Week 7 Deliverables:

Team Details :

- Group Name: The NLP Resume Extractors
- Members:
  - Connor Bryson:
    - Email: cdbryson@iu.edu
    - Country: USA
    - College: Indiana University Indianapolis
    - Specialization: NLP
  - Alison March:
    - Email: alison.march@colorado.edu
    - Country: USA
    - College: University of Colorado, Boulder
    - Specialization: NLP
  - Haseeb Javed:
    - Email: Haseebajved4652@gmail.com
    - Country: USA
    - College: CUNY Queens College
    - Specialization: NLP

GitHub Repository Link: https://github.com/haseebjaved4652/ResumeAtraction

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.

Business understanding:

With the increase in people earning degrees and job applicants on the rise, businesses can utilize NLP to help narrow down their pool of candidates to the ones that meet the criteria. By using NLP, businesses can break down a person's resume and restructure it to then organize candidates by what skills they have and whether they relate with the job description. This project is useful for hiring managers to better understand their candidates and gain data to support the reason they hired one candidate over another.

Project Lifecycle for NLP Resume Extraction

Week 7 due 10/19/2023 Deliverables(Deciding Project Focus)

1. Provide team members details: Group Name, Name, Email, Country, College/Company, Specialization
2. Problem Description
3. Business Understanding
4. Project Life Cycle along with deadline
5. Data Intake Report
6. Github Repo Link - each team member should create their own branch before committing code

Week 8 due 10/26/2023 Deliverables(Brainstorming)

1. What type of data you have got for analysis
2. What are the problems in the data(number of NA values, outliers, skewed etc)
3. What approaches are you trying to apply on your data set to overcome problems like NA value, outlier etc and why?
4. Github Repo Link

Week 9 due 11/2/2023 Deliverables(Data Cleaning + Transformation)

1. Data Cleansing and transformation
2. Try at least 2 techniques to clean the data
    a. For NA values: mean/median/mode/Model based approach
    b. Identify and handle outliers
    c. For NLP projects try different featurization techniques and clean the data using regex and python
    d. Each team member should work on different data cleansing approach

Week 10 due 11/9/2023 Deliverables(EDA + Final Recommendations)

1. EDA performed on the data (we can discuss NLP methods and plans of implementation during our meeting, so methods TBD)
2. Final Recommendations

Week 11 due 11/16/2023 Deliverables(EDA Presentation + Proposed Modeling Technique)

1. EDA presentation for business users
2. Last slide of EDA should be dedicated to technical user which should contain recommended models for this data set

Week 12 due 11/23/2023 Deliverables(Model Selection and Model Building/Dashboard)

1. Select the base model then explore 1 model of each family if it's classification problem,
2. Select 1 model for Linear models, 1 model for Ensemble, 1 model for boosting and other models if we have time
3. Upload the code of each team member and other deliverables in the single repo and share the repo url.

Week 13 due 11/30/2023 Deliverables(Final Project Report and Code)

1. Discuss the solution of each member and select the solution which is best and is per the requirement
2. Create Powerpoint presentation for the project results
3. Deliver and merge final code on Github and provide url