

# Abstract

**College Name:** NYSS Datta Meghe College of Engineering (DMC)

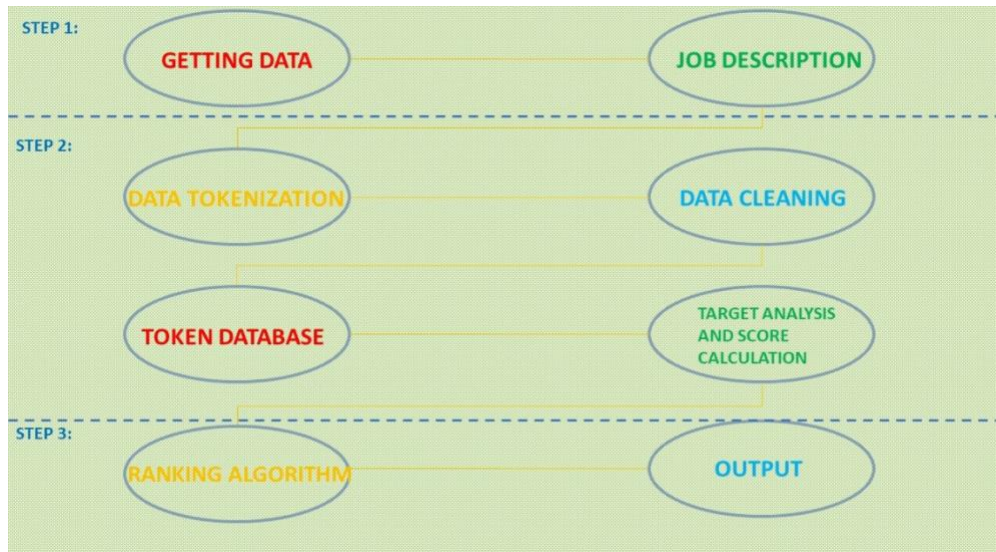
**Team Details:**

Name	Member/Leader	Year(Second/Third/Final)
Sawant Neha Ganesh	Leader	Third
Sankpal Jaykumar Shivappa	Member	Third
Shinde Jayesh Sharad	Member	Third
Singh Karandeep Sandhu	Member	Second

**Problem Selection:** Resume Parser (RPA)

Technology today has made it possible to send a resume within the tap of a button. Hundreds of resumes are being received for a particular job posting. This makes the job of an HR department especially difficult as it is impossible to peruse through each of the resumes and shortlist the candidates. Moreover, each resume has a different format. There is a need to extract the accurate information from the resume and store it in the database. In order to make the job of a recruiter easier, we propose an automated system that does most of the preliminary filtering and presents the data in a standard format. This project is dealing with multiple domains; hence accuracy can't be avoided in this topic as it is one of the key and main features in recruiting of jobs. Information accuracy and extraction plays an important role in resume analytics as the unstructured resumes need to be normalized into a standardized format for further processing and system should have quick retrieval of information. Accuracy and quick retrieval of information is most challenging part of this problem from others. Apart from being challenging this topic is much interesting to work. It has various development scopes not only in backend but also in frontend. ML and UI development which hold good command of us encourage us to work, not only this but to cope up with lacking accuracy in the existing system inspired us to choose this topic.

In this process, the CVs being provided as input would be cleansed to remove special or any junk characters that are there in the CVs. We got the clean dataset after these steps having no special characters, numbers or single letter word. The dataset is split into the tokens using the NLTK tokenization. Further, the pre-processing steps are applied on tokenized dataset such as stop word removal, stemming, and lemmatization. The raw CV file was imported and the data in the resume field was cleansed to remove the numbers and the extra spaces in the date. Then data masking will be done for data like special characters, etc. After completing the above steps, we are going to use processes such as stop words removal and will ensure that all are root words. Then we are going to use Random Forest, Multinomial Naive Bayes, Logistic Regression, Linear Support Vector Machine Classifier for target analysis and score calculation. And in the last step we will use ranking algorithms to rank the resumes by using Content Based Recommendation using Cosine Similarity, k-Nearest Neighbours. Overview of approach is given in diagram given below



Platforms we are going to use are Vs Code, PyCharm and Jupyter notebook. We are going to use Python, HTML, CSS, Js as coding languages and Django, React Js, Bootstrap as frameworks. We are going to use Rest API. The database which we are going to use is PostgreSQL and we are going to download datasets from Kaggle.

Fortnightly targets as follows:

Week 1-2: Home section, models, serializers, database designing

Week 3-4: Applicant screen (designing), backend views, getting data, cleaning data

Week 5-7: Applicant screen (Connection), Recruiter screen design, URL's, training dataset

Week 8-10: Recruiter screen connection, deployment, remaining backend work, testing dataset

Week 11-12: Remaining work and testing application

Our members are experienced and has gained knowledge in particular domain which makes our team strong and work together. The projects which we had undertaken are Alumni portal, Training & Placement portal, Ham-tech, Hospital Automation system which has been successfully deployed and are in day-to-day use. Strengths of our backend developer consists of our team members been the lead in Machine learning domain at Google Developer Student's Club (GDSC) also a member of DMCE-Alumni Tech member, Code chef-DMCE having a rich problem-solving skill, as well as part of technical team member for CSI-DMCE, GDCE-DMCE web-development team. Frontend developer having a very keen observation of particular elements and are quite creative which makes every website very simple and user-friendly to use. Along with this our team has been selected at various in college and inters college level. We have got selected in the top twenty for final round of international project competition under software category organized by DBCE, Goa. We had also put one step in entrepreneurship by using our knowledge and making the website for start-ups. Our team is always ready to explore and learn new things. We always try to keep everything planned and organized. We always look forward to work enthusiastically. Our team has good communication on various discussions forums and platforms. Each member of our team contributes equally in projects while development which makes us grabbing new idea easier and applying them faster. Our team is supportive diverse and organized in work. Along teams plans clear focus on goals and make them achieve.