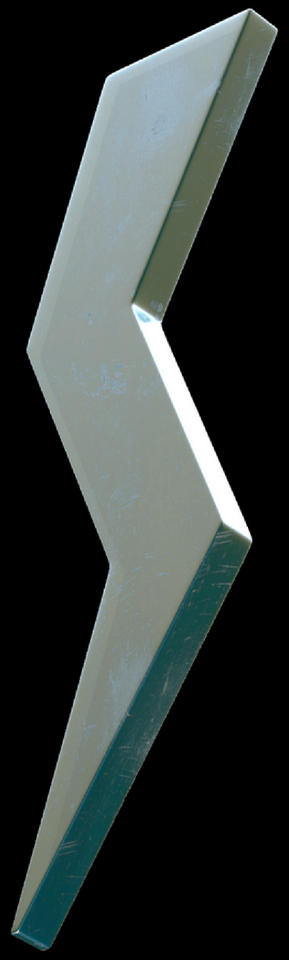
Artificial intelligence

to classify LinkedIn experiences

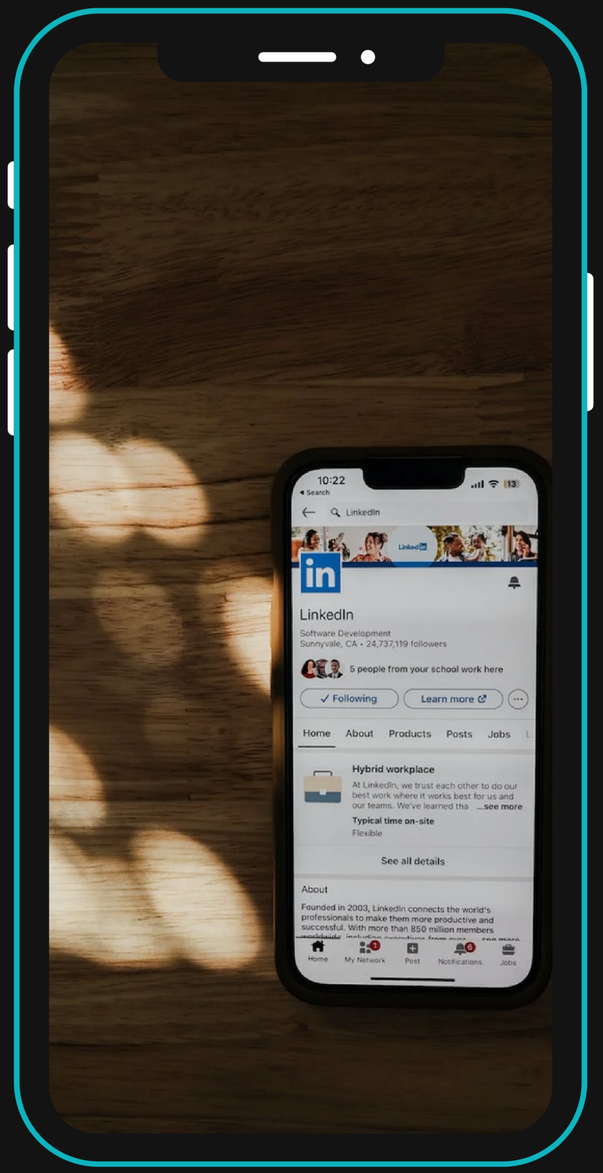
Little Beauty Bisoyi
1906335
Information Technology
Semester exchange student
CY Tech, Cergy, France
Mentor: Thomas DANGUILHEN

Content

What you need to know



1. Introduction
2. Problem Statement
3. Libraries used
4. Dataset
5. Data Cleaning
6. Exploratory data analysis
7. Feature engineering
8. Splitting the dataset
9. ML Models
10. Plotting F1 Score
11. Checking for overfitting
12. Prediction
13. Conclusion
14. Limitations
15. Future Scope



Introduction

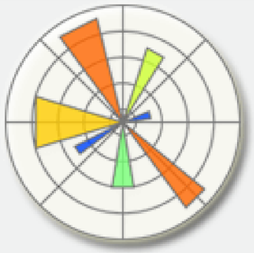
LinkedIn has emerged as a prominent platform for professionals to showcase their skills, experiences, and career aspirations.Manual classification and analysis of LinkedIn experiences can be a time-consuming task. The whole process can be automated by using Artificial Intelligence.

Problem Statement

We have a large number of LinkedIn profiles, and each profile has on average 5 previous experiences. We want to find a method to automatically classify the previous experiences as AI or non-AI.

AI/Non-AI

Libraries used



Matplotlib

Dataset

jobId
Unique key

548398365_0

companyName
No null values

TEHTRIS

companyUrl
10.011111111111111 % null

https://www.linkedin.com/company/tehtris/

jobTitle
0.011111111111111112 % null

R&D Cybersecurity Consultant

dataRange
10.011111111111111 % null

oct. 2021 – Aujourd'hui

location
10.011111111111111 % null

Bordeaux, Nouvelle-Aquitaine, France

description
28.3 % null

Conception et programmation pour une plateform...

logoUrl
11.988888888888889 % null

https://media-exp1.licdn.com/dms/image/C4D0BAQ...

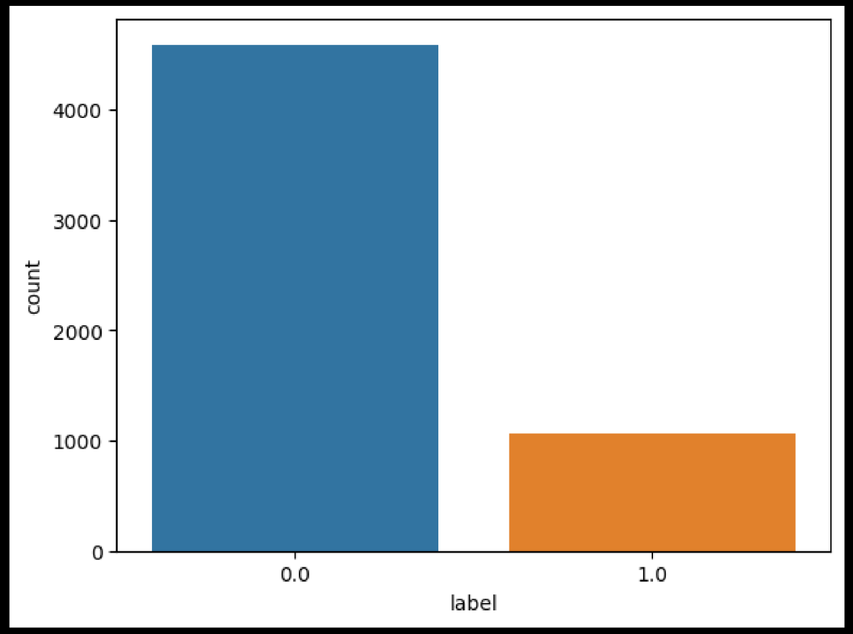
label
37.21111111111111 % null

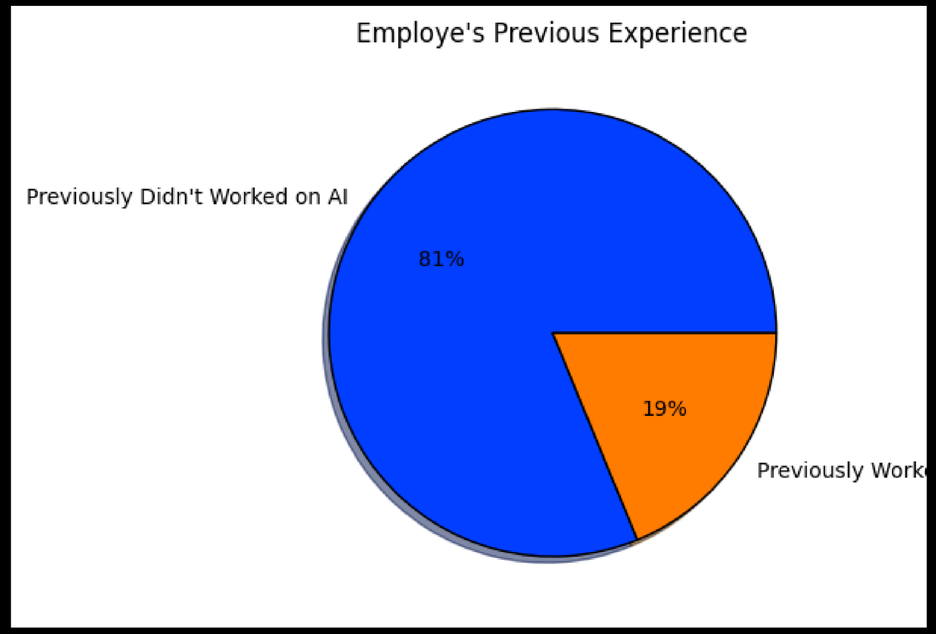
1

DATA CLEANING

1. Removing Rows with label NAN
2. Replace None in Desc col with empty string
3. Concatenating columns jobTitle and desc
4. Converting jobDesc col into lowercase
5. Removing stopwords
6. Removing unecessary cols
7. Relabelling the label col

Exploratory Data Analysis





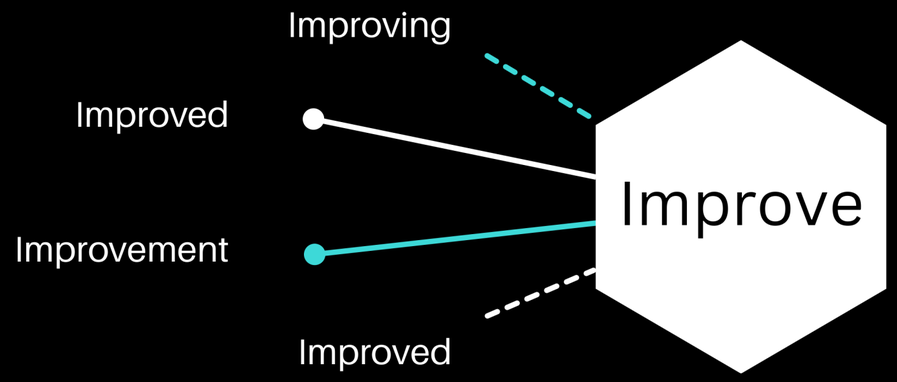
Visualize label column

AI / Non-AI

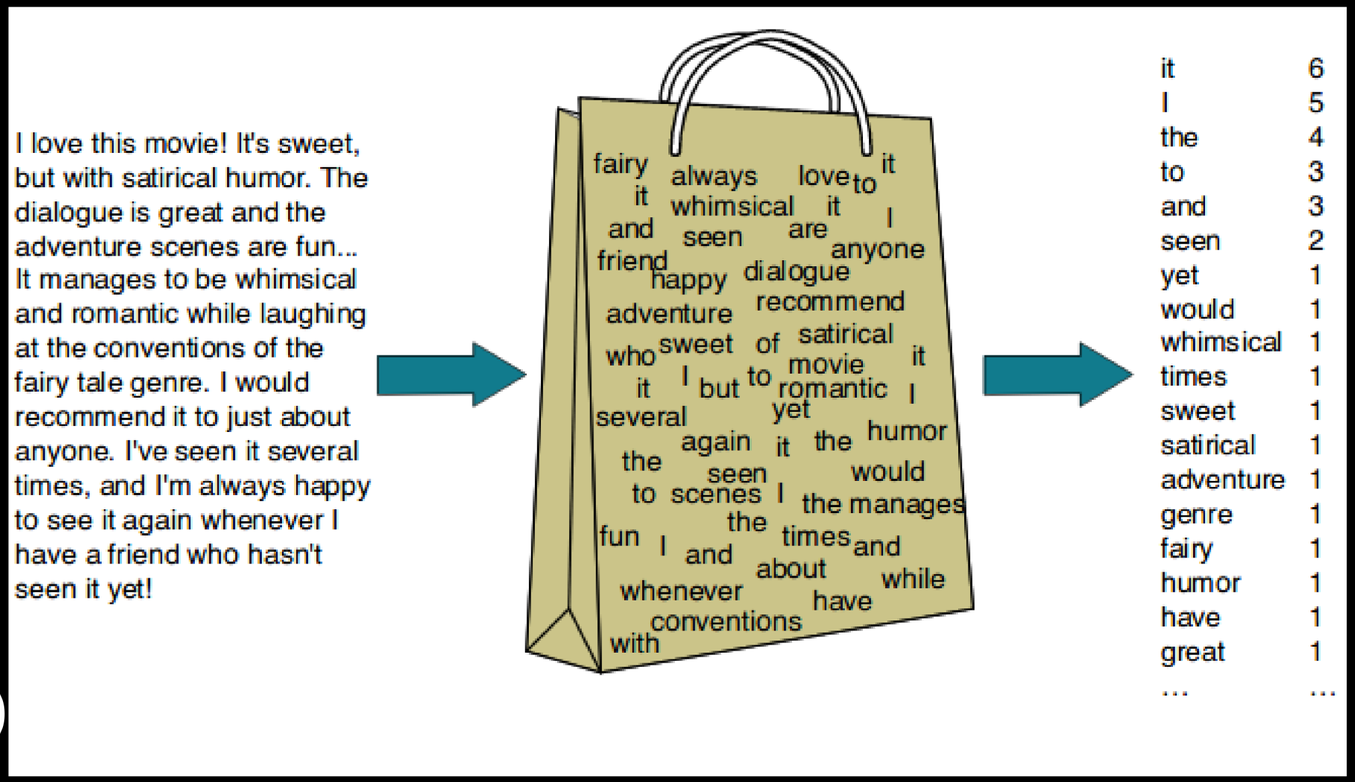
FEATURE ENGINEERING

01 Lemmatization

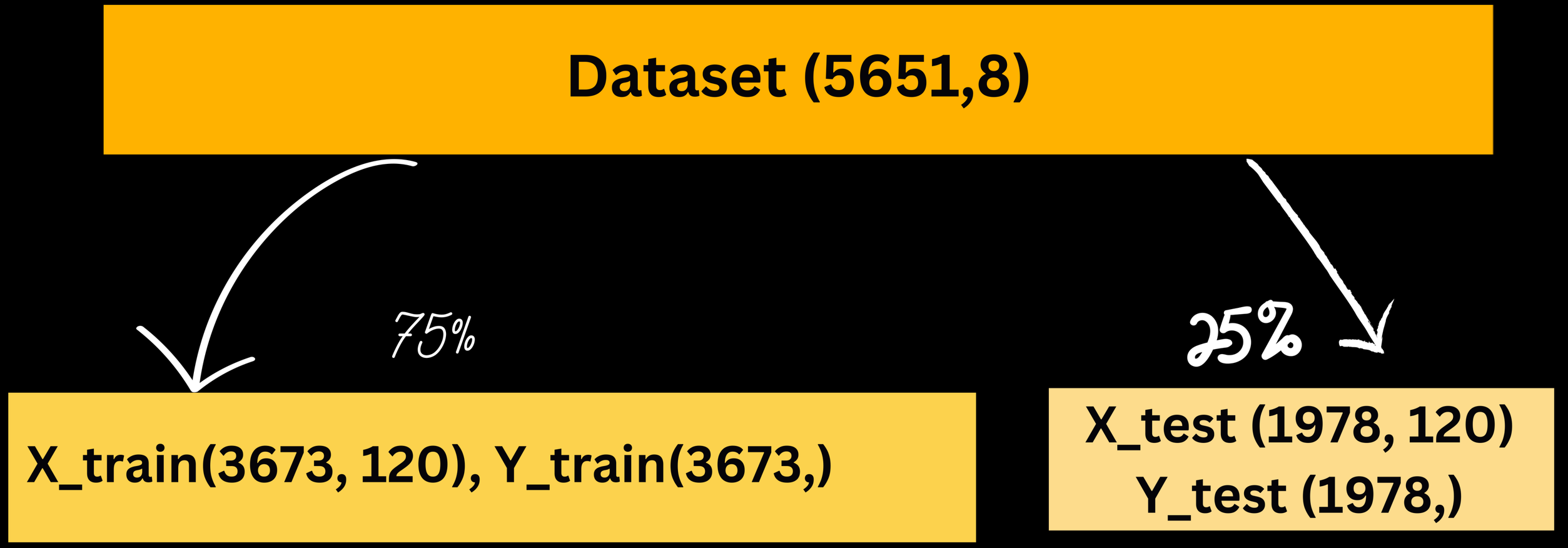
(jobDescription_without_stopwords)



02 Bag of Words



Splitting the dataset



Modeling

F1 Score

$$\frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})}$$

Accuracy

$$\frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Precision

$$\frac{TP}{(TP + FP)}$$

Classification error

1 - Accuracy

MSE

((Y_pred_class - Y_test) ** 2).mean()

Logistic Regression

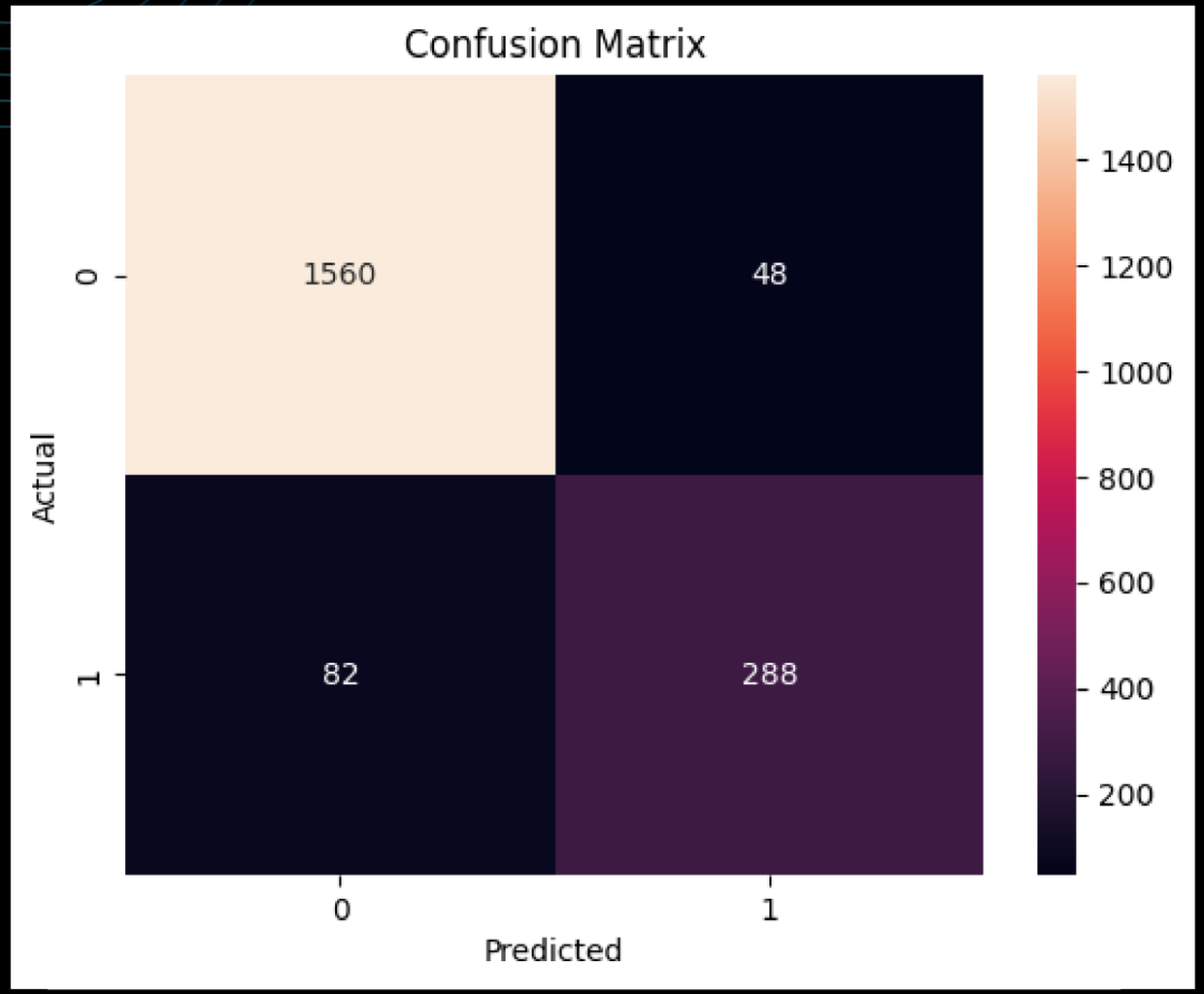
Accuracy: 0.93418

Classification Error: 0.065

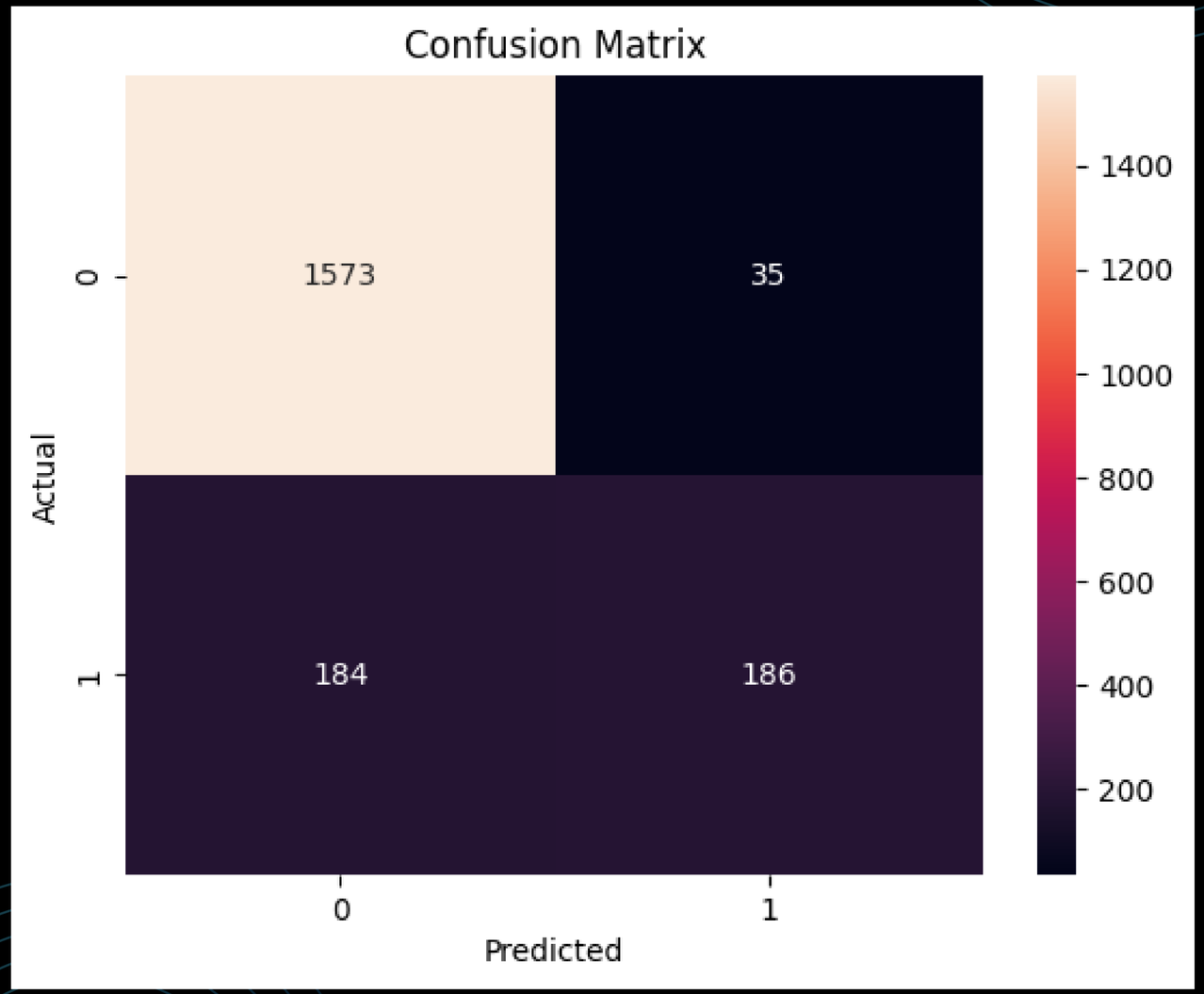
Precision: 0.84

F1 Score: 0.815

MSE : 0.065



k- Nearest Neighbour



Accuracy: 0.88393

Classification Error: 0.1160

Precision: 0.811

F1 Score: 0.629

MSE : 0.11

Decision Tree Classifier

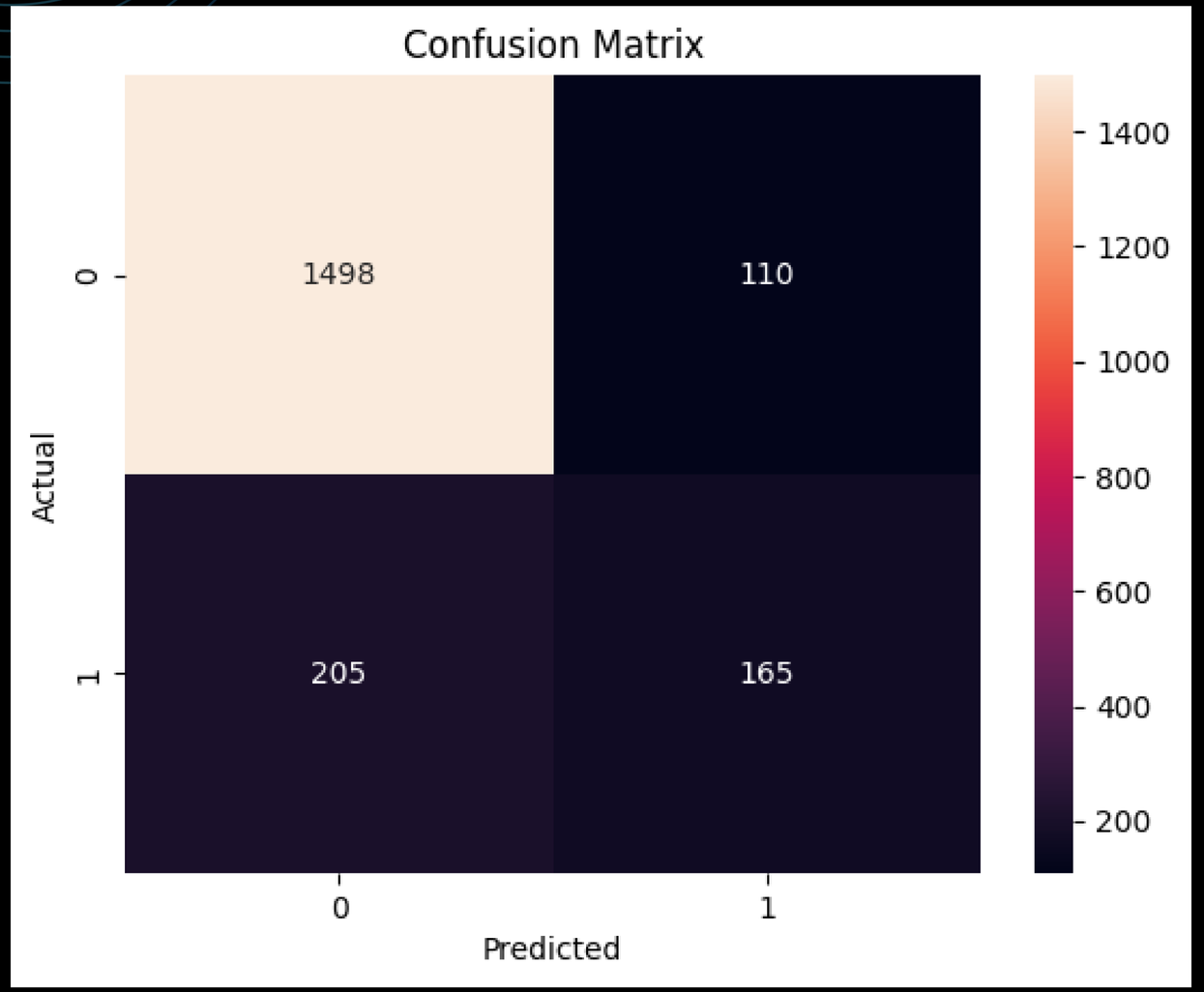
Accuracy:
0.84074

Classification Error:
0.159

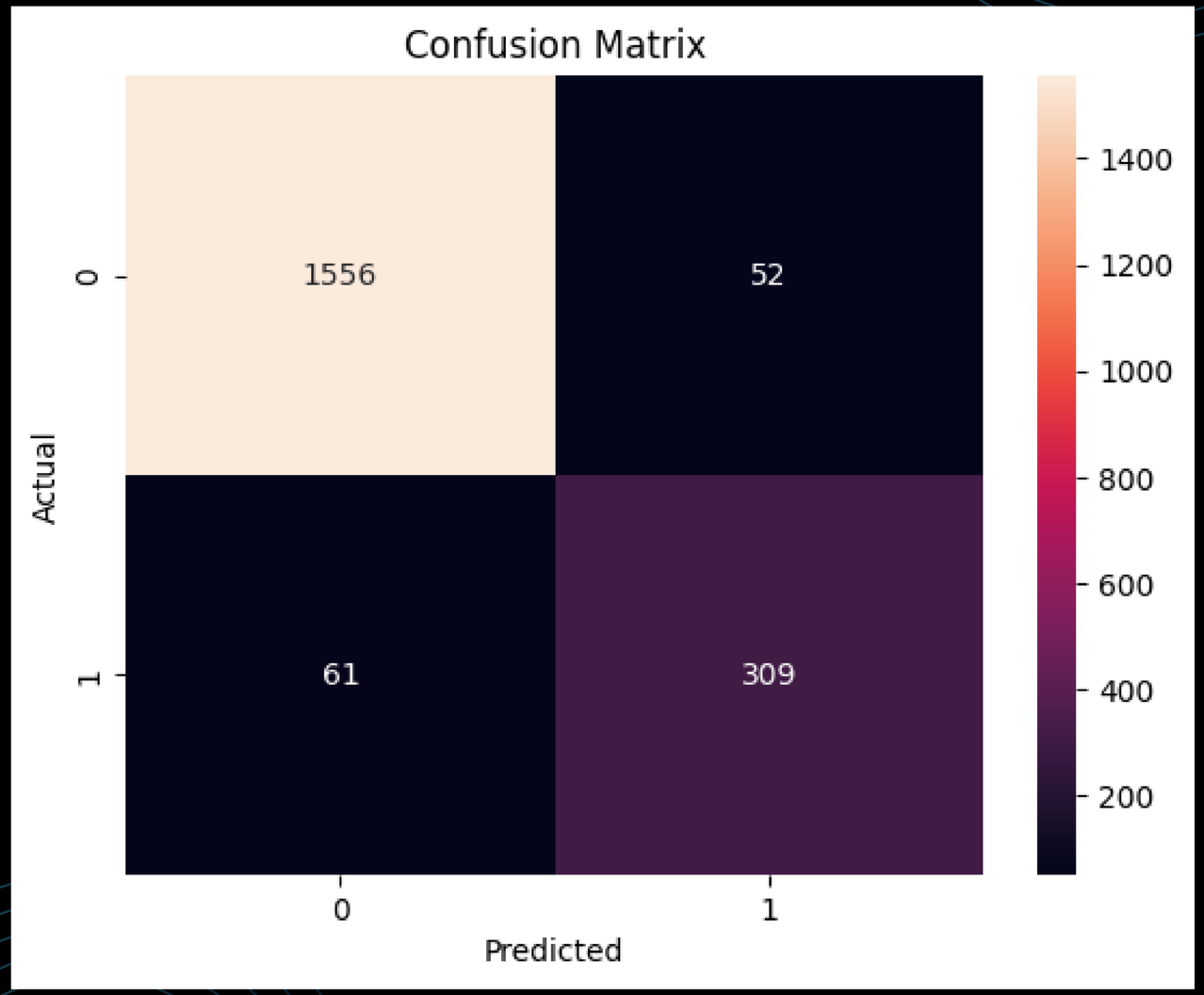
Precision: 0.6

F1 Score:
0.5116

MSE : 0.159



Random Forest Classifier



Accuracy:
0.94287

Classification Error:
0.0571

Precision:
0.8559

F1 Score:
0.84541

MSE : 0.05712

Support Vector Machine

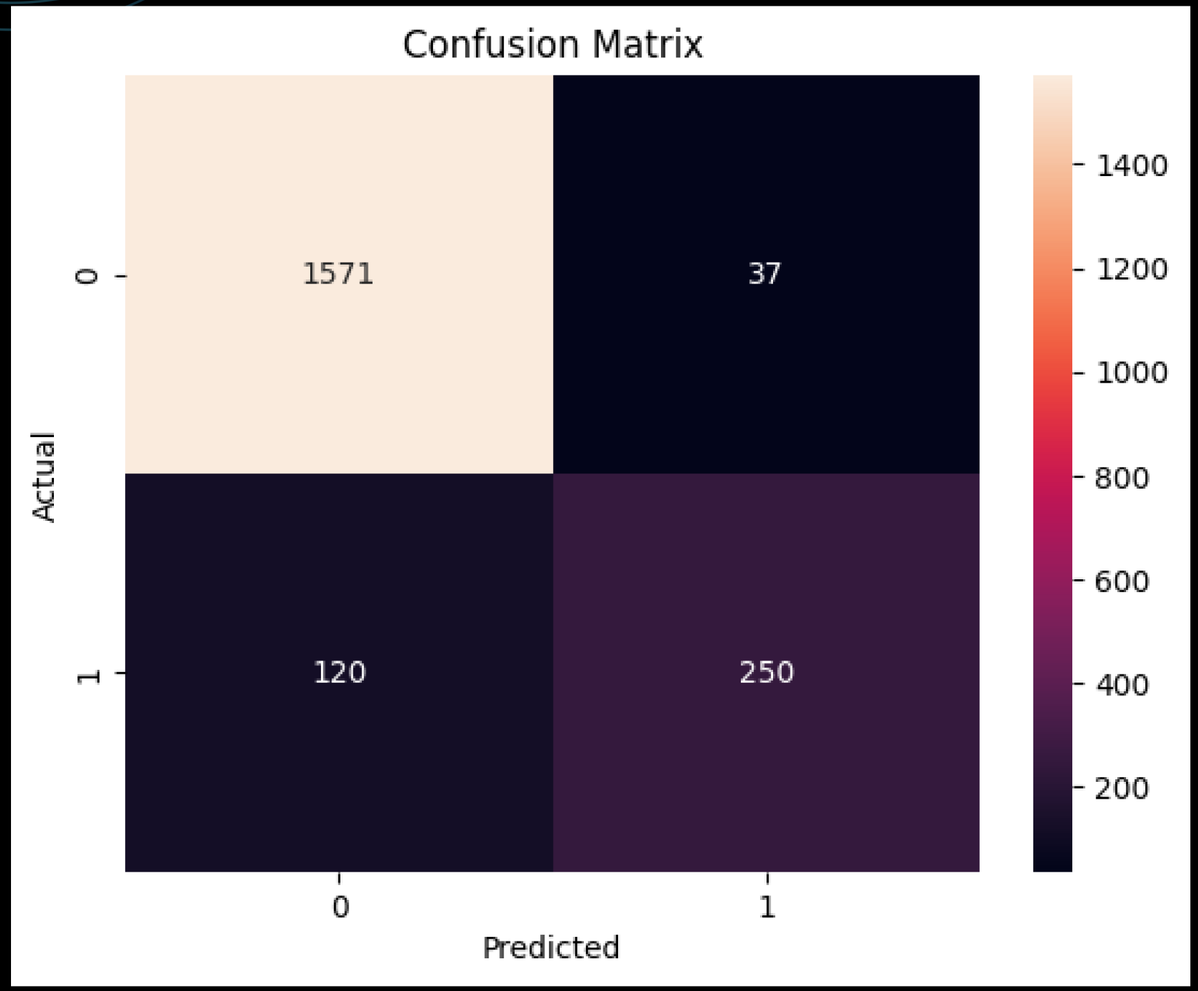
Accuracy:
0.920626

Classification Error:
0.0793

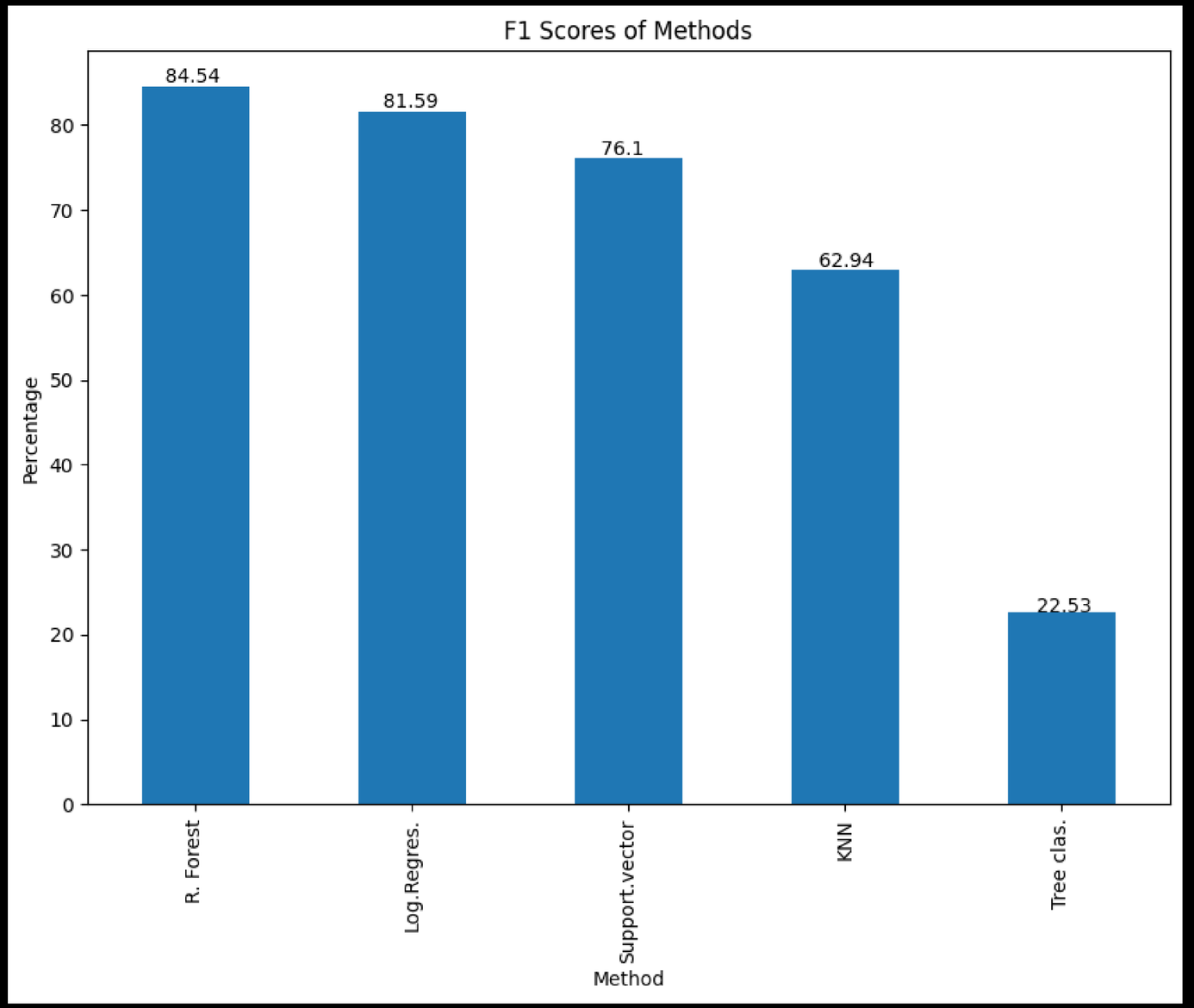
Precision:
0.871

F1 Score:
0.7610

MSE : 0.079



Plotting the F1 scores of all classification Model



Checking for Overfitting

Mean absolute error

Mae_train : 0.05581

Mae_test : 0.065

k-fold cross validation

Average Score: 0.60423

Prediction

	Job_Description	Actual_label	Predicted_label
0	Architecte d'IA : Conçoit et met en place l'in...	1	1
1	Chercheur en IA : Mène de recherches avancées ...	1	1
2	Éthicien (ne) de l'IA : Évalue et analyse le...	1	1
3	Vendeur/Vendeuse : Responsable de l'accueil et...	0	0
4	Secrétaire : Chargé (e) de l'organisation et...	0	0
5	Ingénieur (e) en informatique : Responsable ...	0	0
6	Serveur/Serveuse : Responsable du service et d...	0	0
7	Mécanicien (ne) automobile : Spécialiste de ...	0	0
8	Enseignant (e) : Chargé (e) d'instruire et...	0	0

Conclusion

From the project we can conclude that AI has a great potential to enhance the process of classifying LinkedIn experiences. By utilizing AI algorithms, it becomes possible to automate the categorization and analysis of job descriptions, skills, and qualifications listed in users' LinkedIn profiles.

AI algorithms can provide valuable insights and analytics by identifying trends, patterns, and correlations among different experiences. This information can help companies and recruiters gain a better understanding of the skills and expertise available in the job market, enabling them to make more informed decisions.

LIMITATIONS

Data Quality

Contextual Understanding

Limited Domain expertise

Evolving Terminology

Lack of Human Judgement

Future Scope

This project holds a lots of improvement in future. If given more time then I would have worked on transformers to make the results of the project more accurate and improve the performance of the AI system. I would have worked upon the BERT (Bidirectional Encoder Representations from Transformers) model. The advantage of using BERT and transformers in this context lies in their ability to understand the contextual relationships within the text, considering the surrounding words and phrases. This allows for more accurate classification, as the model can capture subtle nuances and dependencies that might be missed by traditional machine learning approaches. In summary, the use of transformers, particularly the BERT model, in the classification of LinkedIn experiences enhances the AI system's ability to understand the semantic meaning and context of the text. This leads to improved accuracy and performance in categorizing job domains, industries, or skill sets, ultimately benefiting job seekers and recruiters in their search for the right opportunities and talents.

Thank you

By-Little Beauty Bisoyi