Job Automation Risk Prediction

Final Presentation

Secureer

Group-7

Kaung Nyo Lwin- st125066

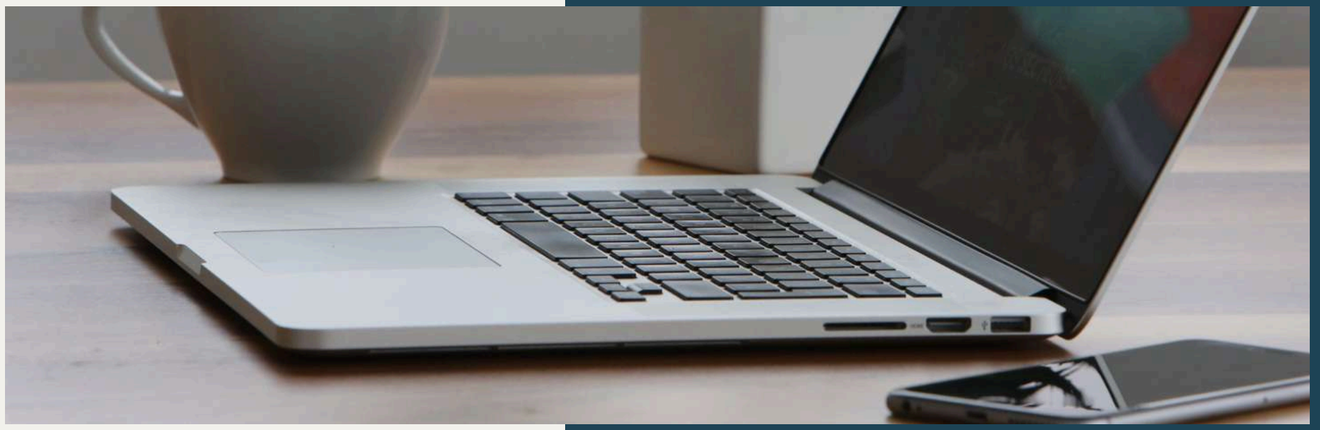
Nyein Chan Aung- st125553

Phone Myint Naing- st124973

Khin Yadanar Hlaing-st124959

December 2024

TABLE OF CONTENT

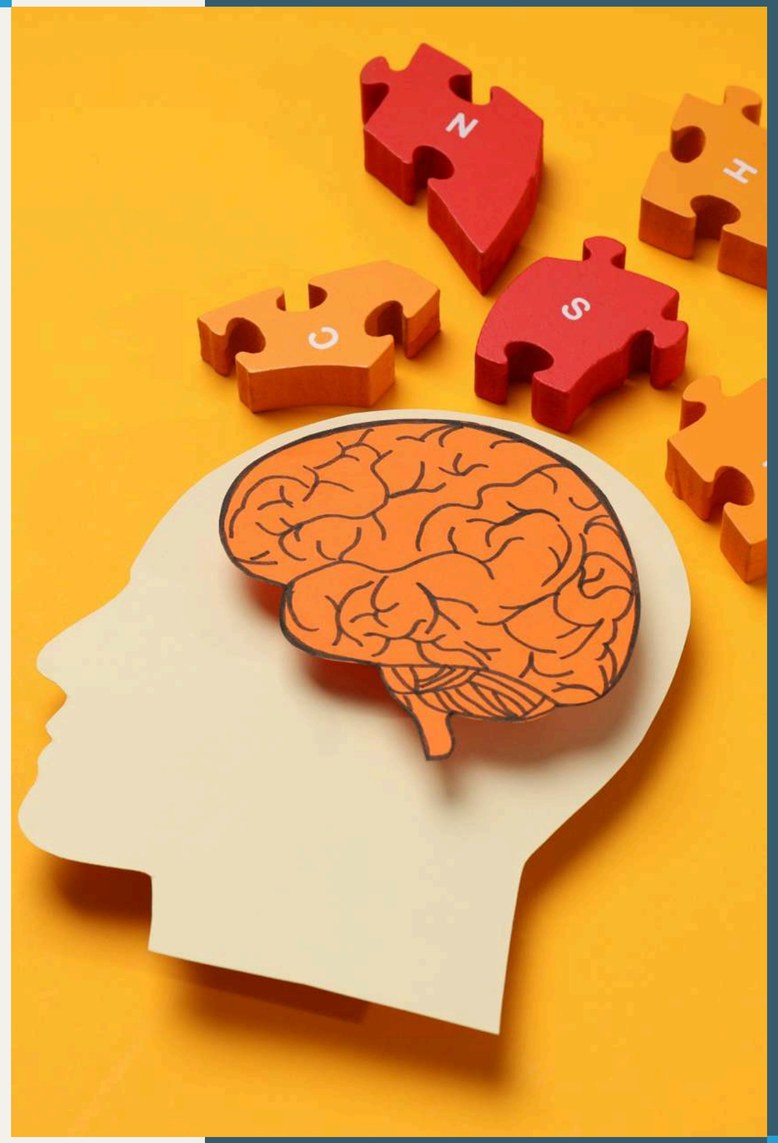


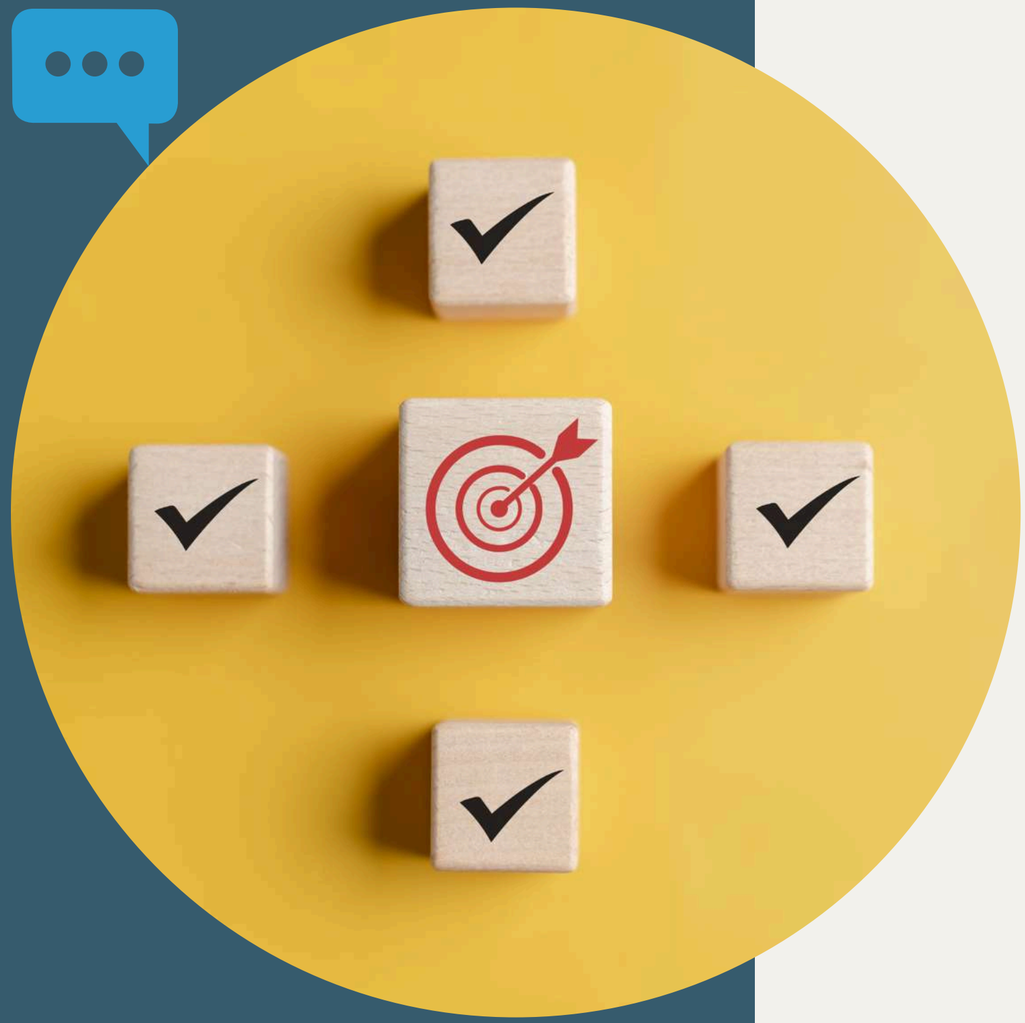
INTRODUCTION

Advancements in AI and automation are reshaping global labor markets, placing jobs at risk and creating new opportunities. Myanmar’s digital transformation presents challenges and opportunities, with industries in early stages of adopting cutting-edge technologies. Leveraging Frey and Osborne’s framework, this project assesses automation risks using Myanmar’s job market data, assuming the methodology remains relevant to the current context.

PROBLEM STATEMENT

- Advancements in digitalization threaten Myanmar's workforce.
- Repetitive and low-skill jobs are highly vulnerable to automation.
- Myanmar lacks a framework to:
  - Assess computerization risk for occupations.
  - Provide career transition guidance.





PROJECT OBJECTIVES

- Develop a Computerization Risk Assessment Model using MyJob.com data.
- Apply Frey and Osborne’s framework to classify job roles by automation risk.
- Give a Job Recommender System to:
  - Suggest alternative low-risk careers.
  - Guide career planning and skill development initiatives.

METHODOLOGY

Secureer

Data Collection

Datasets sponsored by MyJob.com
-Job Listings Dataset-7116 rows
-Job Skills Dataset- 3809 rows

Data Preparation

Clean, normalize, and Feature Engineering

Modeling

TextVectorizer + NN,
Sentence Transformer + NN

DATA PREPARATION

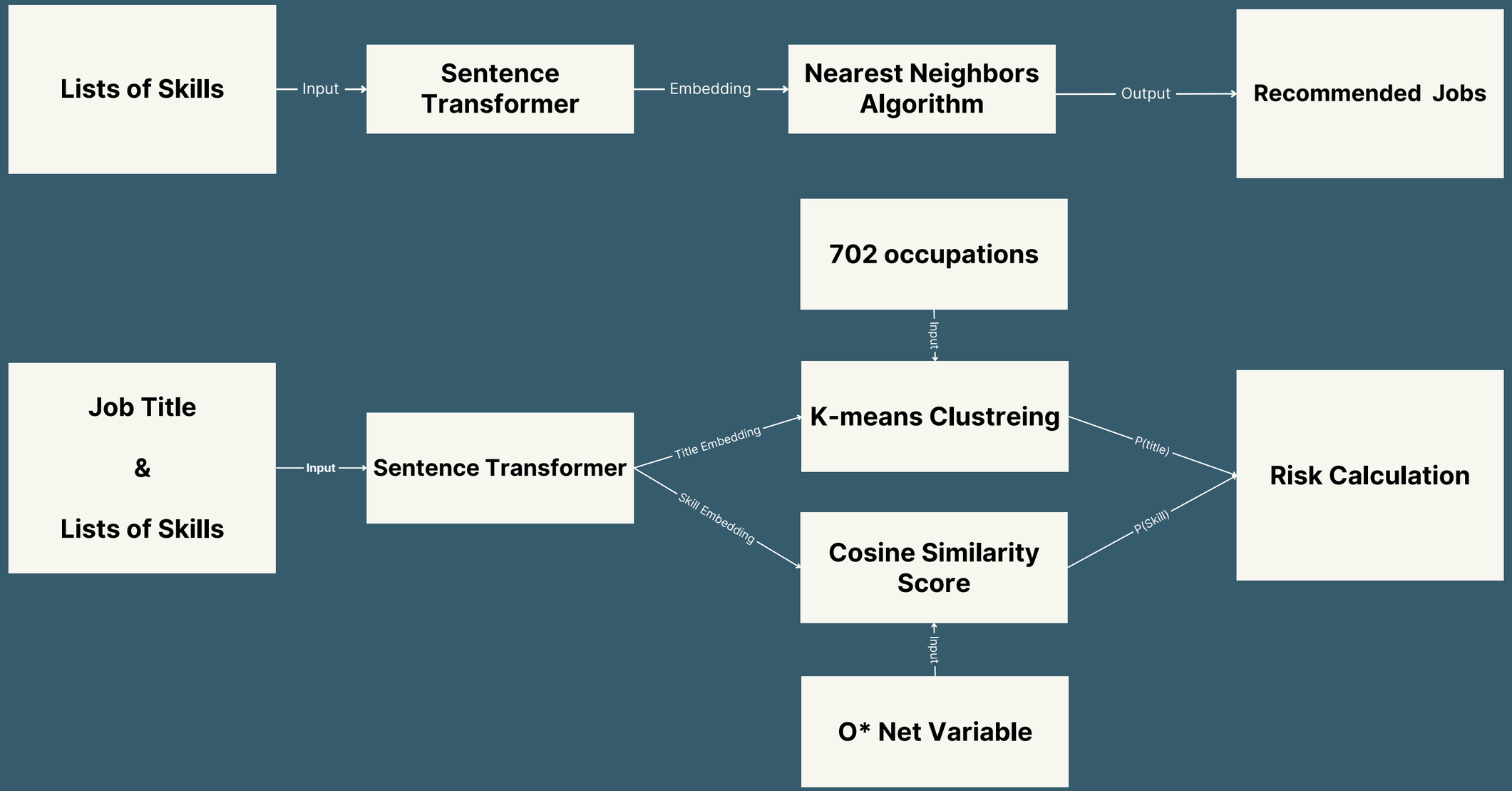
- Feature Selection: Chose essential features: ID, Title, and Industry from the job listings dataset.
- Merging Datasets: Combined with the Job Skills dataset on ID. Selected Title, Industries, and Job Skills for further analysis.
- Data Cleaning: Removed Myanmar text from columns using regex:

```
df['Job Skills'] = df['Job Skills'].apply(lambda x: ' '.join(re.findall(r'[a-zA-Z]+', x)))
df
```

- Grouping and Aggregation: Grouped by Title and aggregated

```
df = df.groupby('Title').agg({
    'Industries': list,
    'Job Skills': list}).reset_index()
```

MODELING DIAGRAM



O* NET VARIABLE

Computerisation bottleneck	O*NET Variable	O*NET Description
Perception and Manipulation	Finger Dexterity	The ability to make precisely coordinated movements of the fingers of one or both hands to grasp, manipulate, or assemble very small objects.
	Manual Dexterity	The ability to quickly move your hand, your hand together with your arm, or your two hands to grasp, manipulate, or assemble objects.
	Cramped Work Space, Awkward Positions	How often does this job require working in cramped work spaces that requires getting into awkward positions?
Creative Intelligence	Originality	The ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem.
	Fine Arts	Knowledge of theory and techniques required to compose, produce, and perform works of music, dance, visual arts, drama, and sculpture.
Social Intelligence	Social Perceptiveness	Being aware of others' reactions and understanding why they react as they do.
	Negotiation	Bringing others together and trying to reconcile differences.
	Persuasion	Persuading others to change their minds or behavior.
	Assisting and Caring for Others	Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients.

TABLE OD702 OCCUPATIONS

The table is ranks occupations according to their probability of computerisation (from least- to most-computerisable)

Rank	Computerisable Probability	Label	SOC code	Occupation
1.	0.0028		29-1125	Recreational Therapists
2.	0.003		49-1011	First-Line Supervisors of Mechanics, Installers, and Repairers
3.	0.003		11-9161	Emergency Management Directors
4.	0.0031		21-1023	Mental Health and Substance Abuse Social Workers
5.	0.0033		29-1181	Audiologists
6.	0.0035		29-1122	Occupational Therapists
7.	0.0035		29-2091	Orthotists and Prosthetists
8.	0.0035		21-1022	Healthcare Social Workers
9.	0.0036		29-1022	Oral and Maxillofacial Surgeons
10.	0.0036		33-1021	First-Line Supervisors of Fire Fighting and Prevention Workers
11.	0.0039		29-1031	Dietitians and Nutritionists
12.	0.0039		11-9081	Lodging Managers
13.	0.004		27-2032	Choreographers
14.	0.0041		41-9031	Sales Engineers
15.	0.0042	0	29-1060	Physicians and Surgeons
16.	0.0042		25-9031	Instructional Coordinators
17.	0.0043		19-3039	Psychologists, All Other
18.	0.0044		33-1012	First-Line Supervisors of Police and Detectives
19.	0.0044	0	29-1021	Dentists, General
20.	0.0044		25-2021	Elementary School Teachers, Except Special Education
21.	0.0045		19-1042	Medical Scientists, Except Epidemiologists
22.	0.0046		11-9032	Education Administrators, Elementary and Secondary School
23.	0.0046		29-1081	Podiatrists

RISK CALCULATION

$$P_{skill} = \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} (1 - S_{ij})$$

$$P_{title} = \frac{1}{m} \sum_{i=1}^{m} P_i$$

$$Risk = \frac{P_{title} + P_{skill}}{2}$$

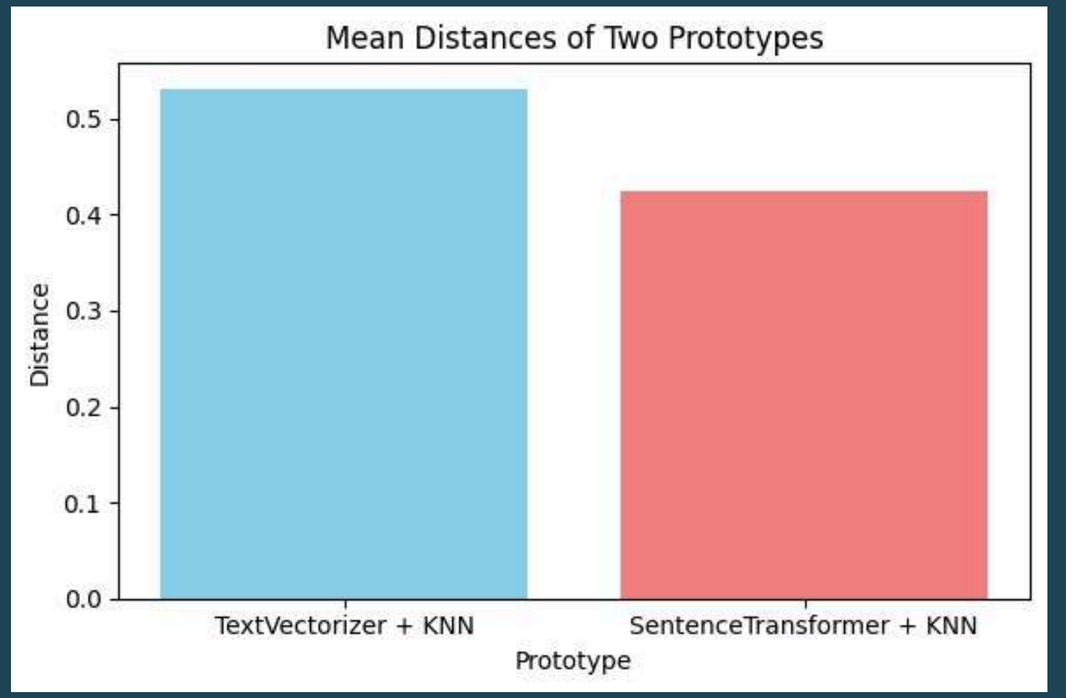
- P is the probability of computerization
- m is the number of samples belonged to the predicted group cluster
- n is the skills
- k is the O*Net variables
- S is the similarity score between skills and O*Net variables

TEXT ENCODING

```python
from sklearn.feature_extraction.text import TfidfVectorizer
tf = TfidfVectorizer()
skill_vectors = tf.fit_transform(df['Job Skills'])
skill_vectors
from sklearn.neighbors import NearestNeighbors

knn = NearestNeighbors(n_neighbors=5, metric='cosine')
knn.fit(skill_vectors)
```

```python
from sentence_transformers import SentenceTransformer

model = SentenceTransformer('sentence-transformers/all-MiniLM-L6-v2')
skill_embeddings = model.encode(df['Job Skills'].to_list())
skill_embeddings
from sklearn.neighbors import NearestNeighbors

knn = NearestNeighbors(n_neighbors=5, metric='cosine')
knn.fit(skill_embeddings)
```

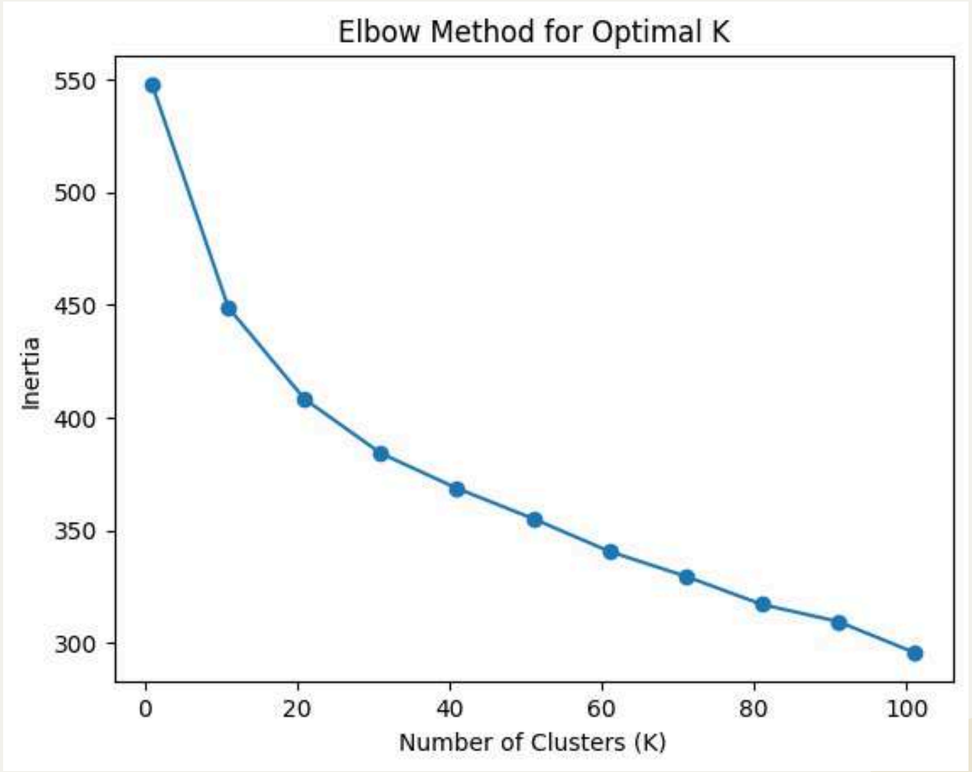
```python
# Data for plotting
types = ['TextVectorizer + KNN', 'SentenceTransformer + KNN']
values = [text_vectorizer_score, transformer_score]

# Plotting the bar chart
plt.figure(figsize=(6, 4))
plt.bar(types, values, color=['skyblue', 'lightcoral'])

# Adding titles and labels
plt.title('Mean Distances of Two Prototypes')
plt.xlabel('Prototype')
plt.ylabel('Distance')

# Show the plot
plt.tight_layout()
plt.show()
```



CLUSTERING

```
inertia = []
k_values = range(1, 102, 10)

for k in k_values:
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(job_embeddings)
    inertia.append(kmeans.inertia_)

# Plot the Elbow Curve
plt.plot(k_values, inertia, marker='o')
plt.xlabel("Number of Clusters (K)")
plt.ylabel("Inertia")
plt.title("Elbow Method for Optimal K")
plt.show()
```

Finding the optimal K for 702 occupation list



EVALUATION

- Dataset: BBC Business Insider dataset, enhanced with job titles and relevant skills to align with the model's approach.

Metrics:

- Mean Squared Error (MSE): Assesses prediction accuracy by measuring average squared differences.
- Mean Absolute Percentage Error (MAPE): Provides a relative measure of prediction performance.

EVALUATION RESULT

```
mean_squared_error(y_true, y_pred)
```

```
9.147299250785093
```

```
mean_absolute_percentage_error(y_true, y_pred)
```

```
3.631028576493364
```

- 365 test samples
- MSE - 9.1
- MAPE - 3.6
- average three percent error from true value of test

APP DEVELOPMENT

Tech Stack:

- The application was built using Django, leveraging its robust framework for backend management and frontend integration.
- Risk prediction machine learning model was embedded within the application to calculate automation risks based on job data.

CI/CD Pipeline:

- A CI/CD pipeline was created using GitHub Actions to streamline the development and deployment processes. Continuous Integration: Merges code changes into the main branch without conflicts.

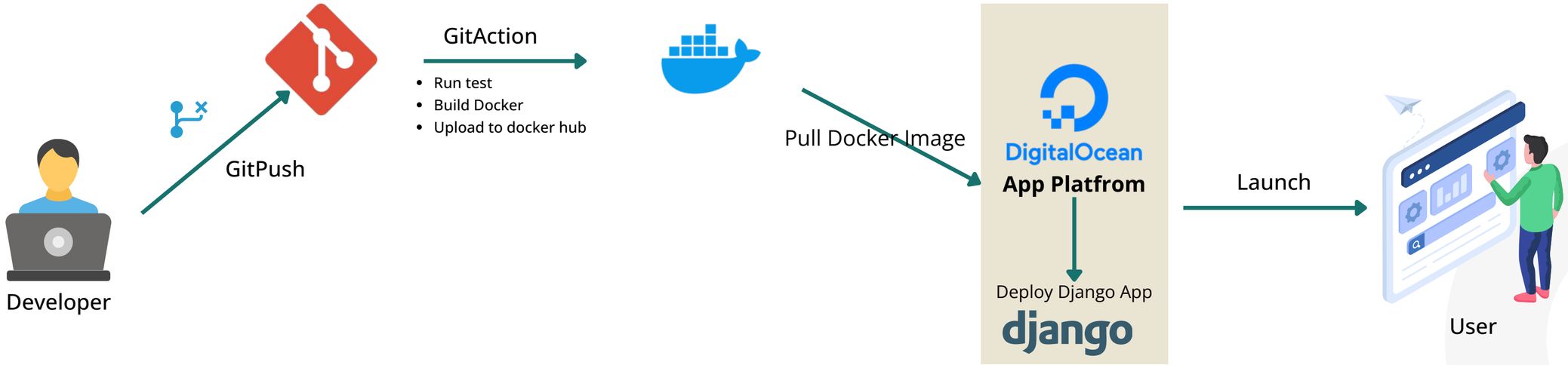
Image Source: https://www.linkedin.com/pulse/mastering-django-essential-design-patterns-best-mohammad-fa-alfard-zsg3f

DEPLOYMENT

Deployment Platform:

The application was deployed on DigitalOcean's App Platform, leveraging its serverless capabilities for:

- Scalability: Automatically adjusts resources based on demand.
- Ease of Use: Simplifies deployment by managing infrastructure.
- Cost-Effectiveness: Optimizes operational costs by scaling resources dynamically.



DEMO

Securing Your Career Path

with Job Automation Risk Prediction

Name

Nyein Chan Aung

Position

Product Manager

Skills

Analyst × Management × Team Building × Programming × Enter your skills

I'm looking for a job

I want to improve my skill

By submitting the form you agree to our privacy policy.

Submit

Securing Your Career Path

with Job Automation Risk Prediction

Estimated risk of automation:

58.50 %

Moderate Risk

Recommended Job Title

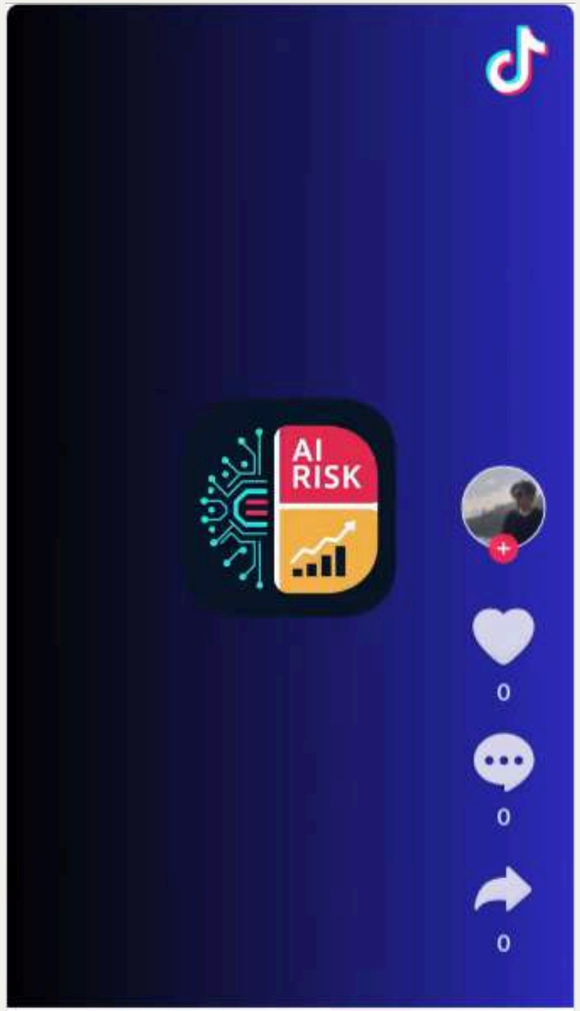
According to you skill, following position are suggested to apply.

	Job Title	Estimated Risk
1.	**Teaching Assistant Young and Adult Learners**	45.5 %
2.	**Regional Sale Manager Male Yangon Posts Tractor Truck Engine oil Generator**	79.5 %
3.	**Regional Sale Manager Male Posts Mandalay**	61.5 %
4.	**Assistant Sale Marketing Manager Position**	59.0 %
5.	**Registered Nurse**	60.0 %

Matched skills

DEMO

https://shark-app-ylemj.ondigitalocean.app



https://www.tiktok.com/@t.bwar/video/7443827162106367239

https://www.tiktok.com/@khinkhin1996/video/7443612845931023632

THANK YOU

FOR YOUR NICE ATTENTION

December 2024