Automated Transcript Analysis to Optimize Recruitment Decisions

BUSSINESS PROBLEM

Research paper

Dataset

Python

DATA UNDERSTANDING

1.Load and Inspect Data
2.Basic Text Cleaning
3.Class Distribution Text
4.Length Analysis Word
5.Frequency Analysis
6.Stopword Analysis
7.N-gram Analysis
8.Sentiment Analysis
9.Named Entity
10.Recognition (NER)
11.Part-of-Speech (POS)
12.Tagging Text Similarity
13.Check Embedding
14.Visualization
15.Class-wise Text
16.Comparison

DATA PREPROCESSING

MODEL EVALUATION

1. Accuracy
2. Precision
3. Recall
4. F1-Score
5. Confusion Matrix
6. ROC-AUC Score
7. Log Loss

MODEL BUILDING

1.Whisper (for speech-to-text)
2.Logistic Regression
3.Support Vector Machine(SVM)
4.spaCy NER
5.BERT-based NER
6.LDA (LatentDirichletAllocation)
7. VADER
8.TextBlob
9.KMeans
10.HDBSCAN
11.XGBoost
12.Random Forest
13.GPT-based LLM (e.g., GPT-4, GPT-3.5)
14.LLaMA

DEPLOYMENT

Streamlit

MONITORING MAINTENANCE

END USERS

Power BI