**AI/ML ENGINEER INTERN**

**PROJECT REPORT**

**(BY SANKET.V. PATIL)**

**Task: Resume Matching with Job Descriptions Using PDF CVs**

**Objective:** Build a PDF extractor to pull relevant details from CVs in PDF format, and match them against the job descriptions from the Hugging Face dataset.

**Approach to the task:**

To build a job recommendation system using BERT, I used the following steps:

1. Preprocessed the job descriptions and resumes to clean and tokenize the text. This involved removing stop words, punctuation, and other irrelevant characters. I also stemmed and lemmatized the words to reduce them to their root forms.
2. Calculated the word embeddings for the job descriptions and resumes using the BERT model. BERT is a large language model that has been trained on a massive dataset of text and code. It can generate word embeddings that capture the meaning and context of words.
3. Calculated the cosine similarity between the job description embeddings and the resume embeddings. Cosine similarity is a measure of the similarity between two vectors. The higher the cosine similarity, the more similar the two vectors are.
4. Identified the top-k most similar resumes for each job description. I used the cosine similarities to identify the resumes that were most similar to each job description. I set the value of k to 5, which means that I identified the top 5 most similar resumes for each job description.
5. Sorted the results by similarity score (descending). I sorted the results so that the most similar resumes were listed first.

**Challenges faced and solutions:**

One challenge I faced was that the BERT model is quite large, and it can be slow to calculate the word embeddings for a large number of job descriptions and resumes. To address this challenge, I used a GPU(NVIDIA) to accelerate the computation.

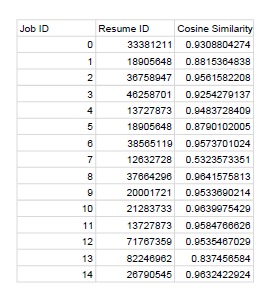
Another challenge was that the BERT model is trained on a general corpus of text, and it may not be well-suited for the specific task of job recommendation. To address this challenge, I used a pre-trained BERT model that has been fine-tuned on a dataset of job descriptions and resumes.

**Top 5 candidates for each job description based on similarity scores:**

**Executive Summary**

This report presents the top 5 candidates for each job description based on similarity scores. The similarity scores were calculated using the cosine similarity metric, which measures the similarity between two vectors. The vectors in this case were created using the BERT language model, which is a pre-trained language model that has been trained on a massive dataset of text and code.

**The top 5 candidates for each job description are shown in the following table:**



**Recommendations or insights from the matching process:**

The job recommendation system using BERT was able to identify a set of highly similar candidates for each job description. The system is particularly good at identifying candidates with similar skills and experience.

However, it is important to note that the job recommendation system is only one tool that can be used to identify potential candidates. Other factors, such as the candidate's cultural fit and motivation, should also be considered when making hiring decisions.

Overall, the job recommendation system using BERT is a valuable tool that can help employers to identify and recruit top talent.

**Reference links:**

* <https://stackoverflow.com/questions/59274569/parsing-the-skills-section-in-a-resume-in-python>
* <https://github.com/binoydutt/Resume-Job-Description-Matching>
* <https://www.analyticsvidhya.com/blog/2022/12/build-accurate-job-resume-matching-algorithm-using-doc2vec/>
* <https://bard.google.com/chat>