RESUME REVEALER

# INTRODUCTION

The ResumeRevealer project is an advanced resume analysis tool that supports various document formats, including DOCX. Beyond standard extraction functionalities, it leverages Natural Language Processing (NLP) techniques to intelligently extract relevant information from resumes. The script utilizes predefined lists and regular expressions for pattern matching, enhancing its ability to identify and extract specific details such as skills, education, and contact details. Users can employ the extracted data for further processing or analysis, making the tool versatile for purposes such as recruitment or career counseling.

# PROJEC T OVERVIEW

ResumeRevealer offers a versatile solution for parsing resumes in different document formats. Key features include:

* Document Format Support: Capable of handling a variety of resume formats, including DOCX,HTML,PDF.
* Information Extraction: Utilizes extraction functions to identify and extract critical details like education, phone numbers, emails, and skills.
* Job Prediction: Employs predictive algorithms to classify and visualize the distribution of job titles within the resumes.
* Sentiment Analysis: Incorporates sentiment analysis to review the tone expressed in the resumes.

# DEPENDENCIES

* Python 3.X
* Libraries used- Imported Libraries:
* docx2txt: Used to extract text from DOCX files.
* os: Standard Python library for interacting with the operating system.
* re: Library for regular expression matching operations in Python.
* nltk: Natural Language Toolkit, a library for natural language processing in Python.
* pandas: Data manipulation and analysis library providing data structures like DataFrame.
* scikit-learn: Machine learning library for data mining and analysis. Used for vectorization (TfidfVectorizer), model selection (train\_test\_split), and implementing machine learning models (LogisticRegression, SVC).
* matplotlib.pyplot: Plotting library for creating visualizations in Python.
* seaborn: Statistical data visualization library based on Matplotlib.
* pdfminer.high\_level: A tool for extracting information from PDF documents. The pdfminer.high\_level module is used for high-level PDF text extraction.

# CODE

The code is accessible in our GitHub repository, with individual scripts for each function. Please ensure you have downloaded all the specified libraries before running the code. Although some functions have not been fully integrated, they operate effectively on their own. Given additional time, we anticipate completing the project in the future to deliver a comprehensive and fully functional solution.

# USAGE

* Utilize the tool to process resumes in various document formats.
* Extract information such as education, phone numbers, emails, skills, and contact details.
* Predict job titles based on the content.
* Assess the sentiment expressed in the resumes.

# DOCUMENT FORMAT SUPPORT

* HTML: HTML documents are supported, enabling users to parse and extract information from resumes presented in web format.
* DOC and DOCX: Microsoft Word documents in both legacy DOC and modern DOCX formats are seamlessly handled. ResumeRevealer employs robust methods to extract relevant details from these widely used document types.
* PDF: Portable Document Format (PDF) is a common format for resumes. ResumeRevealer leverages the pdfminer library to efficiently extract text from PDF files, ensuring compatibility and accurate information retrieval.

# USP

ResumeRevealer actively addresses unethical practices, including the use of white text on a white background to manipulate ATS scanners. Leveraging the **pdfplumber** library, the tool examines font color and size attributes to detect instances of white text against a white background. Upon identification, ResumeRevealer takes action by highlighting or marking the concealed content, ensuring transparency and preventing fraudulent practices during the parsing process. This feature significantly contributes to ethical recruitment practices, enhancing the reliability of the parsed data and maintaining the integrity of the resume analysis.

ResumeRevealer not only safeguards against unethical practices like white text manipulation using advanced techniques but also provides users with a concise summary of key resume details, including skills, education, and work experience. This dual functionality ensures a transparent and efficient resume evaluation process, empowering recruiters with valuable insights while maintaining ethical standards.