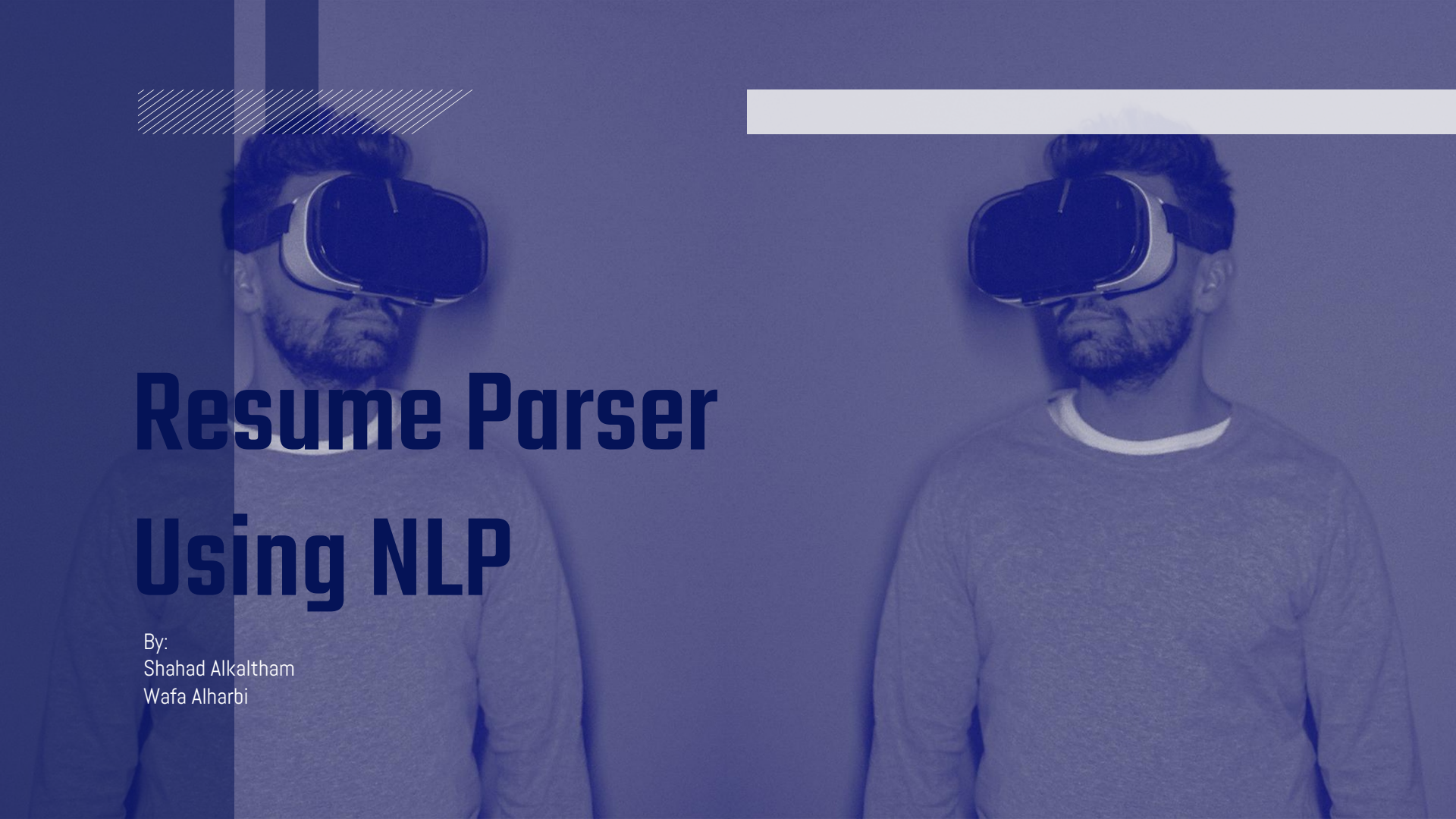


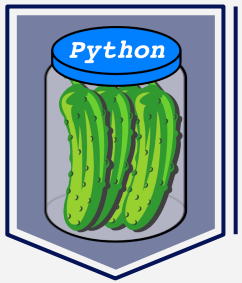
Resume Parser Using NLP

By:
Shahad Alkaltham
Wafa Alharbi

PROBLEM STATEMENT

Finding a job is considered one of the major activities for humans. The purpose of this project is to use the power of Natural Language Processing to parse information from a resume pdf files, find the keywords, cluster them onto sectors based on their keywords, in order to help job seekers to find their potential field according to their resumes contents.

TOOLS

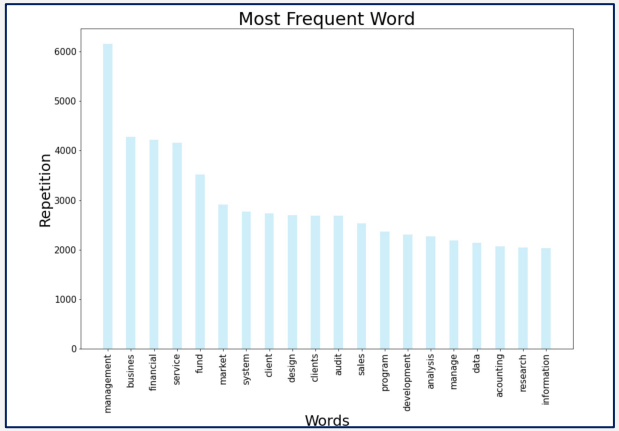


PDF Files

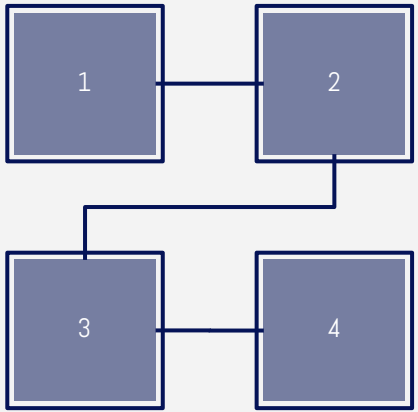
BEFORE 4,440
AFTER 1,534

Two Ready Dataset

BEFORE 1,388
AFTER 1,255



Drop nulls and duplicates



- Remove any digits and special characters
- lemmatize

Remove stop words (domain , English)

Remove Entities by SpaCy

PDF Files

- ✓ NMF + TF-IDF → 3 to 11 Topic
- ✓ Remove Domain Specific Word → **334** words
- ✓ NMF + TF-IDF → 10 to 20 Topic
- ✓ Remove Domain Specific Word → **584** words
- ✓ NMF + TF-IDF → 10 to 20 Topic
- ✓ Remove Domain Specific Word → **595** words
- ✓ NMF + TF-IDF → 7 to 15 Topic
- ✓ NMF + TF-IDF → 11 TOPIC

The Final Topics

- Finance
- Hospitality
- Electrical and Mechanical Engineering
- Sales and Marketing
- Beauty Artist
- Secretarial
- Accounting
- Health
- Accounting
- Investment
- Others

Two Ready Dataset

- ✓ NMF + TF-IDF → 3 to 11 Topic
- ✓ Remove Domain Specific Word → **334** words
- ✓ NMF + TF-IDF → 10 to 20 Topic
- ✓ Remove Domain Specific Word → **584** words
- ✓ NMF + TF-IDF → 10 to 20 Topic
- ✓ NMF + TF-IDF → 13 TOPIC

The Final Topics

- Business Administration
- Sales and Marketing
- Accounting
- Management Information System
- Risk Management
- Law
- Hospitality
- Electrical and Mechanical Engineering
- Health
- Graphics Design
- Finance
- Teaching
- Development

Model Selection

Model	Training Accuracy	Validation Accuracy	F1	Recall	Precision
Logistic Regression	0.6499	0.6774	0.6774	0.6774	0.6774
Decision Tree	0.9085	0.8853	0.8853	0.8853	0.8853
Random Forest	1.0000	0.8960	0.8961	0.8961	0.8961
Extra Trees	1.0000	0.8960	0.8961	0.8961	0.8961
XGB	0.9439	0.8996	0.8996	0.8996	0.8996
Gaussian	0.4114	0.4265	0.4265	0.4265	0.4265
Bernoulli	0.3366	0.2652	0.2652	0.2652	0.2652

Reporting The Final Model

Chosen Model Decision Tree
Training Score 0.9127
Validation Score 0.8351
F1 0.8351
Recall 0.8351
Precision 0.8351

FUTURE WORK

1
Improve our model to be more accurate

2
Implement an auto resume parser

3
Store the passed pdf files to use it as training set in future

4
Integrate our project with a regression model

THANKS!