# 2024S- T3 AML 2304 – Natural Language Processing

## Mid-Point Report

# Resume -Job Description ATS Score Generator


**Guide: Bhavik Gandhi**

**Group A**

Saurabh Gangwar(C0894380)

Rohit Kumar(C0895100)

Rewant Sharma(C0894265)

Beema Sajeev(C0891941)

Kuncheria Tom(C0900973)

Prince Thomas(C0894907)

Shanmuga Priyan Jeevanandam(C0889053)

Sravya Somala(C0907007)

Remya Kannankandari(C0895293)

## Table of Contents



**MECE TABLE**

| **Names** | **Task Assigned** |
|---|---|
| Saurabh Gangwar | Model Development(BERT) |
| Rohit Kumar | Model Development(BERT) |
| Rewant Sharma | Model Development(BERT), Interpretability(Lime) |
| Beema Sajeev | Model Development(Gradient Boost Regressor) |
| Kuncheria Tom | Project Board Creation(GitHub), Ats Score Generation by Google palm(LangChain) |
| Prince Thomas | Ats Score Generation by Paraphrase-MiniLM-L6-v2(Langchain), Model Development(Bi Directional GRU) |
| Shanmuga Priyan Jeevanandam | Model Development(BERT) |
| Sravya Somala | Model Development (TF-IDF), Interpretability(Lime), Presentation PPT |
| Remya Kannankandari | Model Development (BERT), Documentation |

## Abstract

The ATS Score Generation project aims to develop an AI-based tool for accurate resume evaluation against job descriptions. This tool reduces recruitment hassle for candidates and employers. Initial research, data collection, and preprocessing are complete, with initial models showing fair accuracy.

## Introduction

The ATS Score Generation tool evaluates resumes against job descriptions to provide a scoring system, streamlining the recruitment process. The open-source application allows students to verify their ATS scores, enhancing their job application success.

### Domain Overview

With 98% of Fortune 500 companies using ATS systems, the importance of these tools in recruitment cannot be overstated. By providing a free ATS scoring tool, we aim to level the playing field for students.

## Dataset Description

### Resume Dataset

Source: Resume Dataset (kaggle.com)

Content: 2400+ resumes in string and PDF formats.

### LinkedIn Job Postings Dataset

Source: LinkedIn Job Postings (2023 - 2024) (kaggle.com)

Content: 124K+ job postings.

## Objectives

### Expected Outcomes

- An AI-powered ATS scoring tool.
- An open-source platform for students.
- Improved resume quality and higher chances of interview shortlisting.
- Increased awareness of resume optimization importance.

### Scope

Included:

- Data Collection
- Data Preprocessing
- Feature Engineering
- Model Building
- Evaluation
- Deployment

### Out-of-scope:

- External marketing and sales activities

## Methodology

### Data Preprocessing Pipeline

- Handle Digital Symbols
- Handle Slangs/Emojis/Abbreviations
- Named Entity Recognition (NER)
- Spell Corrector
- Part-of-Speech Tagging
- Handle Contractions/Negations
- Remove Stop Words
- Stemming or Lemmatization
- Lowercase Conversion
- Ngrams

### Feature Engineering

Bag of Words: Converts text into count features.

TF-IDF: Balances word frequency against uniqueness.

Word Embeddings: Captures semantic meaning and context.

## Model Development

**TF-IDF Model:** Transforms text into numerical vectors to calculate similarity features for accurate ATS scores.

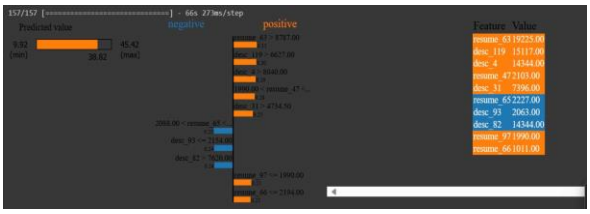
**Gradient Boost Model:** Cleans text data, tokenizes, and vectorizes through TF-IDF or word embeddings. Measures similarity and trains a model for accurate ATS scores.

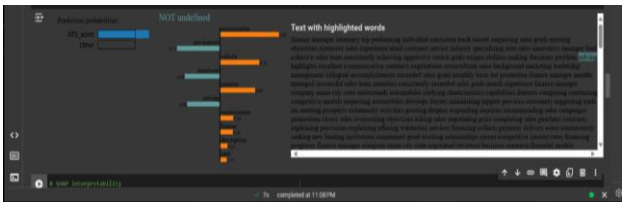
**Bidirectional GRU Model:** Processes text in both forward and backward directions for detailed comprehension, enhancing the gradient boosting model.

**BERT:** Captures semantic meaning and context for effective resume-job description matching.

## Model Interpretability

**LIME (Local Interpretable Model-agnostic Explanations):** Explains individual predictions, enhancing trust and understanding of the model's decisions, which is crucial for fine-tuning and ensuring fairness in the ATS.




**GitHub:** https://github.com/kuncheriatom/ATS_SCANNER

## Project Board Walkthrough

- **Platform Used**: GitHub
- Project Board · ATS Resume Checker Project Board (github.com)

## Work Done

### Tasks and Milestones

1. Research and Data Collection: Gathered various job descriptions and resumes.

2. Project Management: Setup project board, defined roles, and responsibilities.

### Key Highlights

- Diverse datasets for robust model training.

- Continuous improvement of data preprocessing and feature engineering.
- Finalizing model evaluation and website integration.

## Work in Progress

### Active Tasks

1. Improve Data Preprocessing Techniques: Ensure superior quality input data( 80% done).

2. Enhance Feature Engineering Processes: Add methods like deep word embeddings.

3. Iterate Modeling and Evaluation: Improve resume scoring accuracy ( Aiming for a minimum of 85%).

### Timeline

- Data Preprocessing: To be completed in one week.
- Model Evaluation: Intensive tuning in two weeks.
- Website Launch: End of the project.

## Challenges, Issues, and Mitigation Strategies

### Challenges in Merging Resume and Job Datasets

1. Using LangChain for ATS Score Output

2. Using GPT-2 for ATS Score Generation

### Mitigation Strategies:

**Ats Score generation using LLM chain:** LangChain runs natural language processing, while the Prompt Template guides the LLM Chain in reviewing the resumes. Google PaLM provides a language model that understands and compares key skills and experiences. The LLM Chain processes inputs to create an ATS score, indicating how well a candidate's resume fits the job description. This automation enhances the speed and quality of candidate screening.

### Scope of Improvement

1. Implemented robust validation checks before model training to ensure data quality and consistency.

2. Exploration of optimization techniques to better utilize available computing resources, aiming to reduce training time and costs.

## Future Work and Next Steps

1. Completion and Validation of Data Preprocessing and Feature Engineering
2. Model Building and Rigorous Evaluation
3. Integration of the Scoring System into the New Website

## Conclusion

The project is progressing as initially planned, with preliminary models indicating satisfactory scores when ranking CVs. The team remains optimistic and dedicated to achieving the set goals and delivering a viable, functional tool capable of generating accurate ATS scores for users' resumes.