Team member's details :

Group name: one,

name: Srinivas Devarajula,

Email: srinivasdevarajula21@gmail.com,

Country: United kingdom,

Company: Anglia Ruskin university alumni,

Specialization: NLP,

Problem description: Resumes contain surfeit information that is not relevant for the HR/authority, and they have to manually process the resumes to shortlist the promising candidates for them. And, thus making the shortlisting task a herculean task for HR. By making use of the NER(Named Entity Recognition) model of NLP this problem can be solved by finding and classifying the entities that are present in each resume into predefined classes such as person name, college name, academics information, relevant experiences, skill set, etc.

Projected Business advantages:

- Adherence to a drug directly impacts the sales dollars a product can generate
- Optimize and improve the efficacy of clinical trials
- Target specific patient populations more effectively
- Better insight into patient behaviour to improve drug effectiveness

PROJECT LIFECYCLE

- WEEK8(Jan 20 – jan 26)
  - Data understanding
  - Type of data for analysis
  - Problems in the data
- WEEK9(jan 27 – feb 02)

  Data Cleansing and Transformation
- WEEK10(feb 03 – feb 09)

  EDA and Final recommendation
- WEEK11(feb10 – feb16)

  EDA presentation for business users
- WEEK12(feb 17 – feb 23)

  Model Selection and Model building

- WEEK13(feb 24 – feb 28)

  Final Report, Code, Ppt submission

DATA INTAKE REPORT

Name: NLP - Resume Extraction

Report date: 20/01/2023
Internship Batch: LISUM16: 30
Version: <1.0>

Data intake by: Srinivas Devarajula

Data intake reviewer: Srinivas Devarajula

Data storage location: https://github.com/SRINIVASDEVARAJULA/week7/tree/main

Tabular data details:

Total number of observations	
Total number of files	1
Total number of features	
Base format of the file	.json
Size of the data	1.1 mb