

# Resume Scanner

The Resume Scanner tool, aims to simplify and expedite the job application process, benefiting both candidates and recruiters. Through the use of advanced NLP techniques, it seeks to provide more accurate and efficient resume analysis and job role matching.

The problem selected for this project is to build a resume scanner tool. The primary motivation for this project is to address the challenges candidates face when applying for job roles, specifically the time-consuming process of shortlisting candidates based on their resumes. This aims to streamline this process by leveraging NLP techniques to achieve the following objectives:

- **Resume - Job Description Similarity Score:** Calculate the similarity score between a candidate's resume and a job description to assess how well the resume matches the job requirements.
- **Multi-class Resume Classification:** Categorize resumes into predefined classes or job roles, making it easier for candidates to identify suitable job roles for their resume.
- **Resume Analysis:** Provide comprehensive analysis and insights into a candidate's resume, including key skills, qualifications, and experience.

**Dataset:** [https://github.com/florex/resume\\_corpus/tree/master](https://github.com/florex/resume_corpus/tree/master)

## **Key concepts:**

Pretrained NLP (Transformers Based Networks)

Recurrent Network (Any RNN model LSTM, GRU)

**NLP Tasks:** Document Similarity, Document Classification

**Model Customization:** The project will involve fine-tuning pretrained models for specific tasks such as document similarity and classification. While we will initially use off-the-shelf pretrained models, fine-tuning and customization will be essential to adapt the models to the unique requirements of resume analysis.

## **Packages:**

- **PyTorch:** As our primary deep learning framework for building and customizing NLP models.
- **Pandas:** For data preprocessing, manipulation, and organization.
- **NumPy:** For numerical operations and array handling.
- **Scikit-Learn:** To assist in machine learning tasks, such as classification and model evaluation.
- **Seaborn:** For data visualization and performance analysis.

**NLP Tasks:** We will focus on the following NLP tasks:

- a. **Document Similarity:** To compute the similarity score between a candidate's resume and a job description.
- b. **Document Classification:** To categorize resumes into predefined job roles or classes.
- c. **Resume Analysis:** Extracting relevant information from resumes, such as key skills, qualifications, and experience.

# Resume Scanner

## Performance Metrics:

We will judge the performance of our model using the following metrics:

1. Similarity
  - a. Jaccard
  - b. Cosine
2. Classification
  - a. F1 score
  - b. Confusion matrix
  - c. Recall
  - d. Precision

