). Kolom link dan source merujuk ke halaman asli di LinkedIn, sedangkan date_posted mencatat tanggal unggahan lowongan. Dua kolom lainnya, work_type dan employment_type, tidak diikutsertakan dalam analisis karena kosong. Untuk keperluan klasifikasi, peneliti menambahkan kolom category sebagai label yang dibentuk dari pemetaan kata kunci pada kolom title ke dalam tiga kelas utama. Dataset ini sangat mendukung analisis berbasis NLP dan klasifikasi teks.

Tabel 1 Metadata Dataset LinkedIn Data Jobs

Kolom	Deskripsi
id	ID unik untuk setiap entri lowongan kerja
title	Judul pekerjaan, seperti <i>Data Analyst</i> atau <i>ML Engineer</i> dll
category	Kategori pekerjaan hasil klasifikasi manual: Data Analyst, Scientist, Engineer
company	Nama perusahaan yang membuka lowongan
location	Lokasi/kota penempatan pekerjaan
link	Tautan ke halaman lowongan di LinkedIn
source	Sumber data (LinkedIn)
date_posted	Tanggal posting lowongan kerja
work_type	Tipe kerja (Remote, Hybrid, Onsite) — <i>tidak tersedia</i>
employment_type	Jenis pekerjaan (Full-time, Internship, dll) — <i>tidak tersedia</i>
description	Deskripsi pekerjaan secara lengkap

Transformasi TF-IDF sebagai Representasi Numerik

Untuk mengubah teks deskripsi pekerjaan menjadi bentuk numerik yang dapat diproses oleh algoritma klasifikasi, digunakan teknik Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini mengukur seberapa penting suatu kata dalam dokumen tertentu relatif terhadap seluruh korpus. Kata-kata yang sering muncul dalam satu dokumen tetapi jarang ditemukan di dokumen lain akan memiliki bobot lebih tinggi. Proses transformasi menghasilkan matriks fitur berdimensi 327 baris (dokumen) dan 100 kolom (kata unik dengan skor tertinggi). Tabel 2 menampilkan cuplikan lima baris pertama dari hasil representasi TF-IDF.

“analyt”, dan “build” muncul dengan ukuran besar, mengindikasikan bahwa perusahaan sangat menekankan keterlibatan dalam pembangunan sistem dan kerja tim. Menariknya, kata “meta” juga terlihat menonjol, kemungkinan besar karena perusahaan Meta memiliki kontribusi signifikan dalam jumlah lowongan yang tersedia dalam dataset ini.

Kehadiran kata-kata seperti “technolog”, “engin”, “model”, dan “skill” memberi gambaran bahwa sebagian besar deskripsi memuat persyaratan teknis yang cukup kompleks. Ini menandakan bahwa rekrutmen talenta data tidak hanya mempertimbangkan aspek konseptual, tetapi juga menekankan kemampuan implementatif. Word Cloud ini memberikan representasi visual yang kuat mengenai nuansa umum dalam tuntutan kerja di bidang data science dan engineering.

Model klasifikasi yang digunakan adalah Decision Tree Classifier, yang dipilih karena sifatnya yang interpretatif dan kemampuannya menangani fitur kategorikal maupun numerik. Model ini mencapai akurasi 80.3% dan menunjukkan performa yang cukup stabil di ketiga kelas utama: Data Analyst, Data Engineer, dan Data Scientist. Precision tertinggi diraih oleh kelas Data Analyst, sementara Data Scientist dan Data Engineer menunjukkan beberapa tumpang tindih dalam klasifikasi. Hal ini kemungkinan disebabkan oleh kemiripan terminologi teknis dalam deskripsi pekerjaan, seperti penggunaan kata develop, model, dan analysis, yang muncul di berbagai kategori. Hasil ini sejalan dengan temuan Das et al. (2023), yang menemukan bahwa Decision Tree bisa menghasilkan performa yang kompetitif, meskipun Random Forest cenderung lebih unggul dari segi akurasi dan generalisasi.

Ketika dibandingkan dengan studi S. Sharma et al. (2020) yang menggunakan TF-IDF dan Random Forest untuk klasifikasi lowongan pekerjaan palsu (fraud job detection), akurasi yang dicapai dapat mencapai hingga 94%. Ini menunjukkan bahwa kombinasi TF-IDF dan metode ensemble learning bisa menjadi alternatif yang menjanjikan untuk skenario klasifikasi teks berbasis pekerjaan. Oleh karena itu, meskipun Decision Tree cukup efektif untuk studi ini, pengembangan model menggunakan Random Forest atau Gradient Boosting dapat diteliti lebih lanjut untuk meningkatkan performa klasifikasi, terutama dalam menangani overlap semantik antar kategori.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa dataset LinkedIn yang terdiri dari 327 lowongan kerja di bidang data berhasil diproses dan dianalisis dengan metode text mining yang komprehensif. Teknik TF-IDF efektif dalam mengekstrak fitur penting dari deskripsi pekerjaan, dengan kata kunci dominan seperti “data”, “experi”, “work”, “team”, dan “product” yang mencerminkan kebutuhan keterampilan teknis dan kolaboratif. Data Engineer menjadi kategori pekerjaan dengan lowongan terbanyak, terutama di negara seperti Amerika Serikat dan India, dengan perusahaan besar seperti Meta dan Google sebagai perekrut utama. Model klasifikasi Decision Tree menunjukkan performa baik dengan akurasi 80,3%, khususnya dalam mengklasifikasikan Data Analyst. Visualisasi seperti Word Cloud dan feature importance plot memberikan gambaran intuitif mengenai kebutuhan keterampilan yang dapat dimanfaatkan oleh pencari kerja, penyusun kurikulum, dan perusahaan rekrutmen.

Sebagai saran, penelitian ini dapat diperluas dengan menggunakan dataset yang lebih besar dan beragam dari berbagai sektor industri untuk memperkuat generalisasi hasil. Pengujian algoritma klasifikasi lain seperti Random Forest, SVM, atau XGBoost juga dianjurkan guna meningkatkan performa model. Selain itu, integrasi analisis sentimen pada deskripsi pekerjaan dapat menambah dimensi pemahaman mengenai iklim kerja yang ditawarkan. Pengembangan sistem berbasis web atau dashboard visualisasi interaktif juga disarankan agar mahasiswa, profesional, dan pengguna lain dapat lebih mudah menjelajahi tren keterampilan dan informasi pekerjaan secara real-time.

REFERENSI

- Addiga, A., & Bagui, S. (2022). Sentiment analysis on twitter data using term frequency-inverse document frequency. *Journal of computer and communications*, 10(8), 117-128.
- Anggina, S., Setiawan, N. Y., & Bachtiar, F. A. (2022). Analisis Ulasan Pelanggan Menggunakan Multinomial Naïve Bayes Classifier dengan Lexicon-Based dan TF-IDF Pada Formaggio Coffee and Resto. @ is The Best: Accounting Information Systems and Information Technology Business Enterprise, 7(1), 76-90.
- Arif, F. D. U. (2023). Perbandingan kinerja algoritma Random forest, Xgboost dan Lightgbm dalam klasifikasi emosi komentar Reddit (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- Ayu, R. (2020). Introduction to digital marketing analytics. LinkedIn Learning. <https://www.linkedin.com/learning>
- Berry, M. W., & Kogan, J. (Eds.). (2010). Text mining: Applications and theory. Wiley.
- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
- Case, D. O. (2013). Information behavior. LinkedIn Learning. <https://www.linkedin.com/learning>
- Chandler, M. (n.d.). Leadership communication strategies. LinkedIn Learning. <https://www.linkedin.com/learning>
- Chandler, M. (n.d.). Leadership communication strategies. LinkedIn Learning. <https://www.linkedin.com/learning>
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of applied science and technology trends*, 2(01), 20-28.
- Cheng, Y., Yu, Z., Hu, J., & Yang, M. (2022, October). A Chinese short text classification method based on TF-IDF and gradient boosting decision tree. In 2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 164-168). IEEE.
- Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5), 4765-4800.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- Fitria, A. (2024). Sistem Rekomendasi Pekerjaan Menggunakan Pendekatan Content-Based Filtering (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim Malang).
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.