**Resume classifier**

**Syed Hani Haider, Dinesh Sai Pappuru, Nikitha Kanaparthi**

**Introduction**:

To be considered for a job one must prove that they are qualified for it. The first step in doing so is providing a resume which will show their professional skills. Almost every company requests its applicants to provide a resume whenever they are applying for jobs. However, hiring the correct talent is a challenging aspect because a company may get a high volume of resumes. The recruiter would not have time to go through and review each one of the applicants. The task that our team is trying to do is to categorize the resumes and select the best talent among the lot.

**Approach to the solution**:

**Dataset**:

We are going to use two datasets that we found on kaggle. Each contains 2 columns, the first column contains the category that the resume would fall under and the second contains details on the resume ranging from applicants' experience and skills to projects that they have worked on in the past.

**Approach**:

The plan is to use the nltk library to remove unnecessary stop words and special characters from the resumes. The next step is to tokenize and either create a bag of words or use TF-IDF vectorization on the cleaned dataset. After cleaning the dataset, various classification algorithms such as Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Multi-Layer Perceptron, and XGBoost will be applied for training and testing the data. The model will then be run on user input, and the best fit applicant will be determined by filtering out the resumes based on the job description.

**Related Work:**

The first paper we found "Machine Learning Techniques for Resume Screening and Shortlisting” discusses the use of various machine learning techniques to screen and shortlist candidates based on their resumes. The paper explains that Natural Language Processing (NLP) is used to screen resumes and understand the text. Therefore, it can help reduce outsourcing to talent acquisition companies. They use algorithms such as K-Nearest Neighbors, Support Vector Machines, and Named Entity Recognition (NER) to achieve classification. We are planning to use similar techniques for preprocessing and predicting on our dataset. [1]

The second paper we found “Automated Resume Screening system using Natural Language Processing (NLP) and Similarity Analysis techniques” is made to filter job applicants based on their resumes.The resume data is preprocessed to extract relevant information (personal details, education, work experience, and skills). It then uses NLP techniques like tokenization, stemming, and stop-word removal to convert the text into a numerical representation for further analysis. The model then uses a similarity analysis algorithm to compare the resume to a given job description and assigns a score to each candidate depending on how well their resume matches to the job requirements [2].These paper will help with implementing different algorithms and test our hypothesis.

**Assessment methodology**:

We want to compare the results of different models in terms of accuracy, specificity, and sensitivity based on the formulas in Figure A, where TP and TN stand for true positive and true negative, FP and FN stand for false positive and false negative, and the lowercase *n* is the total number of measurements. To validate the model, accuracy, F1 score, precision, and recall will be used. Matplotlib will be used to visualize the legitimacy of the results and determine the best one.

**Citation:**

1. Pokharel, Prasuna. (2022). RESUME PARSER USING NLP. 10.13140/RG.2.2.10323.25127/1.
2. Kinge, B. (n.d.). *Resume screening using machine learning and NLP - IJSRCSEIT*. https://ijsrcseit.com/paper/CSEIT228240.pdf