Project

Resume Parsing and Classification Using Named Entity Recognition (NER)

Data Science Intern

Monisha Shree Senthil Nathan

IU International University of Applied Sciences

Germany

LISUM32, Data Glacier

monishashree.career@gmail.com

Contents

1. Problem Description

HR departments face the challenge of manually processing a large number of resumes, which is both time-consuming and labor-intensive. Each resume contains various sections such as personal details, education, work experience, and skills. By using Named Entity Recognition (NER) models in Natural Language Processing (NLP), we can automate the extraction and classification of these entities, streamlining the resume screening process and making it more efficient and accurate.

2. Business Understanding

Objective:

Develop an automated system to parse resumes, extract relevant information, and classify this information into predefined categories.

Benefits:

- **Efficiency:** Automation reduces the time and effort required by HR personnel.
- **Accuracy:** Ensures consistent and accurate extraction and classification of information.
- **Scalability:** Handles large volumes of resumes, making it suitable for organizations of any size.
- **Cost-Effective:** Reduces manual labor, translating into cost savings.

3. Project Lifecycle and Deadlines

Week	Plan	Deliverables	Date
Week 7	Initial Documentation	– Problem Description – Business Understanding – Project Lifecycle and - Deadlines – Data Intake Report – GitHub Repository Link	18-05-2024
Week 8	Data Analysis Preparation	– Data Understanding Analysis of Data: Type of data, issues such as NA values, outliers, skewness Approaches to handle data issues: NA values, outliers, etc.	25-05-2024
Week 9	Data Cleansing and Transformation and Feature Extraction	– Two Techniques for Data Cleaning (handling NA values, outliers, etc.) – NLP Techniques for Data Cleaning and Featurization – Report and github link	01-06-2024

Week 10	Exploratory Data Analysis (EDA)	– Exploratory Data Analysis – Final Recommendations – Report and github link	08-06-2024
Week 11	EDA Presentation and Modeling Techniques	– EDA Presentation and Proposed Modeling Techniques – Report and github link	15-June-2024
Week 12	Model Selection and Building	– Model Selection and Model Building/Dashboard – Base Model Selection – Explore Models from Different Families (e.g., Linear, Ensemble, Boosting) – Report and github link	22-June-2024
Week 13	Final Project Report and Code	– Presentation – Source code	29-June-2024