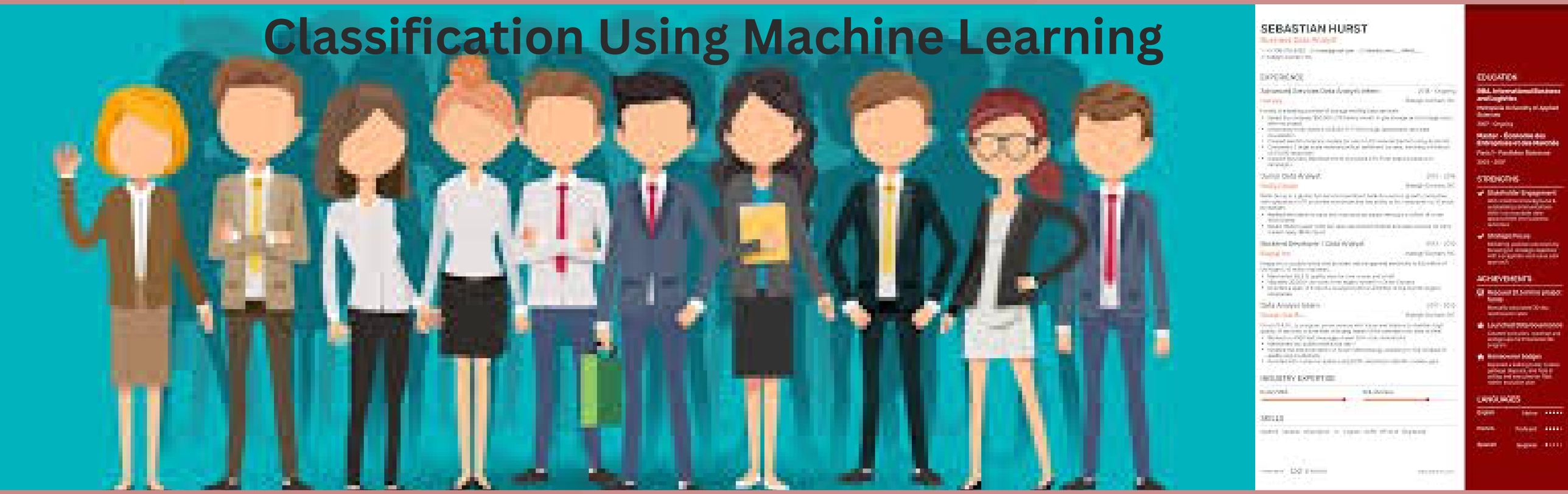
Resume classification and search

Automated Resume Parsing, Skill Extraction, and Classification Using Machine Learning



Mentors:
Snehal Shinde
B Harish

From Group : 2

From : Group 2

Name:

1: Muskawad Raman Sopan

2: Waghmare Priti Siddhart

3: Jahanvi Tammu

4: Mr.Bhanu Prasad

5: Kalpesh Patil

6: Chintha Likhita

Introduction :

Purpose:

Simplify resume screening with automated tools.

Features:

Extract text from resumes (.docx, .pdf).
Identify technical skills and experience.
Classify resumes using machine learning.
Search and rank resumes based on criteria.

```
[3]: # Load the file
     file_path = "Master Resume.docx"

[4]: # Load the .docx file
     document = Document(file_path)

[5]: # Extract text from the .docx file
     content = [paragraph.text.strip() for paragraph in document.paragraphs if paragraph.text.st

[6]: # Create a DataFrame
     df = pd.DataFrame({'content': content, 'profile': ['Master Resume'] * len(content)})
     df
```

[6]:

	content	profile
0	Chinna Subbarayudu M	Master Resume
1	DOB: 06th March 1994	Master Resume

Key Components & Technical Stack :

Technical Stack:

Python libraries: Streamlit, scikit-learn, TfidfVectorizer, PyPDF2.
Predefined technical skills for matching.

Machine Learning:

Gradient Boosting Classifier for classification.

Data Processing:

TfidfVectorizer for text similarity.
Regex for experience extraction.

Features and Functionality :

Key Functionalities:

Resume Extraction: Supports .docx and .pdf formats.
Text Processing: Extracts skills and experience.
Search and Filter: Allows criteria-based filtering.
Model Training: Uses Gradient Boosting Classifier.

EDA and Visualizations :

Exploratory Data Analysis (EDA):

Analyzing the distribution of key features such as skills, experience years, and resume categories.

Visualizations:

Bar Charts: Display skill frequency in resumes.
Word Clouds: Highlight most common skills across resumes.
Experience Distribution: Visualize the years of experience among resumes.

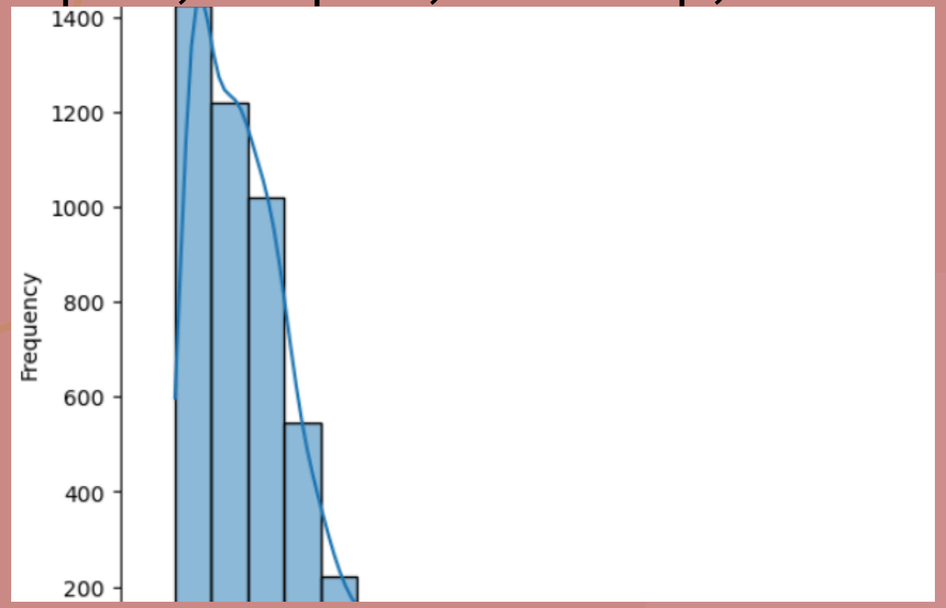
```
[7]: # Display the DataFrame
     print(df.head())
```

```
                                            content        profile
0                              Chinna Subbarayudu M  Master Resume
1                              DOB: 06th March 1994  Master Resume
2                               Nationality: Indian  Master Resume
3                                  PROFILE SUMMARY:  Master Resume
4  Having around 5.1 years of IT experience in de...  Master Resume
```

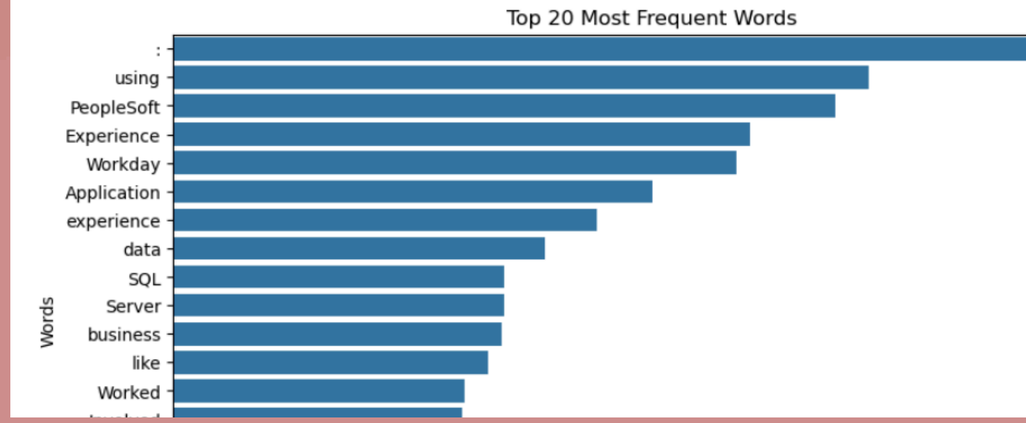
```
[8]: # Create a DataFrame
     df = pd.DataFrame({'content': content, 'profile': ['Master Resume'] * len(content)})
```

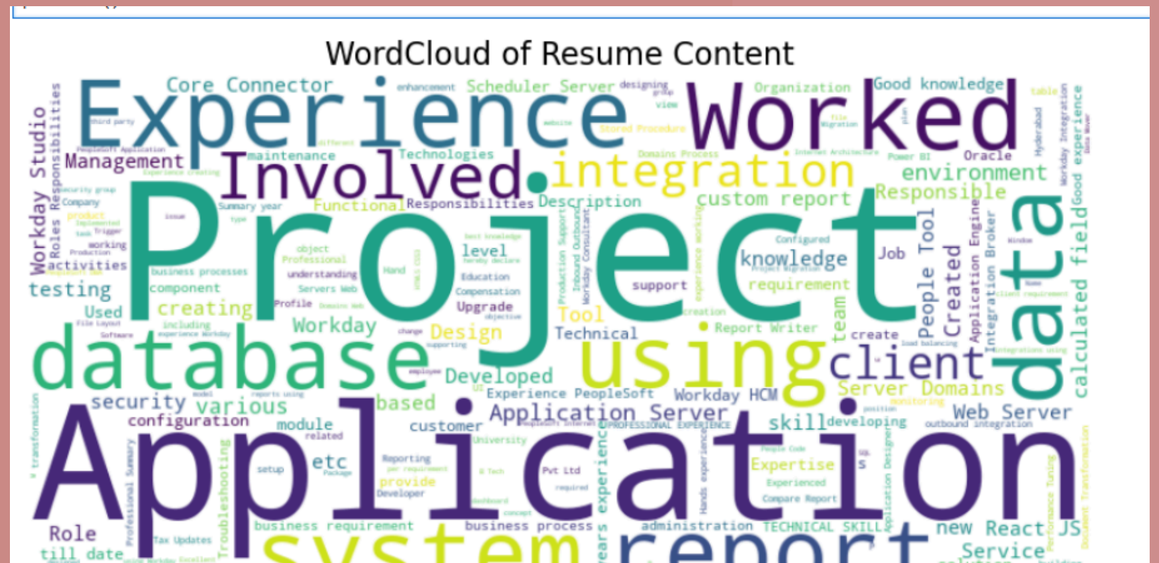
EDA & Visualizations part:

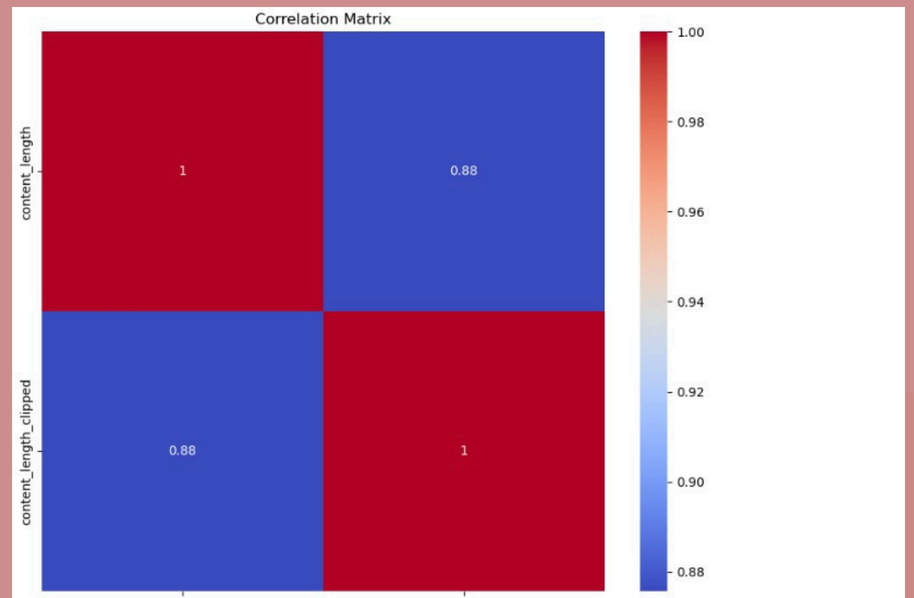
Plots:Histplot, Barplot,Heatmap,WordCloud,etc




```
sns.barplot(x=list(counts), y=list(words))
plt.title('Top 20 Most Frequent Words')
plt.xlabel('Frequency')
plt.ylabel('Words')
plt.show()
```







Model Evaluation & Improvements :

Model Performance:

Accuracy, Precision, Recall, and F1-score.

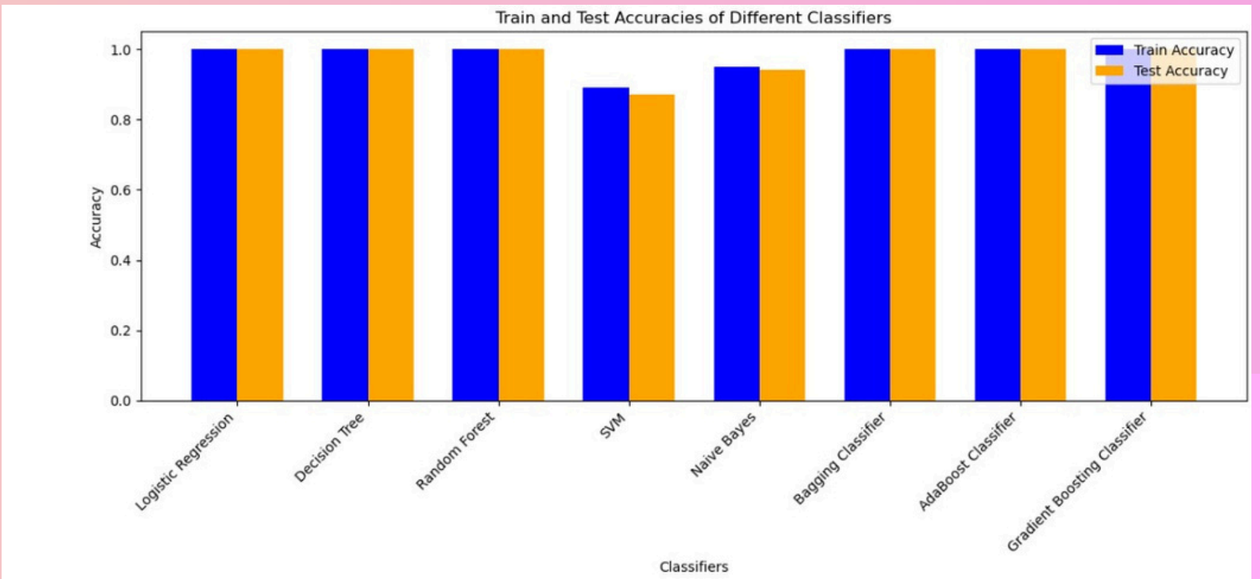
Improvement Techniques:

Hyperparameter Tuning.
Cross-validation to ensure model generalization.

```
train_accuracy = accuracy_score(y_train, y_train
test_accuracy = accuracy_score(y_test, y_test_pr

print(f"Training Accuracy: {train_accuracy:.2f}"
print(f"Testing Accuracy: {test_accuracy:.2f}")
```

```
Training Accuracy: 1.00
Testing Accuracy: 0.96
```



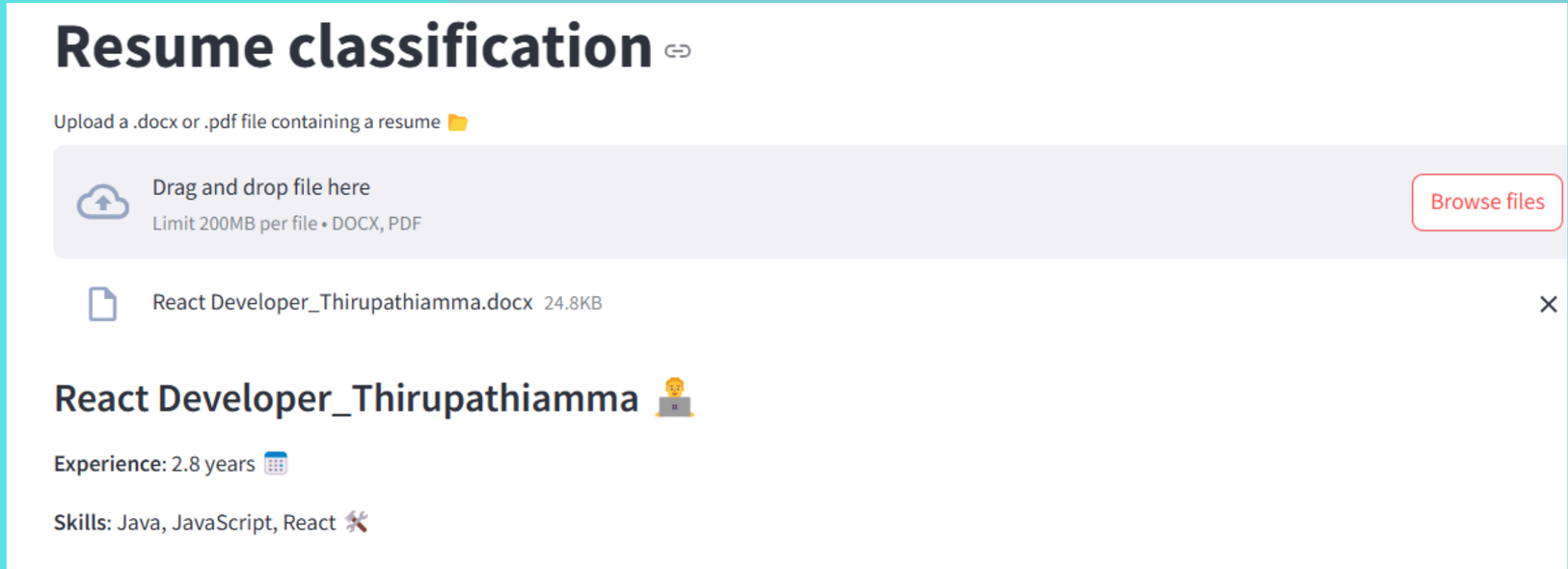
Deployment & Scalability :

Deployment Process:

Deploy the Streamlit app on cloud platforms (AWS, Heroku, Google Cloud).

Scalability:

Ensure scalability for handling large datasets of resumes.
Use parallel processing to speed up processing time.



Challenges and Considerations :

Key Challenges:

Handling varied resume formats.

Extracting structured data from unstructured text.

Considerations:

Balancing search specificity and broad applicability.

Future Enhancements and Conclusion :

Future Enhancements:

Integration with ATS systems.
Multilingual resume support.
Advanced classification techniques.

Conclusion:

Streamlines recruitment processes, saves time, and improves productivity.
Efficient solution for resume parsing and classification using advanced ML techniques.

THANK YOU