

SISTEM KERJACERDAS : PLATFORM PENCARIAN KERJA BERASASKAN AI

¹Nur Izzati Binti Rasman¹ (A194878), Ts. Dr. Nor Samsiah Sani²

*Ijazah Sarjana Muda Sains Komputer dengan Kepujian, Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia*

A194878@siswa.ukm.edu.my

ABSTRAK

KerjaCerdas ialah sistem pintar yang menggunakan teknologi Pemprosesan Bahasa Semulajadi (NLP) untuk memadankan pencari kerja dengan majikan berdasarkan kebolehan dan pengalaman mereka. Sistem ini menggunakan model seperti BERT dan Word2Vec untuk memahami konteks dan makna teks dalam iklan pekerjaan dan resume. Untuk melakukan analisis yang lebih mendalam tentang kesesuaian antara keperluan pekerjaan dan profil pencari kerja, pendekatan ini dilengkapi dengan kaedah seperti Cosine Similarity dan Jaccard Similarity. Pendekatan model metodologi CRISP-DM digunakan dalam proses pembangunan sistem ini. Proses ini termasuk pengumpulan data daripada pelbagai sumber, pembersihan data untuk memastikan maklumat itu tepat, pembangunan model pintar, dan evaluasi keberkesanan sistem. Platform ini dijangka dapat menyediakan padanan kerja yang lebih relevan dan tepat serta mempercepatkan proses pencarian kerja, dan memudahkan majikan mencari calon yang memenuhi keperluan mereka kerana ia menggabungkan teknologi NLP dan kaedah analisis kesamaan. Secara keseluruhan, matlamat KerjaCerdas adalah untuk menjadikan proses mencari pekerjaan dan pengambilan pekerja lebih kreatif dan berkesan. Sistem KerjaCerdas ini dijangka boleh menjadi penyelesaian berkesan untuk meningkatkan pengalaman dan kecekapan dalam proses pencarian dan pengambilan pekerja melalui pendekatan inovatif.

Kata Kunci: KerjaCerdas, NLP, BERT, Word2Vec, Cosine Similarity, Jaccard Similarity, CRISP-DM, padanan kerja, pencari kerja, majikan

PENGENALAN

Di era digital yang pesat berkembang, analitik data telah menjadi alat penting dalam memudahkan proses pencarian pekerjaan dan meningkatkan kecekapan pencocokan antara pencari kerja dan majikan. Melalui penggunaan analitik data yang berasaskan kecerdasan buatan (AI), pelbagai data boleh dianalisis untuk mengenal pasti peluang pekerjaan yang relevan dengan profil individu (A. Drury, 2024). Sistem seperti KerjaCerdas yang menggunakan teknologi AI dapat memproses data daripada pelbagai sumber, termasuk media sosial, laman web pekerjaan, dan profil pencari kerja, bagi menghasilkan cadangan pekerjaan yang lebih tepat dan bersesuaian. Menurut Erin Stertz-Follett (2022), ini bukan sahaja mengurangkan masa yang diambil oleh pencari kerja dalam menelusuri iklan pekerjaan yang banyak, tetapi juga membantu majikan dalam mengenal pasti calon yang memenuhi kriteria dengan lebih berkesan. Melalui analitik data, KerjaCerdas dapat memberikan cadangan yang disesuaikan berdasarkan pengalaman, kemahiran, dan keutamaan pengguna, menjadikan proses pencarian pekerjaan lebih berfokus dan berimpak tinggi dalam suasana persaingan global yang kian mencabar.

KerjaCerdas ini direka bertujuan untuk menangani isu dalam pencarian pekerjaan, seperti kesukaran mengenal pasti peluang pekerjaan yang sesuai dan cabaran dalam menilai sejauh mana calon memenuhi keperluan majikan. Platform ini boleh memberikan cadangan pekerjaan yang lebih sesuai berdasarkan pengalaman, kemahiran, dan keutamaan pencari kerja melalui algoritma AI yang canggih. Hal ini dapat meningkatkan pengalaman pengguna secara keseluruhan dan mempercepatkan proses pencarian tanpa perlu menggunakan tenaga dan masa. Menurut Zaky Zainuddin (2023), masalah ekonomi negara yang kian membimbangkan membuka mata pihak tertentu untuk meningkatkan peluang pekerjaan dalam usaha untuk menangani masalah inflasi dan kenaikan kos sara hidup. Oleh yang demikian, sistem KerjaCerdas ini akan membantu dan memudahkan penggunanya untuk mencari pekerjaan seperti yang dicari. Sebagai contoh, pengguna ingin mencari pekerjaan yang fleksibel, dimana pekerja boleh bekerja mengikut hari dan masa yang mereka boleh bekerja. Menurut Akta 2

Pekerjaan (Pindaan) 2022, untuk mewujudkan persekitaran kerja yang lebih baik di pelbagai industry, pihak berwajib perlu memasukkan peruntukan anti-diskriminasi dan fleksibiliti waktu kerja.

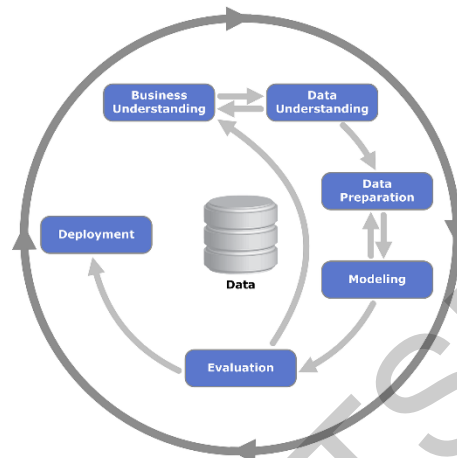
Sistem KerjaCerdas juga mempunyai keupayaan untuk merapatkan hubungan antara individu yang sedang mencari pekerjaan dan Syarikat. Kelebihan utama KerjaCerdas ini adalah membolehkan pekerja dan majikan berhubung secara langsung, berkongsi maklumat dan membina hubungan yang lebih kukuh dengan menyediakan lebih banyak ruang untuk berinteraksi. Pekerja dan majikan juga dapat membincangkan cara untuk menambah baik insentif bagi mewujudkan keseimbangan kerja dan hidup sekaligus memberi kelebihan kepada kedua-dua pihak (M. Fauzi, 2024). Menurut data World Health Organization (WHO), pekerja yang mempunyai keseimbangan hidup dan kerja yang baik mempunyai risiko 30% yang lebih rendah terhadap kemurungan dan kebimbangan. Hal ini penting dalam persekitaran kerja yang sentiasa berubah, di mana keperluan dan jangkaan kerap berubah. KerjaCerdas juga menyediakan panduan resume, nasihat temuduga, dan latihan kemahiran yang membantu pengguna dalam proses mencari pekerjaan. Hal ini dapat memastikan pengguna bukan sahaja mendapatkan pekerjaan, tetapi mereka juga dipersiapkan untuk menghadapi cabaran di tempat kerja. Kelebihan bagi majikan pula, mereka dapat menyaring dan mengenal pasti calon yang paling sesuai dengan keperluan syarikat mereka

METODOLOGI

Perancangan Awal Projek

Bagi projek KerjaCerdas ini, saya memilih untuk menggunakan Metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining). Terdapat beberapa faktor yang menyebabkan saya memilih Model tersebut bagi menyiapkan projek KerjaCerdas saya. Model CRISP-DM merupakan model yang efektif dan mudah untuk digunakan bagi menyelesaikan projek KerjaCerdas ini secara efektif dan mengikut masa yang ditetapkan. Yang pertama, model ini merupakan metod yang sangat mudah dan penting bagi memastikan sesebuah projek dapat dianalisis dengan teliti serta berjalan dengan teratur. Model ini merupakan salah satu metod yang fleksibel kerana prosesnya berlaku

secara berulang dan bertukar mengikut kesesuaian projek. Bagi projek yang revelansi yang sering berubah seperti KerjaCerdas, model CRISP-DM adalah yang paling sesuai.



Rajah 1 Model Metodologi (CRISP-DM)

Sumber: P. Syafruddin, 2021

Fasa Pemahaman dan Pengumpulan Data

Fasa awal pembangunan sistem bermula dengan proses pemahaman domain masalah. Sistem ini menunjukkan bagaimana sistem ini membantu dalam menawarkan pekerjaan yang sesuai dengan resume pengguna. Data dikumpul daripada dua sumber utama: resume pengguna yang akan dimuat naik dan dataset iklan pekerjaan (clean_jobposting_data.csv). Semua maklumat penting seperti tajuk, industri, lokasi dan deskripsi tugas telah dibersihkan dan diproses dalam data pekerjaan. Sebagai tambahan kepada tanda baca, penukaran huruf ke huruf kecil, dan normalisasi teks, Pandas dan NLTK telah digunakan.

```

# --- Fill missing values ---
jobstreet_df.fillna({
    'title': '', 'company': '', 'description': '',
    'location': '', 'industry': '', 'salary': '0'
}, inplace=True)

job_postings_df.fillna({
    'title': '', 'company': '', 'location': '', 'industry': ''
}, inplace=True)

# --- Convert salary to numeric ---
jobstreet_df['salary'] = pd.to_numeric(jobstreet_df['salary'], errors='coerce').fillna(0)

# --- Encode 'industry' from both datasets ---
le = LabelEncoder()

combined_industries = pd.concat([
    jobstreet_df['industry'].astype(str),
    job_postings_df['industry'].astype(str)
])

le.fit(combined_industries)

jobstreet_df['industry_encoded'] = le.transform(jobstreet_df['industry'].astype(str))
job_postings_df['industry_encoded'] = le.transform(job_postings_df['industry'].astype(str))

# --- Compare job titles between the two datasets ---
titles_jobstreet = set(jobstreet_df['title'].str.lower().str.strip())
titles_postings = set(job_postings_df['title'].str.lower().str.strip())

# Common job titles
common_titles = titles_jobstreet.intersection(titles_postings)
print("\n# Number of common job titles (len(common_titles))")
print("Sample common job titles:", list(common_titles)[:10])

# Unique to JobStreet
unique_jobstreet = titles_jobstreet - titles_postings
print("\n# Titles only in JobStreet: (len(unique_jobstreet))")
print("Sample:", list(unique_jobstreet)[:10])

# Unique to JobPostings
unique_postings = titles_postings - titles_jobstreet
print("\n# Titles only in JobPostings: (len(unique_postings))")
print("Sample:", list(unique_postings)[:10])

# Optional: Percentage overlap
percent_overlap = len(common_titles) / len(titles_jobstreet) * 100 if len(titles_jobstreet) > 0 else 0
print("\n# Percentage of JobStreet titles also in JobPostings: (percent_overlap.2f)%"

```

Rajah 2 Pemahaman dan Pembersihan Data

	A	B	C	D	E	F	G	H	I	J	K	L
1	Unnamed: 0	job_id	job_title	company	descriptions	location	category	subcategory	type	salary		
2	0	72761527	Data Engineer	ANHSHIN TE	Design, develop,	Kuala Lumpur	Science & Mathematics	Full time		RM 5,000 - RM 7,000 per month		
3	1	72787241	Machine Learning Engineer (AI)	Accordia C	Design, develop,	Shah Alam	Science & Mathematics	Full time		RM 5,000 - RM 7,000 per month		
4	2	72866732	Senior Risk/Data Analyst	Toyota Car	Analyse data to	Petaling	Banking & Compliance	Full time		RM 5,000 - RM 7,000 per month		
5	3	72851872	Senior Data Engineer (Hybrid Working)	SEEK	Design,	Kuala Lumpur	Information Engineering	Full time		RM 5,000 - RM 7,000 per month		
6	4	72526811	Data Scientist (Hybrid Working)	SEEK Asia	(Research, build,	Kuala Lumpur	Information Developer	Full time		RM 5,000 - RM 7,000 per month		
7	5	72762762	Senior Engineer / Engineer - Data Managen	B. Braun M	Build, maintain	Penang	Engineering Other	Full time		RM 5,000 - RM 7,000 per month		
8	6	72799675	Data Assistant	Wheeler's	Running Reports	Bangsar	Sc Administration	Data Entry	Full time	RM 5,000 - RM 7,000 per month		
9	7	72762141	Executive - Business Intelligence / SAS Ana	Berjaya So	Fulfil requests	Kuala Lumpur	Insurance Actuarial	Full time		RM 5,000 - RM 7,000 per month		
10	8	72851569	Senior Product Data Analyst (Hybrid Workin	SEEK	Act as an	Kuala Lumpur	Information Business	Full time		RM 5,000 - RM 7,000 per month		
11	9	72720034	Software Business Analyst	AIA IT (M)	S Liaise with	Kuala Lumpur	Information Business	Full time		RM 5,000 - RM 7,000 per month		
12	10	72844633	Associate Process Engineer (All Modules)	GlobalFour	Review & analyze	Penang	Information Engineering	Process Er	Full time	RM 5,000 - RM 7,000 per month		
13	11	72750107	IT Business Analyst	Mewah Grn	To develop new	Klang Distr	Information Business	Full time		RM 5,000 - RM 7,000 per month		
14	12	72693889	Contract Business Administration Analyst	RGA Globa	Use RGA	Kuala Lumpur	Science & Mathematics	Contract		RM 4,000 - RM 4,500 per month		
15	13	72871004	Data Scientist	Private Ad	What's your	Kuala Lumpur	Science & Mathematics	Full time		RM 5,000 - RM 7,000 per month		
16	14	72814199	Data Scientist	Continents	Working with	Petaling	Information Database	Full time		RM 5,000 - RM 7,000 per month		
17	15	72867189	Data Engineer	FARMBYTE	Implement and	Kuala Lumpur	Science & Mathematics	Full time		RM 4,000 - RM 6,000 per month		
18	16	72878751	Data Analyst	Taylor's Education Group		Petaling	Consulting Analysts	Full time		RM 5,000 - RM 7,000 per month		
19	17	72696058	DATA ANALYST	LYNX SOLL	Handle	Kuala Lumpur	Marketing, Market Res	Full time		RM 4,000 - RM 4,500 per month		
20	18	72863386	Data Analysts/ Data Engineers (Bangsar So	AG Staffing	Data Collection	Kuala Lumpur	Science & Mathematics	Full time		RM 8,000 - RM 12,000 per month		
21	19	72382969	Machine Learning Engineer (Artificial Intell	FootballCa	Design, develop,	Kuala Lumpur	Science & Mathematics	Full time		RM 4,000 - RM 6,000 per month		
22	20	72856382	Data Analyst	Daikin Mal	Collect, process,	Petaling	Information Business	Full time		RM 4,000 - RM 6,000 per month		
23	21	72753208	Data Architect	CIMB Grou	Act as Data	Kuala Lumpur	Information Architects	Full time		RM 5,000 - RM 7,000 per month		
24	22	72869475	IT Data Engineer (Semiconductor)	Robert Bos	As an IT Data	Penang	Information Database	Full time		RM 5,000 - RM 7,000 per month		
25	23	72854137	Data Analyst	Holiday To	Interpret data,	Kuala Lumpur	Science & Mathematics	Full time		RM 5,000 - RM 7,000 per month		
26	24	72666993	Data Engineer	ISJ TECHN	Develop data	Petaling	Science & Mathematics	Full time		RM 5,000 - RM 7,500 per month		
27	25	72687269	Intern - Data Science	LAM RESE	/ Based under	Seberang	Information Science & Mathematics	Full time		RM 1,500 per month		

Rajah 3 Dataset clean_jobposting_data.csv

Fasa Pembangunan Model

Pendekatan *TF-IDF*, *Cosine Similarity* digunakan untuk membangunkan model cadangan kerja pintar. Bukan itu sahaja, *Jaccard Similarity*, *BERT Embeddings*, dan *Content-Based Filtering* juga digunakan dalam fasa pembangunan model. sistem ini turut menguji penggunaan algoritma XGBoost sebagai model tambahan bagi penilaian relevansi pekerjaan. Ciri-ciri input seperti skor keserupaan, perbezaan lokasi, dan padanan industri digunakan untuk melatih model klasifikasi menggunakan XGBoost. Model ini menunjukkan potensi dalam meningkatkan ketepatan padanan apabila dilatih dengan dataset yang lebih besar dan disokong oleh anotasi pemadanan sebenar. *TF-IDF*

Vectorizer daripada pustaka *scikit-learn* digunakan untuk menghasilkan representasi vektor untuk setiap deskripsi kerja. Tambahan pula, resume yang dimuat naik diubah suai menjadi vektor dan dibandingkan dengan semua vektor kerja. Ianya dapat membina kerangka untuk menguji model yang menggunakan *BERT* di dalam *Google Collab Notebook*.

Cosine Similarity TF-IDF dan BERT

```
# =====
# 1. COSINE SIMILARITY USING TF-IDF
# =====
def cosine_similarity_tfidf(texts):
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform(texts)
    similarity_matrix = cosine_similarity(tfidf_matrix)
    return similarity_matrix

# Example usage
cosine_sim_tfidf = cosine_similarity_tfidf(sample_descriptions)
print("Cosine Similarity using TF-IDF:\n", cosine_sim_tfidf)

# =====
# 2. COSINE SIMILARITY USING BERT
# =====
# Load BERT model and tokenizer
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
model = BertModel.from_pretrained('bert-base-uncased')

# Function to get BERT sentence embedding
def get_bert_embeddings(text):
    inputs = tokenizer(text, return_tensors='pt', truncation=True, padding=True, max_length=512)
    with torch.no_grad():
        outputs = model(**inputs)
    embeddings = outputs.last_hidden_state.mean(dim=1).squeeze()
    return embeddings.numpy()

def cosine_similarity_bert(texts):
    embeddings = np.array([get_bert_embeddings(text) for text in texts])
    similarity_matrix = cosine_similarity(embeddings)
    return similarity_matrix

# Example usage
cosine_sim_bert = cosine_similarity_bert(sample_descriptions)
print("Cosine Similarity using BERT:\n", cosine_sim_bert)
```

Rajah 4 *Cosine similarity TF-IDF dan BERT*

Bahagian ini membincangkan pengiraan tahap persamaan semantik antara penerangan kerja menggunakan dua kaedah pemprosesan bahasa semula jadi (NLP) iaitu TF-IDF dengan Cosine Similarity dan BERT dengan Cosine Similarity. Teknik TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk menukar penerangan kerja kepada vektor berangka. Teknik ini mengukur kepentingan perkataan tertentu dalam dokumen berbanding keseluruhan koleksi. Selepas itu, cosine similarity dikira untuk mengetahui seberapa besar persamaan antara dokumen. Dengan nilai sederhana antara ~0.35 hingga ~0.58, hasil ini menunjukkan bahawa TF-IDF hanya boleh mengenal pasti pertindihan perkataan tetapi tidak memahami konteks ayat. Sebaliknya, BERT (Bidirectional Encoder Representations from Transformers) digunakan untuk membuat embedding ayat yang ringkas dan kontekstual. Embedding ini memasukkan makna penuh ayat, dan ia mendapat skor yang tinggi dan konsisten antara 0.87 dan 1.0 apabila dibandingkan menggunakan cosine similarity. Ini menunjukkan bahawa BERT boleh memahami maksud sebenar teks dalam pelbagai cara. Untuk sistem KerjaCerdas, ciri

ini sangat penting kerana penerangan kerja dan resume mungkin menggunakan perkataan yang sama.

Jaccard Similarity TF-IDF dan BERT

```

# =====
# 1. TF-IDF + Jaccard Similarity
# =====
print("\n== TF-IDF + Jaccard Similarity ==")
# TF-IDF Vectorizer
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(texts).toarray()
# Convert to binary for Jaccard
X_binary = (X > 0).astype(int)
# Compute pairwise Jaccard similarity
tfidf_jaccard_sim = np.zeros((len(X_binary), len(X_binary)))
for i in range(len(X_binary)):
    for j in range(len(X_binary)):
        if np.sum(np.logical_or(X_binary[i], X_binary[j])) == 0:
            tfidf_jaccard_sim[i, j] = 0.0
        else:
            tfidf_jaccard_sim[i, j] = jaccard_score(X_binary[i], X_binary[j])

print(tfidf_jaccard_sim)

# =====
# 2. BERT + Jaccard Similarity
# =====
print("\n== BERT + Jaccard Similarity (Binarized Embeddings) ==")
# Load BERT
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
model = BertModel.from_pretrained('bert-base-uncased')
# Function to get BERT sentence embedding
def get_bert_embeddings(text):
    inputs = tokenizer(text, return_tensors='pt', truncation=True, padding=True, max_length=512)
    with torch.no_grad():
        outputs = model(**inputs)
    return outputs.last_hidden_state.mean(dim=-1).squeeze().numpy()
# Get and binarize BERT embeddings
bert_vectors = [get_bert_embeddings(text) for text in texts]
bert_binary = (np.array(bert_vectors) > 0).astype(int) # Threshold: > 0
# Compute Jaccard similarity for BERT binary vectors
bert_jaccard_sim = np.zeros((len(bert_binary), len(bert_binary)))
for i in range(len(bert_binary)):
    for j in range(len(bert_binary)):
        if np.sum(np.logical_or(bert_binary[i], bert_binary[j])) == 0:
            bert_jaccard_sim[i, j] = 0.0
        else:
            bert_jaccard_sim[i, j] = jaccard_score(bert_binary[i], bert_binary[j])

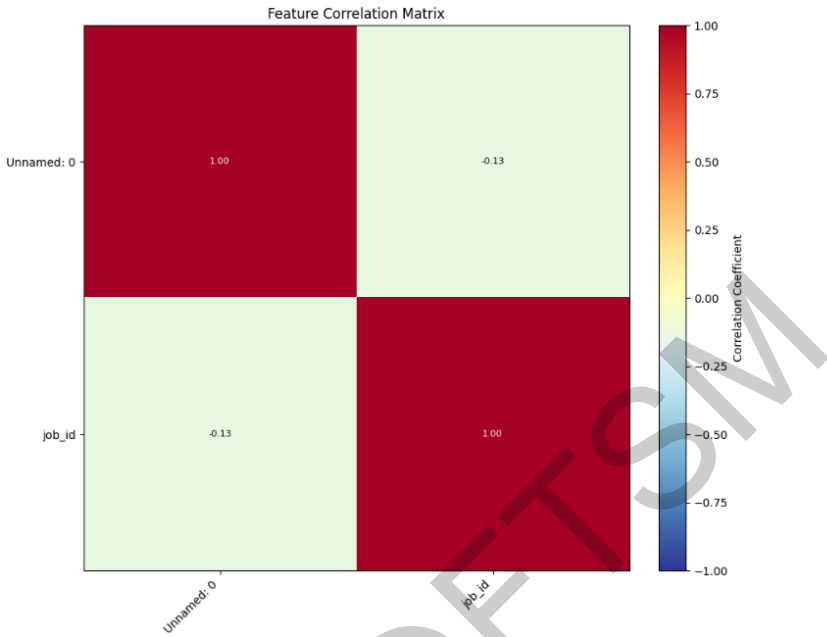
print(bert_jaccard_sim)

```

Rajah 5 Jaccard Similarity TF-IDF dan BERT

Bahagian ini membincangkan teknik Jaccard Similarity. TF-IDF diubah suai kepada bentuk binari iaitu 1 jika wujud perkataan dan 0 jika tidak, dan hubungan Jaccard dikira dengan membandingkan persilangan dan gabungan set perkataan. Metod ini hanya mengukur pertindihan literal dan tidak memahami makna teks, seperti yang ditunjukkan oleh skor yang rendah, kebanyakannya antara 0.08 dan 0.25. BERT yang telah dibinarikan dimasukkan dengan cara yang sama iaitu dengan menetapkan nilai yang lebih besar daripada 0 kepada 1. Hasilnya lebih baik dimana skor antara 0.66 dan 0.78, tetapi masih kurang daripada BERT dengan cosine similarity. Ini disebabkan oleh fakta bahawa kandungan BERT yang kaya dengan makna semantik dikurangkan oleh proses binarisasi. Oleh itu, kaedah ini tidak sesuai untuk digunakan dalam sistem seperti KerjaCerdas, yang memerlukan pepadanan yang didasarkan pada maksud sebenar teks.

Content-Based Filtering TF-IDF dan BERT



Rajah 6 Kolerasi Matriks

Bahagian ini menjalankan analisis data eksploratori dengan menghasilkan matriks korelasi antara ciri berangka seperti gaji dan pengalaman. Fungsi `corr()` digunakan untuk mengira korelasi dan heatmap digunakan untuk divisualisasikan. Hubungan antara dua pembolehubah ditunjukkan oleh setiap sel dalam matriks. Sebagai contoh, ia boleh digunakan untuk memperhalusi logik pemadanan sistem jika gaji berkorelasi dengan pengalaman. Walaupun sistem KerjaCerdas pada masa ini berasaskan teks, pemahaman tentang hubungan antara ciri berangka ini boleh membantu dalam pembinaan model hibrid pada masa akan datang yang menggunakan persamaan data berstruktur dan persamaan teks untuk membuat cadangan yang lebih tepat.

Analisis Gaji Mengikut Tahap Pengalaman

```
import pandas as pd

# Load the dataset
df = pd.read_csv('clean_jobposting_data.csv', encoding='ISO-8859-1')

# Clean the 'salary' column if necessary (e.g., remove non-numeric characters)
df['salary'] = df['salary'].replace({'[^\d.]': ''}, regex=True).astype(float)

# Define a function to categorize experience levels (fresh graduate vs experienced worker)
def categorize_experience(job_title):
    # Keywords indicating a fresh graduate position (you can modify this list based on your dataset)
    fresh_grad_keywords = ['graduate', 'entry', 'junior', 'fresh']

    # Check if any fresh grad keyword exists in the job title (case insensitive)
    if any(keyword.lower() in job_title.lower() for keyword in fresh_grad_keywords):
        return 'Fresh Graduate'
    else:
        return 'Experienced'

# Apply the function to categorize the job titles
df['experience_level'] = df['job_title'].apply(categorize_experience)

# Split the salary data based on experience level
fresh_grads_salary = df[df['experience_level'] == 'Fresh Graduate']['salary']
experienced_salary = df[df['experience_level'] == 'Experienced']['salary']

# Display the salary statistics for each group
print("Fresh Graduate Salary Stats:")
print(fresh_grads_salary.describe())

print("\nExperienced Worker Salary Stats:")
print(experienced_salary.describe())

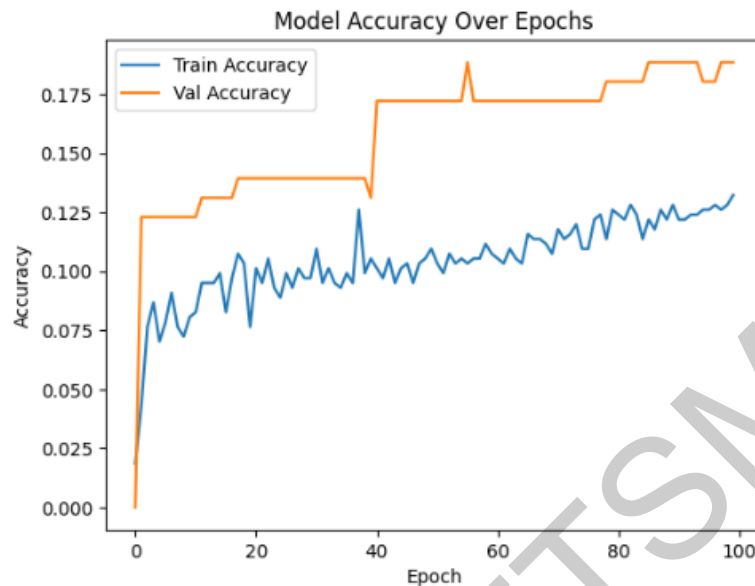
# Optionally: you can save the dataframe with the new 'experience_level' column
df.to_csv('categorized_salaries.csv', index=False)
```

Rajah 7 Analisis Gaji Mengikut Tahap Pengalaman

Bahagian terakhir ini mengkaji perbezaan gaji antara pekerja berpengalaman dan graduan baru. Untuk menentukan jawatan yang sesuai untuk graduan baru, tajuk jawatan dikategorikan menggunakan perkataan seperti "junior", "graduat", atau "entry." Kemudian, gaji dibahagikan mengikut kategori dan statistik deskriptif digunakan untuk menganalisisnya. Analisis ini masih berguna secara konseptual walaupun nilainya tidak tepat. Ia menunjukkan betapa pentingnya menyaring jawatan berdasarkan tahap pengalaman pelanggan. KerjaCerdas boleh menapis dan memaparkan jawatan peringkat permulaan jika sistem mengetahui bahawa pengguna ialah graduan baru. Ini menjadikan cadangan lebih spesifik.

Fasa Pengujian Model

Cadangan yang diberikan diuji secara manual dengan model yang dibangunkan. Beberapa percubaan telah dilakukan untuk membandingkan hasil menggunakan *TF-IDF* dan *embedding BERT*. Untuk menilai relevansi padanan, lima pekerjaan tertinggi akan dicatat.



Fasa Pemilihan dan Penerbitan Model

Cosine Similarity dengan BERT jelas menunjukkan prestasi paling konsisten dan terbaik daripada semua kaedah yang dikaji. Dengan nilai 0.87–1.00, matriks kesamaan yang dihasilkan menunjukkan bahawa model ini menangkap makna semantik dan hubungan kontekstual antara penerangan kerja dengan sangat berkesan. Kerana resume dan deskripsi kerja sering menggunakan perkataan atau frasa yang sama dalam sistem KerjaCerdas, keupayaan ini adalah penting. Sebagai contoh, frasa "membangunkan model pembelajaran mesin" dan "membina algoritma ramalan" menggunakan perkataan yang berbeza tetapi merujuk kepada kebolehan yang serupa. Hubungan penting ini dikenal pasti dengan tepat oleh BERT dan dipadankan dengan lebih bijak. Tambahan pula, pengimejan berasaskan BERT menunjukkan skor kesamaan yang tinggi, dengan kebanyakan nilai melebihi 0.60 dan beberapa hampir 0.70. Ini menunjukkan bahawa BERT boleh dengan tepat memadankan penerangan kerja dan resume pengguna dalam pangkalan data. BERT mampu memahami bahasa semula jadi lebih daripada sekadar pemadanan kata kunci kerana ia dilatih dengan banyak data teks dan berasaskan senibina rangkaian neural yang mendalam. BERT sesuai untuk sistem seperti KerjaCerdas, yang memerlukan pemadanan kerja-resume yang pintar, fleksibel dan kontekstual tanpa bergantung pada kesamaan perkataan.

Fasa Pembangunan Sistem

Akhir sekali, sistem sepenuhnya telah dibangunkan. Ia termasuk antara muka pengguna yang menggunakan HTML, CSS, dan PHP, serta backend Flask API untuk logik

cadangan kerja. Semua fungsi, termasuk muat naik resume, sesi pengguna, dan paparan hasil, telah diintegrasikan. Untuk memastikan bahawa setiap komponen berfungsi dengan baik, pengujian dijalankan dari semasa ke semasa. Sistem ini boleh ditingkatkan pada masa akan datang dengan mudah kerana struktur kodnya yang modular, yang membolehkan integrasi model NLP lanjutan atau sistem pemberitahuan pintar.

Teknologi Digunakan

Pendekatan berasaskan seni bina mikro digunakan dalam pembinaan sistem "KerjaCerdas", yang membahagikan logik pemprosesan kepada dua bahagian berbeza iaitu antara muka pengguna, yang terdiri daripada frontend dan backend PHP, dan sistem pemprosesan pintar, yang menggunakan Flask dan Python. Pendekatan ini memberikan fleksibiliti dalam pembangunan dan penyelenggaraan sistem. Ia juga memudahkan integrasi dengan model NLP yang lebih maju pada masa akan datang. Antara teknologi penting yang digunakan ialah PHP 8.0 dimana PHP 8.0 adalah bahasa pengaturcaraan sisi pelayan yang digunakan untuk mencipta logik frontend bagi sistem KerjaCerdas, seperti pengurusan pengguna, pemuatan resume dan komunikasi dengan API Flask. PHP dipilih kerana ia mudah digunakan dan sesuai untuk aplikasi web berskala sederhana sekaligus digunakan dalam sistem KerjaCerdas untuk membuat antaramuka pengguna yang mudah digunakan dan responsif. Susun atur halaman disesuaikan untuk kedua-dua peranti desktop dan mudah alih. Seterusnya, MySQL dan phpMyAdmin dimana MySQL digunakan sebagai sistem pengurusan pangkalan data untuk menyimpan maklumat pengguna dan senarai pekerjaan, manakala phpMyAdmin digunakan sebagai alat antara muka grafik untuk mengendalikan jadual, pertanyaan SQL dan penyahpejatan data secara visual. Python 3.10 ialah bahasa utama untuk model pembelajaran mesin dan pemprosesan pintar resume.

Table structure Relation view									
#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
☐	1 id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
☐	2 username	varchar(100)	utf8mb4_general_ci		No	None			Change Drop More
☐	3 email	varchar(100)	utf8mb4_general_ci		No	None			Change Drop More
☐	4 password	varchar(255)	utf8mb4_general_ci		No	None			Change Drop More

Rajah 8 phpMyAdmin bagi sistem KerjaCerdas

Python digunakan kerana ia mempunyai pustaka pembelajaran mesin dan pembelajaran NLP yang luas. Flask ialah kerangka mikro web yang boleh digunakan untuk membina API yang menerima resume daripada frontend dan memulangkan hasil padanan pekerjaan. Scikit-learn, Pandas, dan Numpy pula digunakan untuk membina model cadangan kerja. Ini termasuk pengurusan dataset dan pelaksanaan algoritma TF-IDF dan Cosine Similarity.

Proses Pembangunan Kerja

Pendekatan CRISP-DM (Cross Industry Standard Process for Data Mining) digunakan semasa proses pembangunan sistem KerjaCerdas. Pendekatan ini amat sesuai untuk projek yang melibatkan analisis data dan pemodelan pembelajaran mesin. Pemahaman perniagaan, pemahaman data, penyediaan data, pemodelan, penilaian, dan penerapan adalah enam fasa utama metodologi ini yang dilaksanakan secara berperingkat.

Fasa pertama ialah pemahaman perniagaan, di mana matlamat utama projek digariskan. Sistem cadangan KerjaCerdas akan mencari pekerjaan terbaik kepada pengguna berdasarkan resume mereka. Selain itu, dalam peringkat ini, aliran penggunaan sistem dirancang dengan teliti. Ini termasuk proses log masuk, pemuatan naik resume, proses pendaftaran pengguna dan paparan hasil padanan pekerjaan. Kemudian, dalam Fasa Pemahaman Data, dataset iklan kerja yang diperolehi daripada fail `clean_jobposting_data.csv` dianalisis untuk mencari maklumat penting seperti tajuk pekerjaan, industri, lokasi dan deskripsi tugas. Sebagai data input, resume yang dimuat naik oleh pengguna akan dibandingkan dengan iklan kerja ini. Untuk memastikan proses padanan terhadap ciri yang relevan, adalah penting untuk memahami data ini. Proses pembersihan dan normalisasi data dilakukan pada fasa penyediaan data menggunakan pustaka Python seperti Pandas dan NLTK. Ini termasuk membahagikan data untuk latihan dan penilaian, mengeluarkan tanda baca dan kata-kata tidak penting (stopwords), dan menukar semua teks ke huruf kecil. Untuk memastikan pencocokan yang tepat dan konsisten, proses pembersihan juga digunakan untuk resume yang dimuat naik.

Fasa pemodelan melibatkan penciptaan fungsi padanan utama dalam API Flask yang dipanggil `get_top_matches()`. Algoritma TF-IDF digunakan dalam model ini untuk

menukar teks kepada bentuk vektor, dan Cosine Similarity digunakan untuk menentukan sejauh mana persamaan antara resume dan setiap deskripsi kerja. Sebagai hasil, lima pekerjaan dengan persamaan tertinggi akan dipilih dan dihantar kembali kepada sistem PHP dalam format JSON. Selain itu, LabelEncoder digunakan untuk membina beberapa fungsi analitik tambahan untuk menentukan pekerjaan dengan gaji purata tertinggi dan permintaan tertinggi dalam dataset. Dalam fasa penilaian, sistem diuji dengan resume contoh untuk menilai keberkesanan dan ketepatan model cadangan. Untuk menilai peningkatan prestasi jika model NLP lanjutan digunakan, perbandingan juga dibuat antara pendekatan TF-IDF sahaja dan pendekatan TF-IDF yang digabungkan dengan BERT.

Kemudian, hasil padanan disemak secara manual untuk memastikan ia sepadan dengan kandungan resume sebenar. Akhir sekali, fasa penerapan berlaku di mana model yang telah dibangunkan diintegrasikan ke dalam sistem penuh dengan menggunakan API Flask. Resume dihantar ke API Python melalui permintaan HTTP POST menggunakan `upload_resume.php`. API itu kemudiannya menangani resume dan memulangkan lima padanan kerja dalam bentuk JSON. Hasilnya dipaparkan kepada pengguna dalam halaman `results.php`. Reka bentuk sistem yang modular membolehkan fleksibiliti untuk penambahbaikan seterusnya, seperti penggunaan model yang lebih canggih, notifikasi pintar, dan ciri personalisasi.

Pecahan sistem modul adalah seperti berikut:

- Modul untuk pengurusan akaun pengguna dan keselamatan akses adalah `register.php` dan `login.php`.
- Antara muka yang digunakan oleh pengguna untuk memuat naik resume adalah `upload_resume.php`.
- `Results.php` akan mempamerkan hasil padanan kerja yang dihantar oleh API Python.
- `FYP.ipynb` mengandungi kod Python untuk penciptaan model cadangan, preprocessing data kerja, dan API Flask.

`Includes.php` dan `functions.php` akan menyediakan logik tambahan seperti pengurusan sesi dan validasi input.

Spesifikasi Keperluan Perisian

Semua perisian yang diperlukan disenaraikan dalam spesifikasi keperluan perisian ini adalah bagi membangunkan dan menjalankan aplikasi Sistem KerjaCerdas. Dengan menggunakan teknologi kecerdasan buatan seperti algoritma *Similarity* dan model *BERT*, aplikasi ini dapat membantu padanan antara majikan dan pencari kerja. Perisian yang diperlukan oleh pengguna akhir untuk menggunakan aplikasi dan perisian yang diperlukan untuk pembangunan sistem juga termasuk dalam keperluan ini.

Bahasa Pengaturcaraan

a. *Phyton*

Phyton adalah bahasa pengaturcaraan yang mudah difahami dan mudah dipelajari (Faradilla, 2024). Ianya merupakan bahasa pengaturcaraan yang terbaik yang banyak diguna pakai di dalam pelbagai aplikasi seperti analisis data, pembelajaran mesin dan pembangunan web. Selain itu, *Phyton* sangat disarankan bagi pemula kerana ianya mudah digunakan. Terdapat banyak *library* yang dapat digunakan di dalamnya. Antaranya adalah *NumPy*, *Pandas* dan *TensorFlow*. Bukan itu sahaja, ianya juga sesuai bagi pengaturcaraan backend, sains data, dan kecerdasan buatan (AI).

b. *PHP (Backend)*

Dart merupakan salah satu bahasa pengaturcaraan yang dibangunkan oleh *Google* (Nurosoft, 2024) dan digunakan bersama *Flutter* bagi membangunkan aplikasi seperti KerjaCerdas. Ianya juga berfungsi bagi membangunkan pelbagai platform yang lain seperti *Android*, *iOS*, dan web. Prestasi yang tinggi digabungkan dengan keupayaan compiled ke dalam kod asli adalah matlamat utamanya. Antara kelebihan *Dart* adalah ianya boleh diakses melalui telefon atau platform web dalam pembuatan aplikasi berasaskan *Flutter*. Ianya juga membenarkan pembangunan aplikasi yang menggunakan satu kod asas yang berfungsi dalam pelbagai platform.

c. *Flutter (Frontend)*

Flutter adalah rangka kerja antara muka pengguna sumber terbuka yang dibangunkan oleh *Google*, digunakan dalam pembangunan *frontend* sistem KerjaCerdas. Menggunakan satu kod asas, *Flutter* membolehkan aplikasi *cross-platform* seperti *Android*, *iOS*, dan web dibangunkan. *Flutter* digunakan dalam projek ini untuk

membina antara muka untuk fungsi seperti log masuk, daftar pengguna, muat naik resume dan paparan keputusan padanan kerja. Walaupun Flutter menggunakan bahasa Dart secara asas, ia berinteraksi dengan backend berasaskan PHP melalui HTTP request seperti POST dan GET. Dengan *Flutter*, data pengguna dihantar ke PHP dan diterima dalam bentuk JSON. Sokongan *Hot Reload*, antaramuka pengguna yang responsif, dan prestasi tinggi untuk rendering antaramuka menggunakan enjin Skia adalah antara kelebihan utama *Flutter*.

Perisian Pengaturcaraan

a. Google Colab

Selain *Dart*, *Google Colab* juga merupakan platform yang diwujudkan oleh *Google* bagi menjalankan notebook *Jupyter* melalui internet. Dengan sokongan GPU dan TPU bagi aplikasi pembelajaran mesin, ianya membolehkan penulisan dan pelaksanaan kod *Python*. *Google Colab* ini juga sesuai bagi pemula yang ingin mempelajari *Python* (Aorinka, 2023). *Google Colab* menyediakan sumber pengkomputeran seperti TPU dan GPU bagi latihan model pembelajaran mesin. Ianya juga sangat mudah digunakan dan boleh berkongsi projek dengan orang lain.

b. Microsoft Visual Studio Code

Microsoft Visual Studio Code atau juga dikenali sebagai VS Code merupakan editor kod sumber yang ringkas dan percuma yang dibangunkan oleh Microsoft. VS Code mempunyai plugin yang besar bagi pelbagai teknologi dan boleh menyokong pelbagai jenis bahasa pengaturcaraan. VS Code merupakan editor kod yang mempunyai pelbagai ciri yang lengkap berbanding aplikasi yang lain serta boleh beroperasi selain daripada Windows (Elly, 2024). Antara bahasa pengaturcaraan yang boleh digunakan di VS Code adalah *Python*, *JavaScript*, dan *C#*. Ianya juga boleh berfungsi seperti debugging, integrasi Git dan intellisense.

c. MySQL

Dalam sistem KerjaCerdas, *MySQL* digunakan sebagai RDBMS, atau sistem pengurusan pangkalan data relasi, untuk menyimpan dan mengurus maklumat pengguna serta sejarah padanan kerja. Pangkalan data *MySQL* secara teratur menyimpan semua data, termasuk maklumat akaun pengguna, resume yang telah

dimuat naik dan keputusan padanan kerja, yang boleh diakses semula pada bila-bila masa. Permintaan data yang cekap, penyimpanan berskala, dan kebolehpercayaan data boleh dicapai dengan MySQL. Ia disambungkan dengan PHP, yang berfungsi sebagai perantara antara pangkalan data dan antaramuka Flutter. Untuk memastikan integriti data sentiasa terpelihara, struktur pangkalan data menggunakan jadual berkaitan yang saling berkait.

Reka Bentuk Antaramuka

Reka bentuk antaramuka merujuk kepada proses mereka bentuk dan mencipta antara muka yang melaluinya pengguna berinteraksi dengan sistem, aplikasi atau peranti. Ia memainkan peranan penting dalam memastikan pengalaman pengguna yang hebat dan mudah. Reka bentuk antaramuka seharusnya berpaksikan dari pengguna dimana kesemua jenis pengguna diambil kira bagi mendapat sokongan dari pengguna (Zaidatun, 2014). Kerjasama antara pereka bentuk antaramuka (UI) dan pereka pengalaman pengguna (UX), adalah sangat penting di dalam reka bentuk antaramuka bagi mencipta sistem KerjaCerdas yang menarik, praktikal, dan cekap. Terdapat beberapa elemen yang diambil kira di dalam reka bentuk antaramuka KerjaCerdas, antaranya adalah visual, kebolegunaan, dan reka bentuk interaktif. Tujuan utama reka bentuk antaramuka ini adalah bagi meningkatkan pengalaman pengguna, meningkatkan keberkesanan serta aksesibiliti, dan bagi mendapatkan sistem yang mesra pengguna. Reka bentuk antaramuka KerjaCerdas berfokus kepada aplikasi mudah alih dan laman web.

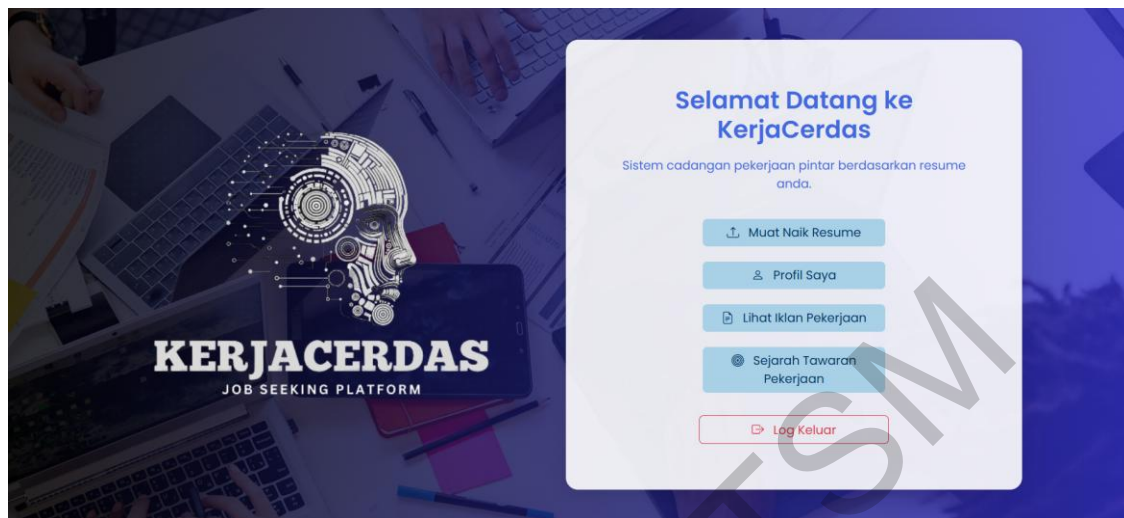
Halaman Pendaftaran Akaun dan Log Masuk

The image displays two screenshots of the KerjaCerdas Job Seeking Platform interface. The top screenshot shows the 'Daftar Akaun KerjaCerdas' (Register Account) form. It includes fields for 'Nama Penuh', 'E-mel', 'Kata Laluan', and 'Jenis Pengguna' (with 'Fresh Graduate' selected). A blue 'Daftar' button is at the bottom, along with a link 'Sudah ada akaun? Log Masuk'. The bottom screenshot shows the 'Log Masuk KerjaCerdas' (Login) form, featuring 'E-mel' and 'Kata Laluan' fields, a blue 'Log Masuk' button, and a link 'Belum ada akaun? Daftar di sini'. Both forms are overlaid on a background image of a desk with a laptop and a stylized head logo.

Rajah 8 Halaman Pendaftaran Akaun dan Log Masuk

Merujuk kepada rajah 8, rajah tersebut menunjukkan halaman bagi pendaftaran akaun dan log masuk. Bagi pengguna pertama kali, mereka dikehendaki untuk mendaftarkan akaun dengan butiran seperti nama, e-mel, nombor telefon dan kata laluan. Setelah membuat pendaftaran akaun, pengguna boleh log masuk ke dalam akaun yang telah didaftarkan. Jika pengguna hendak mengguna sistem pada waktu lain, pengguna boleh terus log masuk ke dalam akaun tanpa perlu mendaftar lagi.

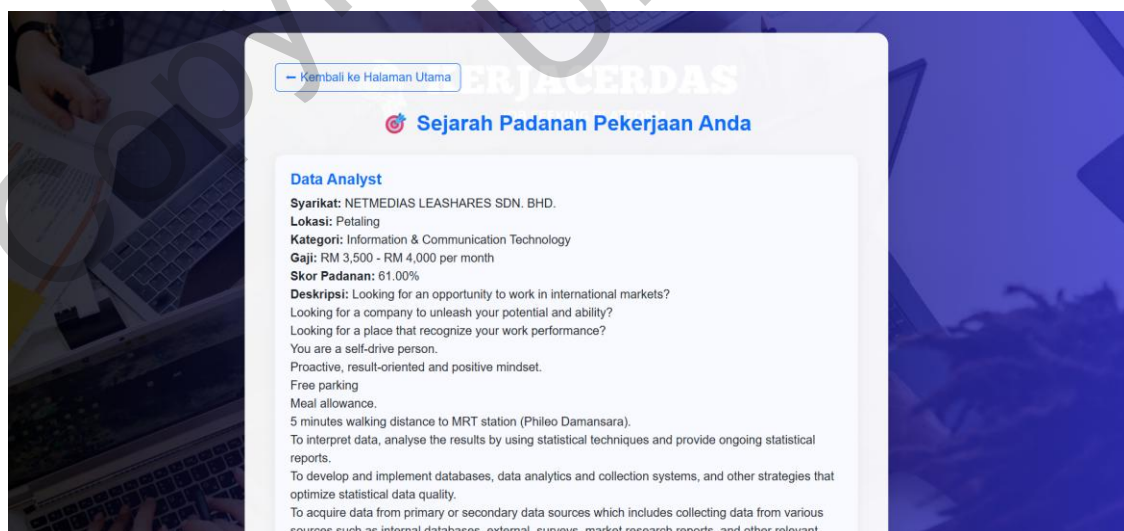
Halaman Utama KerjaCerdas



Rajah 9 Halaman Utama KerjaCerdas

Merujuk kepada rajah 9, berikut merupakan halaman utama bagi sistem KerjaCerdas dimana pengguna boleh membuat carian sendiri dan iklan-iklan pekerjaan akan yang dimasukkan oleh mana-mana syarikat akan dipaparkan secara rawak di halaman utama tanpa mengira bidang, lokasi dan kerja. Pada bahagian bawah halaman utama, terdapat empat halaman lain yang boleh diakses pengguna. Pengguna juga boleh membuat carian sendiri pada halaman utama.

Halaman Sejarah Padanan KerjaCerdas



Rajah 10 Halaman Sejarah Padanan KerjaCerdas

Merujuk kepada rajah 10, ianya menunjukkan halaman Sejarah padanan kerja bagi KerjaCerdas. Pengguna boleh menyimpan syarikat atau kerja yang mungkin diminati di bahagian halaman simpanan ini. Sebagai contoh, setiap kali pengguna memuat naik

resume, setiap keputusan dan padanan akan disimpan, jadi pengguna boleh melihat sejarah padanan hanya dengan log masuk ke dalam sistem mengikut akaun mereka.

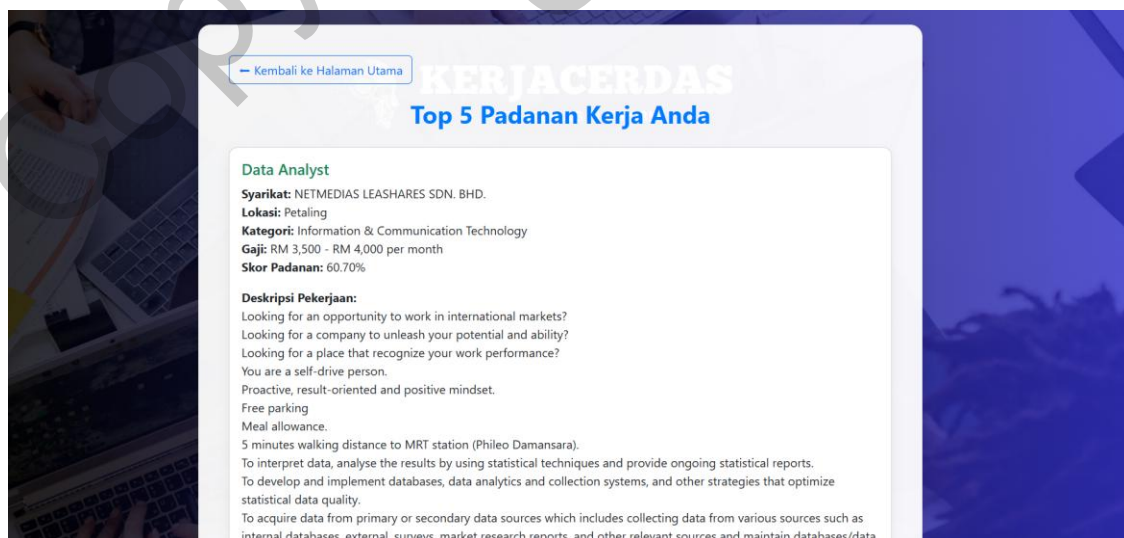
Halaman Muat Naik Resume KerjaCerdas



Rajah 11 Halaman Muat Naik Resume KerjaCerdas

Merujuk kepada rajah 11, halaman tersebut merupakan halaman muat naik resume, dimana pengguna boleh memuat naik resume berbentuk pdf atau perkataan. Jika resume yang dimuat naik menepati format sistem, pengguna boleh menghantar resume untuk dibuat padanan yang akan dihantar ke halaman muka keputusan KerjaCerdas.

Halaman Keputusan Padanan KerjaCerdas

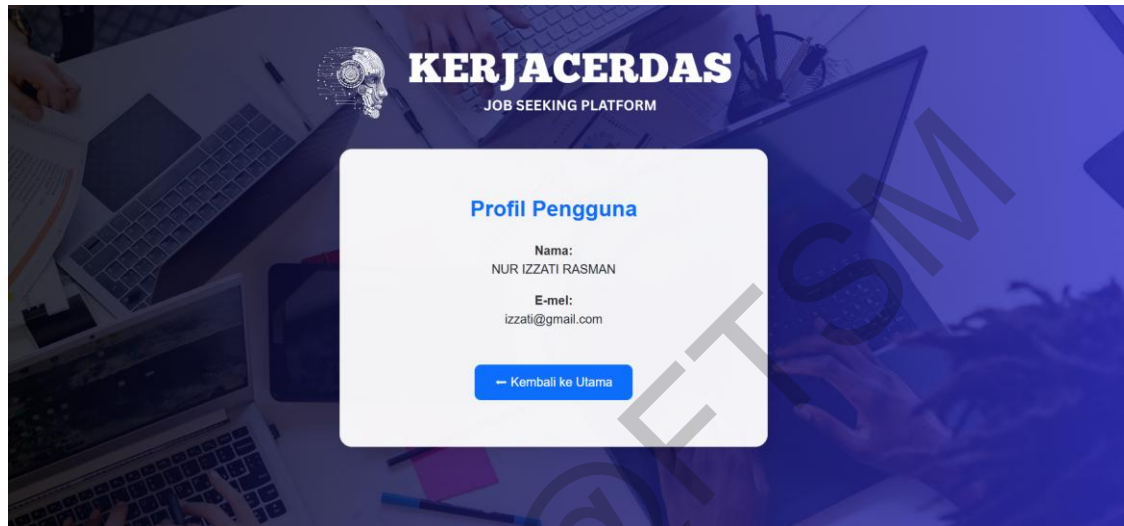


Rajah 12 Halaman Keputusan Padanan KerjaCerdas

Merujuk kepada rajah 12, berikut merupakan halaman keputusan bagi padanan kerja yang telah dibuat berdasarkan resume. Skor padanan terhasil daripada padanan antara

kata kunci di dalam resume dan data iklan pekerjaan di dalam dataset. Pengguna akan mendapat lima padanan kerja yang mempunyai nilai skor padanan yang tertinggi.

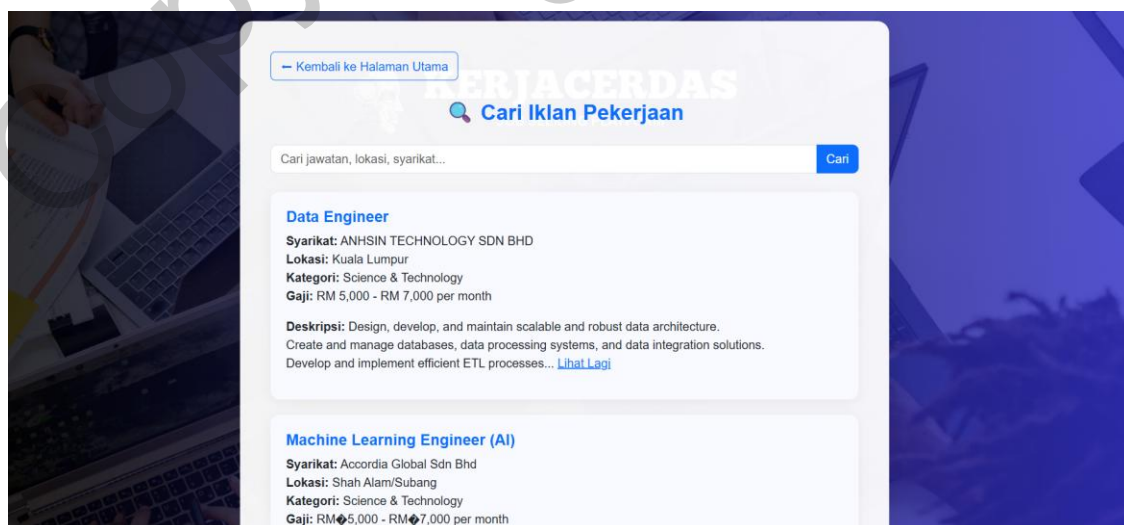
Halaman Profil Pengguna KerjaCerdas



Rajah 13 Halaman Profil Pengguna KerjaCerdas

Merujuk kepada rajah 13, berikut merupakan halaman profil pengguna bagi sistem KerjaCerdas. Bahagian profil pengguna akan memaparkan nama dan e-mel yang akan diambil dari pangkalan data selepas pengguna mendaftar. Pengguna hanya boleh melihat profil pengguna selepas log masuk akaun mereka kedalam sistem.

Halaman Iklan Pekerjaan KerjaCerdas



Rajah 14 Halaman Iklan Pekerjaan KerjaCerdas

Merujuk kepada rajah 14, berikut merupakan halaman iklan pekerjaan KerjaCerdas, dimana pengguna boleh melayari iklan pekerjaan mengikut kata kunci yang mengikut spesifikasi yang mereka mahukan seperti lokasi, gaji dan Syarikat.

KEPUTUSAN DAN PERBINCANGAN

Hasil ujian menunjukkan bahawa sistem KerjaCerdas berjaya memenuhi keperluan fungsi dan bukannya fungsi yang dirancang. Dengan skor cosine similarity 0.87 hingga 1.00 dalam ujian pepadanan resume menggunakan algoritma *BERT*, ia menunjukkan ketepatan yang tinggi dalam mencadangkan pekerjaan yang sesuai. Sebagai perbandingan, pepadanan menggunakan *TF-IDF* dan *cosine similarity* mendapat hanya 0.35 hingga 0.58, yang menunjukkan bahawa *BERT* adalah pilihan yang lebih baik untuk sistem ini.

Z	Julat Nilai Kesamaan / Skor	Ketepatan Semantik	Kelebihan	Kekurangan
TF-IDF + Cosine Similarity	0.34 – 0.58	Rendah	Mudah dilaksanakan dan cepat diproses	Tidak memahami konteks ayat, hanya bergantung pada kekerapan perkataan
TF-IDF + Jaccard Similarity	0.08 – 0.25	Sangat rendah	Sesuai untuk data binari dan analisis ringkas	Lemah untuk teks panjang dan tidak mengambil

				kira makna ayat
BERT + Cosine Similarity	0.87 – 1.00	Tinggi	Memahami makna mendalam dan konteks, sangat sesuai untuk NLP resume- job matching KerjaCerdas	Proses lebih lambat sedikit tetapi sangat tepat
BERT + Jaccard Similarity	0.66 – 0.78	Sederhana	Lebih baik daripada Jaccard biasa, masih menangkap sedikit makna	Binarisasi vektor BERT mengurangkan konteks asal
Content- Based Filtering (TF-IDF)	0.02 – 0.49 (majoriti < 0.3)	Rendah	Boleh digunakan untuk cadangan awal berdasarkan perkataan	Tidak sesuai untuk kes resume yang berbeza tetapi bermakna sama
Content- Based Filtering (BERT)	0.57 – 0.69	Sederhana – Tinggi	Berkesan dalam mengenal pasti padanan resume dan kerja dengan konteks serupa	Masih kurang sedikit daripada cosine similarity terus

Jadual 1 Keputusan Pemadanan KerjaCerdas

Kaedah *BERT* dan *Cosine Similarity* jelas menunjukkan prestasi terbaik dengan skor kesamaan tertinggi dan konsisten (0.87–1.00) daripada jadual. Ia sangat sesuai untuk sistem KerjaCerdas, yang menilai resume dan kerja berdasarkan makna, bukannya perkataan. Kaedah *TF-IDF* dan *Jaccard* dan *TF-IDF* dan *Cosine*, yang bergantung hanya pada perkataan permukaan, menunjukkan prestasi yang jauh lebih rendah. Oleh itu, *BERT* dan *Cosine* adalah teknik utama padanan yang dipilih oleh KerjaCerdas.

Kekuatan Sistem

Keberkesanan Sistem KerjaCerdas sebagai platform pemadanan kerja berasaskan teknologi pintar disokong oleh beberapa kekuatan utamanya. Penggunaan model pemrosesan bahasa semula jadi (NLP) yang canggih, *BERT* (Representations Bidirectional Encoder from Transformers), bersama kaedah cosine similarity adalah salah satu kelebihan yang paling ketara. Dengan gabungan ini, sistem tidak hanya dapat melihat persamaan kata kunci, tetapi juga dapat memahami maksud tersirat dalam resume dan hubungannya dengan deskripsi kerja. Ini menghasilkan padanan kerja yang lebih tepat, relevan dan kontekstual. Selain itu, antara muka pengguna yang mudah dan mudah digunakan sistem menjadikan proses memuat naik resume dan mendapatkan hasil padanan kerja lebih mudah. Selain itu, sistem menyokong format fail PDF yang biasa digunakan apabila memohon jawatan. Untuk membantu pengguna menilai kesesuaian kerja yang dicadangkan, skor padanan disediakan. Penekanan pada teknik penapisan berasaskan kandungan, juga dikenali sebagai penapisan berasaskan kandungan, menunjukkan bahawa sistem mengutamakan pengalaman pengguna secara eksklusif, tanpa mengambil kira data pengguna lain.

Kekangan Sistem

Walaupun ia mempunyai keupayaan teknikal yang tinggi, sistem KerjaCerdas masih mempunyai beberapa kelemahan yang memerlukan penambahbaikan untuk meningkatkan kualitinya. Keterbatasan dalam jumlah pangkalan data pekerjaan yang digunakan adalah antara halangan utama. Sistem sukar untuk memberikan cadangan yang benar-benar relevan jika data kerja tidak dikemaskini atau terdapat sedikit variasi. Selain itu, kerana sistem tidak menyokong penjejakan sejarah interaksi atau profil pengguna, ia tidak dapat

mengetahui keutamaan pengguna untuk padanan masa hadapan. Proses pemadanan sangat bergantung kepada kualiti resume yang dimuat naik; resume yang terlalu ringkas atau tidak lengkap boleh menyebabkan sistem memberikan cadangan yang tidak tepat. Tambahan pula, sistem tidak dapat memproses format fail lain seperti.docx atau gambar, tetapi ia hanya menyokong beberapa format fail, seperti PDF dan teks biasa. Akhir sekali, keperluan bukan teks seperti lokasi, tahap pengalaman, dan anggaran gaji masih tidak diambil kira oleh sistem dalam logik pemadanan. Ini adalah elemen penting dalam proses pencarian kerja sebenar.

Cadangan Penambahbaikan

Walaupun sistem KerjaCerdas telah mencapai objektif pembangunan utamanya, iaitu menawarkan pekerjaan yang paling sesuai berdasarkan resume pengguna, masih ada beberapa komponen yang boleh ditambah baik untuk meningkatkan pengalaman pengguna. Memperbaiki antara muka pengguna (UI/UX) adalah cadangan utama untuk menjadikannya lebih mudah difahami, moden, dan mesra pengguna untuk semua lapisan masyarakat, termasuk pengguna bukan teknikal. Untuk membantu pengguna memahami hasil cadangan sistem, ciri seperti pratonton resume, penanda aras padanan kerja (yang boleh dipaparkan dengan skor yang sepadan), dan panduan penggunaan boleh ditambah. Dari segi teknikal, sistem boleh dipertingkatkan dengan meningkatkan keupayaan analisis dokumen. Ini termasuk sokongan untuk pelbagai format fail, seperti.docx dan.odt, serta pengecaman automatik struktur resume, termasuk struktur pengalaman, pendidikan dan kemahiran. Tambahan pula, mencuba model NLP yang lebih berkuasa seperti RoBERTa, DistilBERT, atau SBERT bersaiz besar, yang telah ditunjukkan untuk berfungsi dengan lebih baik dalam tugas pemadanan semantik, boleh membantu meningkatkan pemilihan model pemadanan. Fungsi padanan berasaskan profil, juga dikenali sebagai nasihat berasaskan profil, boleh diperkenalkan. Fungsi ini memerlukan sistem melihat teks resume serta maklumat seperti lokasi pengguna, gaji yang diminta, dan tahap pengalaman pengguna.

Memperkenalkan mekanisme pembelajaran berasaskan maklum balas pengguna juga dikenali sebagai loop maklum balas pengguna adalah satu komponen yang sangat penting. Ini menunjukkan bahawa sistem akan mengumpul maklum balas tentang relevansi padanan yang diberikan dan menggunakan maklum balas ini untuk melatih semula atau mengubah suai model cadangan pada masa hadapan. Dengan itu, sistem

akan menjadi lebih cekap dan boleh menyesuaikan diri dengan perubahan keperluan pengguna. Akhir sekali, KerjaCerdas mungkin lebih praktikal dan berdaya saing dalam dunia sebenar jika ia digabungkan dengan platform pekerjaan sebenar seperti Jobstreet atau LinkedIn melalui API.

KESIMPULAN

Secara keseluruhan, sistem KerjaCerdas ialah platform pintar yang menggunakan teknologi pemprosesan bahasa semula jadi (NLP) untuk menyelesaikan isu pencarian kerja dengan lebih cekap dan pantas. Resume yang dihantar dalam bentuk fail diterima oleh sistem, yang kemudiannya memilih lima pekerjaan terbaik berdasarkan hubungan semantik antara teks resume dan deskripsi kerja yang disimpan dalam pangkalan data. Sistem telah menunjukkan ketepatan dan prestasi yang sangat memuaskan, dengan skor kesamaan yang konsisten tinggi dan relevan secara kontekstual, melalui penggunaan model BERT bersama cosine similarity. Hasil pengujian menunjukkan bahawa semua fungsi utama sistem berfungsi dengan baik dan tidak mempunyai ralat yang ketara. Sistem ini mendapat penerimaan yang baik dan dianggap berguna oleh pengguna, menurut ujian sistem, ujian kefungsian, ujian ketepatan dan ujian penerimaan pengguna (UAT) melalui soal selidik. Menurut analisis prestasi pemadanan, kaedah BERT + Cosine lebih baik daripada kaedah lain seperti TF-IDF dan Jaccard, menjadikannya pilihan terbaik untuk algoritma pemadanan dalam sistem ini. Dengan membina sistem KerjaCerdas, terbukti bahawa pendekatan berasaskan NLP boleh membantu pencari kerja membuat keputusan dengan lebih cepat dan tepat mengenai masalah pemadanan kerja. Sistem ini boleh menjadi platform sokongan kerjaya yang popular di Malaysia dan di seluruh dunia dengan penambahbaikan teknikal dan fungsi masa hadapan.

PENGHARGAAN

Segala puji dan syukur saya panjatkan ke hadrat Allah SWT kerana dengan limpah kurnia, rahmat, dan izin-Nya, saya telah berjaya menyiapkan Projek Tahun Akhir ini dengan sempurna. Saya ingin merakamkan setinggi-tinggi penghargaan dan ucapan terima kasih kepada penyelia projek saya, prof.madya ts. dr. nor samsiah binti sani, atas segala bimbingan, nasihat, dan dorongan yang telah diberikan sepanjang tempoh pelaksanaan projek ini. Sokongan beliau amatlah saya hargai dan menjadi tunjang kepada kelancaran projek ini. Ucapan terima kasih juga saya tujukan kepada semua pensyarah dan staf akademik di Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia yang telah banyak mencurahkan ilmu serta memberikan sokongan sepanjang pengajian saya.

Saya juga ingin merakamkan jutaan terima kasih kepada rakan-rakan seperjuangan saya atas kerjasama, sokongan moral, idea, dan semangat yang telah dikongsi sepanjang perjalanan projek ini. Persahabatan dan bantuan kalian amat bermakna bagi saya. Tidak dilupakan juga keluarga tercinta, ibu saya, Faridah binti Mohd Yatim, ayah saya, Encik Rasman bin Mat, serta adik-beradik saya yang sentiasa menjadi tulang belakang saya. Doa, kasih sayang, dan dorongan yang kalian berikan menjadi sumber kekuatan saya untuk terus berusaha.

Sekian, terima kasih kepada semua yang terlibat secara langsung atau tidak langsung dalam membantu. menjayakan projek ini. Semoga segala jasa dan kebaikan kalian dibalas dengan ganjaran yang setimpal oleh Allah SWT.

RUJUKAN

- Jose Luis, A. (2022). *Agile Development Process*.
<https://www.krasamo.com/agile-development-process/>
- Zakaria, N (2020). *Pengangguran di Malaysia : Kesan Pandemik Covid 19*
<https://jpp.mohr.gov.my/files/pdf/notes/14%20PENGANGGURAN%20DI%20MALAYSIA%20KESAN%20PANDEMIK%20COVID-19.pdf>
- Abd Hamid, A. (2023). *Demi Menampung Hidup, 1.3 Juta Pelajar IPT Kerja Sambilan*
<https://bebasnews.my/2023/04/07/demi-menampung-hidup-1-3-juta-pelajarippt-kerja-sambilan/>
- Muhd Salleh, W. (2021). *Aplikasi Teknologi dan Ruang Kerja Dalam Dunia Pekerja Pada Masa Kini*.
https://www.researchgate.net/publication/350487559_APLIKASI_TEKNOLOGI_DAN_RUANG_KERJA_DALAM_DUNIA_PEKERJAAN_DI_ERA_MASA_KINI
- Nor Hidayah, S. (2018). *Kepuasan Kerja dalam Kalangan Pekerja di Jabatan Keselamatan Universiti Sains Malaysia*.
<https://core.ac.uk/download/pdf/229279739.pdf>
- Amy, D. Micheal, R. (2024) *Data Analytics:What It Is, How It's Used, and 4 Basic Techniques*.
<https://www.investopedia.com/terms/d/data-analytics.asp>
- Erin Stertz, F. (2019) *The Advantages of an Employment Application*
<https://smallbusiness.chron.com/advantages-employment-application-17127.html>
- Mohd Zaky, Z. (2023) *Tiga isu utama beri kesan negative ekonomi negara*

<https://www.bharian.com.my/bisnes/lain-lain/2023/02/1066744/tiga-isu-utama-beri-kesan-negatif-ekonomi-negara>

Md Fauzi, A. (2024) *Kerajaan, majikan dan pekerja perlu komited seimbangkan kerja dan kehidupan.*

<https://news.uthm.edu.my/ms/2024/07/kerajaan-majikan-dan-pekerja-perlu-komited-seimbangkan-kerja-dan-kehidupan/>

Kashyap, K. (2020) *All about Embeddings*

<https://medium.com/@kashyapkathrani/all-about-embeddings-829c8ff0bf5b>

Liubov, Z. (2024) *Choosing the Best Approach to Building an NLP-Based Text Recommendation System.*

<https://mobidev.biz/blog/how-to-build-text-based-recommender-system-with-nlp>

Bernama, (2021) *Penggunaan teknologi digital meningkat Ketika pandemic COVID-19.*

<https://www.astroawani.com/berita-malaysia/penggunaan-teknologi-digitalmeningkat-ketika-pandemik-covid-19---lppkn-298134>

Cole, S, Jim, H. (2024) *Apa itu NLP (natural language processing atau pemprosesan Bahasa alami) ?*

<https://www.ibm.com/id-id/topics/natural-language-processing>

Satrio, T. (2020) *Natural Language Processing (NLP)*

<https://shiftacademy.id/blog/natural-language-processing/>

Vince, L. (2024) *Traditional Natural Language Processing (NLP) Tasks:*

Understanding What Large Language Models (LLMs) Can Do For You

<https://medium.com/thedeephub/traditional-natural-language-processing-nlp-tasks-understanding-what-large-language-models-e8899161f202>

Liyana, L. (2024) *Tips Penyelidikan & Contoh Metodologi Kajian*

<https://ecentral.my/contoh-metodologi-kajian/#:~:text=Maksud%20Metodologi%20Kajian&text=Dalam%20erti%20kata%20lain%2C%20metodologi,yang%20boleh%20menyokong%20sesuatu%20kajian>

Suhaimi, S. (2009) *Proses Pembangunan Laman Web (WDLC)*

<https://blogapaja.wordpress.com/2009/01/25/proses-pembangunan-laman-web-%E2%80%93-langkah-kitaran-hayat-wdlc/>

Faradilla, A. (2024) *Apa Itu Phyton: Mengenal Salah Satu Bahasa Pemrograman Terpopuler*

<https://www.hostinger.co.id/tutorial/python-adalah>

Aorinka, A. (2024) *Apa itu Flutter? Ini Pengertian, Fungsi, dan Cara Kerjanya!*

<https://www.dewaweb.com/blog/apa-itu-flutter/>

Aorinka, A. (2023) *Mengenal Apa Itu Google Colab dan Cara Menggunakannya*

<https://www.dewaweb.com/blog/mengenal-google-colab/>

Elly, S. (2024) *VSCODE Adalah – Pengertian, Fitur, Kelebihan, dan Cara Menggunakannya*

<https://idwebhost.com/blog/vscode-adalah/>

Jade, M. (2023) *Apakah itu UML Use Case Diagram: Simbol, Template, Alat dan Tutorial*

<https://www.mindonmap.com/ms/blog/what-is-a-uml-use-case-diagram/>

Jabatan Digital Negara. (2023) *4.10 Penyediaan Spesifikasi Reka bentuk Sistem*

<https://sqa.jdn.gov.my/index.php/ms/4-10-penyediaan-spesifikasi-reka-bentuk-sistem-f3-6>

John, L. (2021) *Seni Bina Komputer*

https://ms.wikipedia.org/wiki/Seni_bina_komputer

Zaidatun, T. (2014) *Asas Reka Bentuk Laman Web (Antaramuka)*

<https://utmcdex.uttm.my/wp-content/uploads/2014/01/03-InterfaceDesign.pdf>