**PSG COLLEGE OF TECHNOLOGY**

**Department of Applied Mathematics and Computational Sciences**

**IX Semester M.Sc DS 2020 - 2021**

**15XD96 Information Retrieval Lab**

**Package Abstract**

**Phase 2**

**Title:** Resume Short-listing : Ranking and Querying

**Roll No:** 16PD22

**Name:** Nehal M

**PHASE 2 IMPLEMENTATION**

**Vectorizing:**

With the help of indexing we build document vectors using tf-idf approach(both tf and idf normalized). We get a 570\*3500 dimenstional matrix where 570 is the no of documents and 3500 is the dimension of each vector.

**Querying**:

Here query is of two types. We can either query documents based on keywords(eg: skillset,college,etc) or based on job description document. For the latter approach we have to upload JD do all the pre-processing (as in phase1) and then vectorize the query. We integrated it in django framework.

**Ranking**:

For ranking initially cosine similarity is used to retrieve top N documents. List of selected candidates is available for cross verification. Also Binomial model was used to rank the documents. Both the ranking methods were evaluated using precision@R metric.

**User interface:**

UI for querying is built with the help of django framework. Database for resume uploading is also created.