



Analisis Proyeksi Kebutuhan Tenaga Kerja Berdasarkan *Skills* Yang Dibutuhkan Menggunakan Algoritma *Naive Bayes Classifier*

Nur Azizah Firdausa^{a,1,*}, Ribka Rifanny Br Girsang^{b,2}, Dela Oktaviana^{b,3},

Astri Wahyuningsiam^{b,4}, Muhammad Habibi^{b,5}

^a Informatika, FTTI Unjaya, Yogyakarta, Indonesia

¹ azizahfirdaus88@gmail.com*; ² ribkarifanny2002@gmail.com; ³ laadelaa102@gmail.com, ⁴ astriwahyu29@gmail.com, ⁵ Muhammadhabibi17@gmail.com

* corresponding author

ABSTRAK

ARTICLE INFO

Pada Agustus 2023, Indonesia menghadapi jumlah pengangguran mencapai 7,86 juta orang walaupun tidak memungkirinya bahwa persentase pengangguran telah menurun dari tahun sebelumnya. Dari data tersebut dikelompokkan menjadi empat kelompok yaitu pengangguran melibatkan mereka yang mencari pekerjaan, berusaha mendirikan usaha, merasa sulit mendapatkan pekerjaan, dan bahkan yang sudah bekerja namun belum mulai. Pandemi Covid-19 mengubah paradigma pekerjaan menjadi *remote*, namun kebutuhan akan informasi pekerjaan tetap menjadi kunci. Proyeksi kebutuhan tenaga kerja memberikan wawasan jangka panjang tentang sektor dan bidang yang menjanjikan, memandu pencari kerja untuk mengembangkan keterampilan sesuai tren pasar kerja. Penelitian ini dilakukan dengan menggunakan klasifikasi naive bayes, yaitu sebuah metode klasifikasi teks berdasarkan probabilitas kata kunci dalam membandingkan dokumen latih dan dokumen uji. Dengan metode klasifikasi ini diharapkan dapat membantu mengurangi tingkat pengangguran dan menyelaraskan keterampilan individu dengan kebutuhan industri, memberikan kontribusi pada kebijakan pendidikan dan pelatihan untuk membuat keputusan karir yang cerdas di era digital.

Article history

Received: 1 April 2024
Revised: 24 May 2024
Accepted: 1 August 2024

Keywords

Pengangguran, Strategi Pengurangan Pengangguran, Naive Bayes Classifier, Klasifikasi, Proyeksi Kebutuhan Tenaga Kerja.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

1. Pendahuluan

Jumlah pengangguran di Indonesia mencapai 7,86 juta orang pada Agustus 2023, menurut Badan Pusat Statistik (BPS). Hal ini mencakup empat kelompok: orang-orang yang tidak memiliki pekerjaan dan sedang mencari pekerjaan, orang-orang yang tidak memiliki pekerjaan dan berusaha membangun bisnis, orang-orang yang tidak memiliki pekerjaan dan tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan, dan orang-orang yang sudah memiliki pekerjaan tetapi belum mulai bekerja. Banyak faktor yang mendasari seseorang menjadi pengangguran, salah satunya adalah kekurangan informasi pekerjaan yang dapat mempengaruhi efikasi diri individu dalam pengambilan keputusan karir yang lebih baik. Kekurangan informasi pekerjaan dapat menyebabkan kesulitan dalam pengambilan keputusan karir, sehingga individu perlu dilengkapi dengan bimbingan karir yang membantu mereka mengembangkan kemampuan dan keputusan karir yang lebih baik.



Perubahan dalam dinamika pasar kerja, terutama yang dipicu oleh faktor eksternal yang telah dialami sebelumnya seperti pandemi Covid-19, semakin mempertegas pentingnya akses terhadap informasi pekerjaan. Pandemi Covid-19 telah membawa perubahan dalam bidang pekerjaan. Dari yang awalnya on-site, pekerjaan bisa dikerjakan secara *remote* dari rumah dan dari mana saja. Bahkan para pencari kerja tidak harus bekerja dari negaranya sendiri, mereka bisa mendaftar pekerjaan dari berbagai negara lain dan bekerja secara *remote*. Di era digital yang semakin canggih, hal ini telah memudahkan para pencari kerja. Namun, mereka masih memerlukan informasi yang akurat dan tepat waktu untuk mendapatkan pekerjaan yang sesuai dengan keterampilan dan keahlian mereka. Informasi yang tepat dapat membantu mereka dalam mengembangkan strategi karir yang lebih efektif dan meningkatkan kesempatan mendapatkan pekerjaan.

Beberapa penelitian sebelumnya telah menyoroti pentingnya informasi karir dalam mengatasi pengangguran. Menurut *World Economic Forum* (2023), adopsi teknologi dan digitalisasi dalam pekerjaan dapat menyebabkan *churn* pasar tenaga kerja, namun juga menciptakan peluang baru bagi pekerja yang memiliki akses informasi karir yang tepat dan komprehensif. Penelitian dari *Asian Development Bank* (2023) menunjukkan bahwa bimbingan karir yang efektif dapat membantu individu memahami pasar kerja dan mempersiapkan mereka untuk memenuhi tuntutan pekerjaan melalui program pelatihan dan bimbingan karir yang tepat waktu.

Untuk menjawab tantangan ini, penelitian ini menggunakan metode *Naive Bayes* untuk mengklasifikasikan lowongan pekerjaan dari *Google* dan *YouTube*. Tujuannya adalah untuk mengidentifikasi kata-kata yang sering muncul dalam lowongan pekerjaan tersebut, sehingga pencari kerja dapat memahami keterampilan apa yang paling dibutuhkan oleh perusahaan-perusahaan tersebut. Berdasarkan pemahaman ini, pencari kerja dapat lebih mempersiapkan diri dengan keterampilan yang relevan, meningkatkan peluang mereka untuk mendapatkan pekerjaan yang sesuai.

2. Metode Penelitian

Dengan menggunakan data dari *Kaggle* dan metode klasifikasi *Naive Bayes*, penelitian ini dimulai dengan latar belakang masalah, pemetaan proses, dan analisis data mengenai Data Kebutuhan Tenaga Kerja. Dengan menggunakan metode klasifikasi *Naive Bayes Classification* dan *Confusion Matrix* diharapkan dapat memberikan dasar yang kuat untuk mendukung temuan dan kesimpulan dalam analisis penelitian ini. Metode *Naive Bayes* dipilih karena kemampuannya dalam menghadapi kompleksitas data dan menghasilkan prediksi yang akurat. Penelitian dalam *Jurnal Teknik Informatika* menunjukkan bahwa metode *Naive Bayes* dapat digunakan dalam sistem pendukung keputusan penentuan status gizi balita dengan keuntungan yang signifikan karena membutuhkan jumlah data pelatihan yang kecil. Hasil penelitian dalam *Jurnal Siliwangi* menunjukkan bahwa *Naive Bayes* memiliki akurasi, presisi, dan recall yang hampir sama dengan metode *Deep Learning*, yang merupakan metode yang lebih kompleks dan membutuhkan lebih banyak data pelatihan. Kesimpulan dari penelitian-penelitian tersebut adalah bahwa metode *Naive Bayes* dapat digunakan dengan baik. Proses metode penelitian digambarkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Dataset

Data Collection adalah proses pengambilan data, di mana data tersebut diambil dari *Kaggle*, berupa hasil dari sekumpulan postingan kebutuhan lowongan pekerjaan dari perusahaan *Google* dan *Youtube*. Kedua perusahaan tersebut merupakan perusahaan berskala besar di bidang industri, teknologi dan media sosial, masing-masing memainkan peran yang sangat signifikan di dunia digital.

Data yang dikumpulkan mencakup beberapa kolom yang relevan, berikut adalah penjelasan lebih lanjut mengenai makna setiap kolom dalam dataset ini :

- *Company* : Mengidentifikasi perusahaan yang membuka lowongan pekerjaan. Kolom ini memberikan konteks penting dalam lingkungan kerja perusahaan.
- *Title* : Menyediakan nama posisi atau jabatan yang ditawarkan. Informasi ini membantu dalam mengidentifikasi jenis pekerjaan yang sedang dibuka.
- *Category* : Mengelompokkan pekerjaan berdasarkan bidang tertentu, berguna untuk memahami distribusi pekerjaan di bidang divisi dalam perusahaan.
- *Responsibilities* : Menjelaskan tugas dan tanggung jawab utama yang akan dilakukan oleh kandidat yang terpilih.
- *Minimum Qualifications* : Menyebutkan kualifikasi minimum yang harus dipenuhi oleh kandidat untuk dapat mempertimbangkan posisi tersebut.
- *Preferred Qualifications* : Memberikan rincian mengenai kualifikasi tambahan yang diinginkan dari kandidat. Kualifikasi ini biasanya mencakup ketrampilan tambahan, sertifikasi, atau pengalaman yang dapat memberikan nilai tambah bagi kandidat.
- Data lain yang berkaitan : Mencakup informasi tambahan yang relevan dengan persyaratan lowongan pekerjaan.

Dalam proses klasifikasi, kolom *Title*, *Category*, *Responsibilities*, *Minimum Qualifications*, dan *Preferred Qualifications* sering dianggap sebagai fitur utama karena mencakup informasi yang kritis untuk memprediksi atau mengelompokkan pekerjaan. Kolom *Company* digunakan untuk analisis perbandingan antara *Google* dan *YouTube* jika diperlukan, untuk melihat perbedaan atau kesamaan

dalam persyaratan pekerjaan antara kedua perusahaan. Semua kolom ini dipertimbangkan sebagai fitur dalam proses klasifikasi untuk memprediksi kebutuhan tenaga kerja.

2.2. Preprocessing Data

Untuk melakukan proses klasifikasi data kebutuhan tenaga kerja, maka data yang dikumpulkan diperlukan menjalani tahap pembersihan dan penyeragaman data agar menghasilkan fakta baru yang akurat. *Preprocessing* data dilakukan dengan menggunakan NLTK, yaitu sebuah platform berbasis Python untuk memproses data. Proses yang dilakukan meliputi *text cleaning* yaitu pembersihan data untuk menghilangkan informasi yang tidak relevan dengan text, *tokenizing* yaitu proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian di analisis, normalisasi yaitu menyamakan format atau bentuk dari unit-unit teks, *stopword removal* adalah tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting) dan juga *stemming* yang merupakan proses menghilangkan infleksi kata ke bentuk dasarnya, namun bentuk dasar tersebut tidak berarti sama dengan akar kata (*root word*). Selain itu, dilakukan pengecekan data untuk mengetahui apakah ada data yang kosong atau tidak. Berikut adalah hasil *preprocessing* dari data kebutuhan tenaga kerja dapat dilihat pada Gambar 2.

```
import pandas as pd
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
# df = pd.read_csv('data_lowker.csv')
df = pd.read_csv('job_skills.csv', encoding='latin1')

# bersihkan nilai yg tidak valid
def clean_text(text):
    return str(text)
df = df.head(20000)
df['Preferred Qualifications'] = df['Preferred Qualifications'].apply(clean_text)

# tokenisasi
def skills_tokenize(text):
    tokens = word_tokenize(text)
    return tokens

# normalisasi
def skills_normalized(tokens):
    normalized_tokens = [token.lower() for token in tokens if token.isalpha()]
    return normalized_tokens

# stop words
def skills_stop_removed(tokens):
    stop_words = set(stopwords.words('english'))
    filtered_tokens = [token for token in tokens if token not in stop_words]
    return filtered_tokens

# stemming (Bahasa Inggris)
def skills_stemmed(tokens):
    ps = PorterStemmer()
    stemmed_tokens = [ps.stem(token) for token in tokens]
    return stemmed_tokens

# clean teks
def skills_clean(text):
    cleaned_text = ' '.join(e.lower() for e in text if e.isalnum() or e.isspace())
    return cleaned_text

df.rename(index=str, columns={'Minimum Qualifications':'Minimum_Qualifications', 'Preferred Qualifications':'Preferred_Qualifications'})

df = df.dropna()
df['Preferred_Qualifications_tokenize'] = df['Preferred_Qualifications'].apply(skills_tokenize)
df['Preferred_Qualifications_normalized'] = df['Preferred_Qualifications_tokenize'].apply(skills_normalized)
df['Preferred_Qualifications_stop_removed'] = df['Preferred_Qualifications_normalized'].apply(skills_stop_removed)
df['Preferred_Qualifications_stemmed'] = df['Preferred_Qualifications_stop_removed'].apply(skills_stemmed)
df['Preferred_Qualifications_clean'] = df['Preferred_Qualifications_stemmed'].apply(skills_clean)
df[['Title', 'Company', 'Category', 'Location', 'Responsibilities', 'Preferred_Qualifications', 'Preferred_Qualifications_tokenize']]
```

Gambar 2. Preprocessing Data

2.3. Pelabelan

Pada tahap ini, konversi label digunakan untuk mengubah informasi kategori atau kelas yang dimiliki oleh setiap data menjadi format biner. Konversi ini dilakukan berdasarkan kolom *Company*

dengan angka 0 sebagai *YouTube* dan angka 1 sebagai *Google*. Alasan dari konversi label ini menjadi format biner adalah untuk mempermudah proses pelatihan model dalam *Naive Bayes Classification* (NBC). Dengan mengonversi label menjadi biner, model dapat lebih mudah memahami pola atau distribusi probabilitas fitur-fitur yang terkait dengan setiap kelas.

Company	Category	Label
YouTube	Business Strategy	0
YouTube	Sales & Account Management	0
YouTube	Product & Customer Support	0
YouTube	Partnerships	0
YouTube	Business Strategy	0
...
Google	Real Estate & Workplace Services	1
Google	Real Estate & Workplace Services	1
Google	Real Estate & Workplace Services	1
Google	Real Estate & Workplace Services	1
Google	Real Estate & Workplace Services	1

[1235 rows x 2 columns]

Gambar 3. Labelling Data

2.4. Training Data

Training dalam sistem klasifikasi kebutuhan tenaga kerja melibatkan penggunaan data teks dari kolom-kolom seperti *Title*, *Category*, *Responsibilities*, *Minimum Qualifications*, dan *Preferred Qualification*. Pendekatan ini bertujuan untuk menghasilkan model klasifikasi yang mampu secara otomatis mengidentifikasi ketrampilan dan kebutuhan tenaga kerja berdasarkan diskripsi pekerjaan. Pada proses ini, menggunakan metode *Term Frequency-Invers Document Frequency* (TF-IDF) untuk ekstraksi fitur data teks. Metode TF-IDF membantu sistem mendapatkan representasi numerik yang merefleksikan pentingnya setiap kata dalam dokumen, memungkinkan pembuatan model yang cerdas dalam mengklasifikasikan data berdasarkan kategori yang telah ditentukan.

2.5. Testing

Pengujian pada sistem klasifikasi kebutuhan tenaga kerja dilaksanakan untuk mengevaluasi tingkat akurasi dari model yang telah dibangun selama proses pelatihan dengan menggunakan data uji. Pada penelitian ini, dataset dibagi menjadi dua bagian data training dan data *testing*. Pembagian data dilakukan dengan rasio 80:20 atau 70:30, di mana sebagian besar data digunakan untuk model (*training*) dan sisanya digunakan untuk menguji performa model (*testing*). Fitur-fitur yang digunakan dalam proses ini mencakup *Title*, *Category*, *Responsibilities*, *Minimum Qualifications*, dan *Preferred Qualification*. Pengujian bertujuan mengukur kemampuan model dalam memprediksi kelas atau label dari data uji. Evaluasi ini memberikan gambaran tentang kehandalan model dan kemampuannya dalam menggeneralisasi informasi dari deskripsi pekerjaan baru, membantu mengidentifikasi keterampilan yang dibutuhkan untuk berbagai posisi pekerjaan.

2.6. Klasifikasi Naive Bayes

Naive Bayes adalah metode klasifikasi yang didasarkan pada teorema probabilitas *Bayes*. Dalam konteks sistem analisis kebutuhan tenaga kerja, *Naive Bayes* digunakan untuk memprediksi kelas atau label dari data postingan lowongan pekerjaan berdasarkan kebutuhan keterampilan yang telah ditentukan selama tahap pelatihan. Penerapan metode ini melibatkan perhitungan probabilitas untuk masing-masing kelas yang mungkin diberikan atribut tertentu dari data. Formula utama yang digunakan dalam *Naive Bayes* adalah:

$$P(C | X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Di mana:

- $P(C | X)$ adalah probabilitas kelas C diberikan atribut X
- $P(X | C)$ adalah probabilitas atribut X diberikan kelas C
- $P(C)$ adalah probabilitas awal dari kelas C
- $P(X)$ adalah probabilitas awal dari kelas X

Dalam kasus ini, atribut X terdiri dari informasi pekerjaan seperti *Title*, *Category*, *Responsibilities*, *Minimum Qualifications*, dan *Preferred Qualifications* yang diambil dari *dataset* yang dikumpulkan. Proses klasifikasi dimulai dengan *preprocessing data*, termasuk pembersihan teks, tokenisasi, normalisasi, penghapusan *stopword*, dan *stemming* untuk menghasilkan representasi data yang bersih dan seragam.

Selanjutnya, data dikonversi ke dalam format numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF mengukur seberapa penting sebuah kata dalam dokumen tertentu relatif terhadap seluruh dataset, memungkinkan identifikasi kata-kata kunci yang relevan untuk klasifikasi.

Setelah *preprocessing* dan ekstraksi fitur, data dibagi menjadi dua set: data pelatihan dan data pengujian. Model *Naive Bayes* dilatih menggunakan data pelatihan untuk mempelajari distribusi probabilitas dari fitur-fitur yang diberikan masing-masing kelas. Model ini kemudian diuji menggunakan data pengujian untuk mengevaluasi akurasi dalam memprediksi kelas berdasarkan deskripsi pekerjaan. Dengan menggunakan model ini, dapat diidentifikasi secara otomatis keahlian dan keterampilan yang diperlukan oleh perusahaan dan memudahkan dalam menyaring dan mengelompokkan informasi terkait kebutuhan tenaga kerja.

3. Hasil dan Pembahasan

3.1. Hasil Evaluasi Klasifikasi Data

Evaluasi klasifikasi menggunakan data keseluruhan, yang terdiri dari 1.235 data yang telah dilabeli. Dari data keseluruhan, 1.212 google dan 23 youtube dilabeli. Jumlah tersebut ditunjukkan pada Gambar 4.



Gambar 4. Histogram

Dari 1.235 data yang diklasifikasikan, Google mendapatkan lebih banyak sentimen daripada *YouTube*. Hal ini disebabkan oleh kekuatan teknologi *Google*, yang menawarkan sejumlah produk dan layanan, seperti mesin pencari *Google*, *Google Maps*, *Gmail*, dan *Google Drive*.

Untuk mengevaluasi performa model klasifikasi, digunakan beberapa parameter pengujian utama seperti akurasi, presisi, *recall*, dan *F1-score*. Akurasi mengukur persentase prediksi yang benar dari total data yang diuji. Presisi mengukur berapa banyak dari data yang diprediksi sebagai kelas tertentu yang benar-benar termasuk dalam kelas tersebut. *Recall* mengukur berapa banyak dari data yang benar-benar termasuk dalam kelas tertentu yang berhasil diidentifikasi dengan benar oleh model. *F1-score* merupakan harmonisasi rata-rata dari presisi dan *recall*, memberikan gambaran yang lebih seimbang tentang performa model.

Selain itu, *Confusion Matrix* digunakan untuk memberikan visualisasi detail dari hasil klasifikasi, menunjukkan jumlah *true positives* (TP), *true negatives* (TN), *false positives* (FP), dan *false negatives* (FN). Dari *Confusion Matrix* ini, dapat dianalisis secara lebih mendalam mengenai kesalahan prediksi yang terjadi, baik kesalahan tipe I (*false positive*) maupun kesalahan tipe II (*false negative*). Evaluasi menyeluruh ini membantu dalam mengidentifikasi kekuatan dan kelemahan model, serta memberikan panduan untuk perbaikan dan pengoptimalan lebih lanjut.

3.2. Hasil Evaluasi Testing Data

Hasil evaluasi klasifikasi untuk data pengujian dilakukan melalui *cross-validation* dari perhitungan *confusion matrix*. Tujuan dari evaluasi ini adalah untuk menghitung nilai akurasi dan menentukan perbedaan antara data pelatihan dan pengujian. Data pengujian akan mencakup 1.212 data kebutuhan tenaga kerja yang telah diberi kelas berdasarkan perusahaan *Google* dan 12 label data kebutuhan tenaga kerja berdasarkan perusahaan *YouTube*.

Metode *k-fold cross-validation* dipilih untuk memastikan bahwa model dievaluasi secara menyeluruh dan hasilnya dapat digeneralisasi dengan baik. *k-fold cross-validation* membagi data menjadi *k* subset yang sama besar, kemudian menjalankan *k* iterasi pelatihan dan pengujian di mana setiap subset digunakan sekali sebagai data pengujian dan *k-1 subset* lainnya digunakan sebagai data pelatihan. Dalam penelitian ini, nilai *k* yang dipilih adalah 10 karena *10-fold cross-validation* merupakan standar yang umum digunakan dalam evaluasi model *machine learning*. Ini memberikan keseimbangan yang baik antara bias dan varians dalam hasil evaluasi, serta memastikan setiap data digunakan untuk pengujian. Pemilihan nilai *k* dapat berdasarkan kebutuhan riset. Menurut penelitian Irawan *et al.*, (2024) menggunakan *k = 10* untuk mendapatkan akurasi optimal.

Tabel 1 menunjukkan hasil perhitungan akurasi dan *F-measure* dari setiap *fold* dalam *10-fold cross-validation*. Berdasarkan Tabel 1, hasil evaluasi klasifikasi menunjukkan akurasi yang konsisten tinggi, yaitu sekitar 98-99%, dengan *F-measure* juga berada di kisaran 97-98%.

Tabel 1. Hasil Perhitungan *Confusion Matrix*

<i>10 K Fold Validation</i>	<i>Accuracy</i>	<i>F-measure</i>
Fold 1	98%	97%
Fold 2	98%	97%
Fold 3	98%	97%
Fold 4	98%	97%
Fold 5	98%	97%
Fold 6	98%	97%
Fold 7	98%	97%
Fold 8	98%	97%
Fold 9	99%	98%
Fold 10	99%	98%

Penggunaan *10-fold cross-validation* dalam penelitian ini bertujuan untuk memaksimalkan penggunaan data yang tersedia dan mengurangi kemungkinan *overfitting*. Hasil evaluasi menunjukkan bahwa model klasifikasi yang digunakan mampu memberikan prediksi yang sangat akurat dengan nilai akurasi rata-rata sebesar 98% dan *F-measure* sebesar 97%. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengidentifikasi kebutuhan tenaga kerja berdasarkan deskripsi pekerjaan yang ada. Dengan hasil ini, kita dapat menyimpulkan bahwa model *Naive Bayes* yang dikembangkan dalam penelitian ini cukup andal untuk digunakan dalam klasifikasi kebutuhan tenaga kerja.

3.3. Hasil Wordcloud

Setelah data requirements keterampilan diproses menggunakan metode-metode yang telah disebutkan sebelumnya, setiap kata yang termasuk dalam data *requirements* keterampilan tersebut divisualisasikan dalam bentuk *wordcloud*. Visualisasi ini berfungsi untuk menunjukkan frekuensi kemunculan setiap kata, memberikan gambaran tentang keterampilan yang paling dibutuhkan oleh perusahaan.



Gambar 5. Wordcloud Google Company

"Experi", "Analyst", "Project", "Manag", "Comput Scienc", "Commun", "Skill", dan "Abil", adalah kata-kata yang sering digunakan oleh *Google Company*, seperti yang ditunjukkan pada Gambar 5. Maka dari itu dapat disimpulkan inti dari seluruh kata tersebut berkaitan dengan kualifikasi atau kemampuan yang diinginkan oleh perusahaan *Google*. Secara khusus, perusahaan *Google* mungkin sedang mencari individu yang memiliki pengalaman (*experience*), kemampuan analitis (*analytical*), kemampuan manajemen proyek (*project management*), latar belakang dalam ilmu komputer (*computer science*), kemampuan komunikasi yang baik (*communication*), serta keterampilan umum (*skill*) dan kemampuan (*ability*). Ini menunjukkan fokus pada aspek-aspek ini sebagai bagian dari kriteria pencarian bakat yang diinginkan oleh *Google*.



Gambar 6. Wordcloud YouTube Company

Berdasarkan kata-kata yang sering muncul seperti "Experi", "Analyt", "Data", "Manag", "Strong", "Commun", "market", "knowledge", "Collabor", dan "effect" pada wordcloud untuk *YouTube Company* yang ditunjukkan pada Gambar 6, dapat disimpulkan bahwa inti dari seluruh kata tersebut berkaitan dengan kualifikasi atau karakteristik yang diinginkan oleh perusahaan *YouTube*. Kemungkinan, perusahaan *YouTube* sedang mencari individu yang memiliki pengalaman (*experience*), kemampuan analitis (*analytical*), pemahaman data (*data*), kemampuan manajemen (*management*), kekuatan dalam kemampuan kerja (*strong*), kemampuan komunikasi yang baik (*communication*), pemahaman pasar (*market*), pengetahuan (*knowledge*), kemampuan berkolaborasi (*collaboration*), dan efektivitas (*effectiveness*).

Hasil *wordcloud* memberikan wawasan tentang keterampilan yang paling sering dibutuhkan oleh perusahaan berdasarkan frekuensi kata. Namun, untuk memahami lebih dalam dan mengelompokkan data secara sistematis, analisis klasifikasi diperlukan. Analisis klasifikasi membantu dalam mengkategorikan postingan pekerjaan ke dalam kelas-kelas tertentu berdasarkan keterampilan yang dibutuhkan, seperti yang ditentukan oleh model *Naive Bayes*.

4. Kesimpulan

Penelitian ini memberikan gambaran mendalam tentang proyeksi kebutuhan tenaga kerja di sektor teknologi, dengan fokus pada data pengangguran di Indonesia yang mencapai 7,86 juta orang pada Agustus 2023. Pandemi Covid-19 telah mengubah paradigma kerja menjadi lebih fleksibel dengan meningkatnya pekerjaan *remote*, membuat kebutuhan informasi pekerjaan yang akurat semakin krusial. Analisis menggunakan *Naive Bayes Classification* menyoroti keterampilan yang dibutuhkan di era digital ini, dengan mayoritas permintaan keterampilan cenderung kepada *soft skills* seperti komunikasi dan kolaborasi serta *hard skills* seperti kemampuan yang berkaitan dengan digitalisasi, dengan akurasi model mencapai 98%. Visualisasi data menggunakan *word cloud* memperkuat hasil klasifikasi, menunjukkan kata-kata kunci seperti "experience", "analytical", "project management", dan "computer science" untuk *Google*, serta "data", "management", "market", dan "collaboration" untuk *YouTube*. Integrasi hasil klasifikasi dan *word cloud* ini memberikan panduan berharga untuk pengembangan sumber daya manusia, membantu individu mengembangkan keterampilan sesuai tuntutan pasar kerja, serta memberikan wawasan bagi pembuat kebijakan dalam merancang program pendidikan dan pelatihan yang efektif. Proyeksi ini membantu mengarahkan langkah strategis untuk mengurangi tingkat pengangguran dan meningkatkan kesesuaian antara keterampilan individu dan kebutuhan industri. Hal ini menunjukkan bagaimana metode klasifikasi *Naive Bayes* dan visualisasi *word cloud* dapat digunakan untuk menganalisis dan memproyeksikan kebutuhan keterampilan di sektor teknologi, serta memberikan solusi konkret dalam menjawab masalah pengangguran dan kesenjangan keterampilan di era digital.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada semua orang yang telah membantu menyusun jurnal ini, terutama kepada tim penulis yang sudah bekerja keras dengan penuh dedikasi untuk menyelesaikannya. sehingga menghasilkan laporan yang informatif dan bermanfaat. Kepuasan dalam pencapaian ini tidak terlepas dari dukungan, bimbingan, dan kolaborasi yang berharga dari semua pihak terlibat. Semua kontribusi ini berkontribusi pada pembentukan karya yang diharapkan penulis akan memberikan nilai tambahan kepada pembaca dan pemangku kepentingan. Kami ingin mengucapkan terima kasih kepada seluruh tim atas kerja sama yang luar biasa dan dedikasinya.

Referensi

- [1] A. V. D. Sano, A. A. Stefanus, E. D. Madyatmadja, H. Nindito, A. Purnomo, and C. P. M. Sianipar, "Proposing a visualized comparative review analysis model on tourism domain using Naive Bayes classifier," *Procedia Computer Science*, vol. 227, pp. 482–489, Jan. 2023, doi: 10.1016/j.procs.2023.10.549.
- [2] M. A. Valle, S. Varas, and G. A. Ruz, "Job performance prediction in a call center using a naive Bayes classifier," *Expert Systems With Applications*, vol. 39, no. 11, pp. 9939–9945, Sep. 2012, doi: 10.1016/j.eswa.2011.11.126.
- [3] W. H. Purba, P. Poningsih, D. Suhendro, I. S. Damanik, and I. S. Saragih, "Penerapan Algoritma Naive Bayes pada Penentuan Kelayakan Calon Tenaga Kerja Indonesia," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 1, p. 771, Sep. 2019, doi: 10.30645/senaris.v1i0.83.

- [4] A. V. D. Sano, A. A. Stefanus, E. D. Madyatmadja, H. Nindito, A. Purnomo, and C. P. M. Sianipar, "Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier," *Procedia Computer Science*, vol. 227, pp. 482–489, Jan. 2023, doi: 10.1016/j.procs.2023.10.549.
- [5] R. F. Akbar, M. Habibi, P. W. Cahyo, and N. A. Sa'diya, "Metode Hybrid Menggunakan Pendekatan Lexicon Based dan Naive Bayes Classifier Untuk Analisis Sentimen Terkait Jaminan Hari Tua," *Teknomatika*, vol. 16, no. 2, pp. 73–79, Dec. 2023, doi: 10.30989/teknomatika.v16i2.1247.
- [6] P. P. Aziztiya, M. Habibi, and N. I. Kusumaningtyas, "Analisis Sentimen Berdasarkan Topik Terkait Wabah Covid-19 di Twitter Menggunakan Latent Dirichlet Allocation (LDA) dan Naive Bayes Classifier (NBC)," *Teknomatika*, vol. 15, no. 2, pp. 76–85, Oct. 2022, doi: 10.30989/teknomatika.v15i2.1098.
- [7] A. Wahyudi, M. B. T. Assyamiri, W. A. Aluf, M. R. Fadhillah, S. Yolanda, and M. I. Anshori, "Dampak transformasi era digital terhadap manajemen sumber daya manusia," *Jubima*, vol. 1, pp. 99–111, Oct. 2023, doi: 10.55606/jubima.v1i4.2222.
- [8] K. Bakhsar, W. Saputra, dan H.S. Tambunan, "Penerapan Algoritma K-Means Dalam Pengelompokan Pekerja Tetap Perusahaan Konstruksi Menurut Provinsi," *Kesatria*, vol. 1, pp. 146–153, Oct. 2020, doi: 10.30645/kesatria.v1i4.40.
- [9] S. Sulistyanto, F. Mutohhari, A. Kurniawan, dan D. Ratnawati, "Kebutuhan kompetensi di era revolusi industri 4.0: review perspektif pendidikan vokasional," *Jurnal Taman Vokasi*, vol. 9, pp. 25–35, Juni. 2020, doi: 10.30738/jtv.v9i1.7742.
- [10] S. A. Harahap, R. Aprilia, and R. S. Lubis, "PENERAPAN METODE NAÏVE BAYES DALAM KLASIFIKASI KABUPATEN/KOTA DI SUMATERA UTARA BERDASARKAN PRODUKTIVITAS PANGAN PADI," *Jurnal Lebesgue*, vol. 4, no. 1, pp. 79–90, Feb. 2023, doi: 10.46306/lb.v4i1.202.
- [11] J.M.Br. Sembiring, dan Hendry, "NAÏVE BAYES ALGORITHM CLASSIFICATION IN SENTIMENT ANALYSIS COVID-19 WIKIPEDIA," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, No. 4, pp. 869-875, Agust. 2022, doi: 10.20884/1.jutif.2022.3.4.311.
- [12] S. Mulyani and R. Novita, "IMPLEMENTATION OF THE NAIVE BAYES CLASSIFIER ALGORITHM FOR CLASSIFICATION OF COMMUNITY SENTIMENT ABOUT DEPRESSION ON YOUTUBE," *Jurnal Teknik Informatika*, vol. 3, no. 5, pp. 1355–1361, Oct. 2022, doi: 10.20884/1.jutif.2022.3.5.374.
- [13] P. W. Rahayu, I. G. I. Sudipa, Suryani, and I. M. D. M. Sanjaya, "BUKU AJAR DATA MINING," *Research Gate*, Jan. 2024, [Online]. Available: https://www.researchgate.net/publication/377415198_BUKU_AJAR_DATA_MINING
- [14] S. Supangat, M. Z. B. Saringat, and M. Y. F. Rochman, "Predicting Handling Covid-19 Opinion using Naive Bayes and TF-IDF for Polarity Detection," *Matrik*, vol. 22, no. 2, pp. 173–184, Mar. 2023, doi: 10.30812/matrik.v22i2.2227.
- [15] Irawan, R. N., Hindrayani, K. M., & Idhom, M. (2024). Penerapan Cross Validation sebagai Analisis Sentimen Pelayanan Publik Kereta Api Lokal Daop 8 Menggunakan Metode Multinomial Naïve Bayes. *G-Tech: Jurnal Teknologi Terapan*, 8(2), 954–963. <https://doi.org/10.33379/gtech.v8i2.4117>
- [16] W. I. Rahayu, C. Prianto and E. A. Novia, "PERBANDINGAN ALGORITMA K-MEANS DAN NAÏVE BAYES UNTUK MEMPREDIKSI PRIORITAS PEMBAYARAN TAGIHAN RUMAH SAKIT BERDASARKAN TINGKAT KEPENTINGAN PADA PT. PERTAMINA (PERSERO)", *Jurnal Teknik Informatika*, Vol. 13, No. 2, April. 2021
- [17] N. Ramdhani, R. H. Al-Fadillah, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP BELAJAR DARING SELAMA PANDEMI COVID-19 DENGAN DEEP LEARNING," *Jurnal Siliwangi*, vol.7, no.2, Dec. 2021, doi: 10.37058/jssainstek.v7i2.4281.
- [18] World Economic Forum, "The Future of Jobs Report 2023," World Economic Forum, 2023.
- [19] Asian Development Bank, "Reducing Youth Not in Employment, Education, or Training through Career Coaching and Internships," Asian Development Bank, 2023.