

NLP Project

(RESME)

Ebtesam Sultan - Nuha Aljohani - Hanaa Alqarni

Abstract

Finding a job is considered one of the major activities for everyone.

Typically, large companies do not have enough time to open each CV, so we are help them by use machine learning algorithms for the Resume Screening task.

The purpose of this project is to use the Natural Language Processing to parse information from a resume PDF files, find the keywords, cluster them onto sectors based on their keywords, to helping human resources to fill jobs with the right person.

Data Description

- We extract our data set from Kaggle .com with 2485 rows and 4 columns.
- The Columns include
 - ID: Unique identifier and file name for the respective pdf.
 - Resume_str: Contains the resume text only in string format.
 - Resume_html: Contains the resume data in html format as present while web scrapping.
 - Category: Category of the job the resume was used to apply.

Algorithms:

1. Topic Modeling:

- Non-negative Matrix Factorization (NMF).
- Latent Dirichlet Allocation (LDA).

2. Clustering:

- Hierarchical clustering
- K-means

Tools

- **Technologies:** Python, Jupyter Notebook, PowerPoint
- **Libraries:** Pandas, Numpy, Re, Nltk, Spacy, Genism, Sklearn.Decomposition, Sklearn.Feature_Extraction.Text, Sklearn.Model_Selection, Pyldavis, Pyldavis.Sklearn, Matplotlib.Pyplot, Warnings.

Communication:

This wordcloud chart presented the most frequently used words in resume project.



This spider chart presents the best model of our project it is shows the relationship between the topic of resume that achieve the goal of the project.

