

## **Project MVP**

I am working on my Resume Parser/Screenner with NLP Project.

In this project user will upload a txt or pdf file than resumes are categorized and summarized and out will be showed on web app.

My project is divided into two phases:

Phase 1: Categorization of Resumes and showing it in streamlit web app

Phase 2: Summarized the Resume and showing it in streamlit web app

Finally I have to integrate python files with streamlit to show output on web.

### **So Far what I did ?**

I worked on Resume Categorization with have following steps:

- 1) Exploratory Data Analysis : Here in this step I explored my dataset and made some visualizations for better understanding.
- 2) Data Pre-processing: In this step I made my data clean by removing all the hashtags, URL's and unwanted characters. As now data is clean I converted them into tokens. After generating tokens I used TFIDF Vectorizer to vectorize the tokens.
- 3) Train-Test-Data and modelling: Here comes the splitting of data as train and test data. I allocated 80% of my data for training and 20% data for testing. After that I founded the Training and Validation accuracy.
- 4) Analyzing: At last I analyzed my data by finding most occurring keywords.

### **Conflicts I faced ?**

So far there was no issue with the project but I think they may a rise as I go further.

## **Next Step ?**

Next Step of the project is Phase 2 and Streamlit app (Web app):  
I need to work on summarization by taking the dataset and training it.  
After that I need to integrate both Phase 1 and Phase 2 with Web app.