Summarization Engine Documentation

**\*\*Imp\*\* –** These files are also present in the backend folder for easier integration with the backend. The files present in Summarization Engine folder can be used for reviewing purposes.

**Requirements** – All the required python libraries are mentioned in the pip install.txt

**Description of Files** –

1) SalaryExtract.py – Salary is extracted from the Job description using the convention followed normally, for eg, CTC – 7-10 lacs, Rs. 10 lakhs, Rs. 70,000 – 80,000 and more cases have been added to handle.

2) VacancyExtract.py – Vacancies, if mentioned, are extracted from the Job description using keywords like ‘vacancy’, ‘posts’, ‘positions’ etc.

3) DateExtract.py – Any deadlines mentioned in the job description(of form DD-MM-YYYY) will be extracted. Any date of format DD/MM/YYYY or DD-MM-YYYY can be extracted.

4) JobTitleExtract.py – The job title will be extracted from the job description by matching words with the existing database consisting of 3000+ job title records. The closest one will be selected by a scoring system which determines how relevant a title is.

5) Job\_Title.txt – A list of 3000+ prominent job titles from around the world.

6) summary.py- The job summary will be extracted from the job description, using nltk + summa library. The non-important stop-words are removed and a short summary is generated.

7) code.py – The main python file which imports the above .py files as modules. Along with date, vacancies, salary, job title, job summary, the skills listed in the job description can be distracted. The location is also extracted directly by using a list of all prominent cities and matches the words in Job description with the list.

8) pincode(reduced).csv – It contains the list of all prominent cities.

9) pip install list.txt – Contains all the required dependencies.

10) Recommendation.py – Contains the job recommendation code which is used for sending recommendations to a user based on the skills entered by the user. It works on the principle of cosine similarity, which is used to list out recommendations. Currently only 1 job is listed for a user, but we can set a limit on how many jobs will be posted to him as recommendation.

Note – LocationExtract(Advanced\_Feature).py file contains the Trie Data Structure Code for extracting Location from the uploaded JD.