Week 7 Deliverables

Group Name: Team Fields

Name: Kenneth Fields

Email: kfields110@gmail.com

Country: United States

College: University of Illinois

Specialization: Resume Extraction: NLP

Problem Statement:

> Resumes contain surface 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.

Project Lifecycle:

1) Data Discovery -
    a) Accumulate data relevant to the problem statement.
        i) Resume Data Set given
2) Data Preprocessing-
    a) Clean and prepare data to be used in models.
3) Plan the Model
    a) Develop a model to extract relevant data from resumes.
    b) Save data from model.
4) Deploy the Model
    a) Deploy a model that can be used in real-time.
5) Deadline - 9/30/22

Data Intake Report

Name: Resume Extraction

Report date: 9/19/22

Internship Batch: LISUM11:30

Version: 1.0

Data intake by:Kenneth Fields

Data intake reviewer: Kenneth Fields

Data storage location: https://github.com/kfields110/Data_Glacier_Project

Tabular data details:

Total number of observations 200

Total number of files: 1

Total number of features 2

Base format of the file .json

Size of the data: 1.1 Mb

Note: Replicate same table with file name if you have more than one file.

Proposed Approach: Parse json file with juypter notebook and extract relevant details of each resume.