

Resume

Maha Alharqan
Elaf Talal
Maha Aladwani

Junior Data Scientists
AT SDAIA BOOTCAMP T5



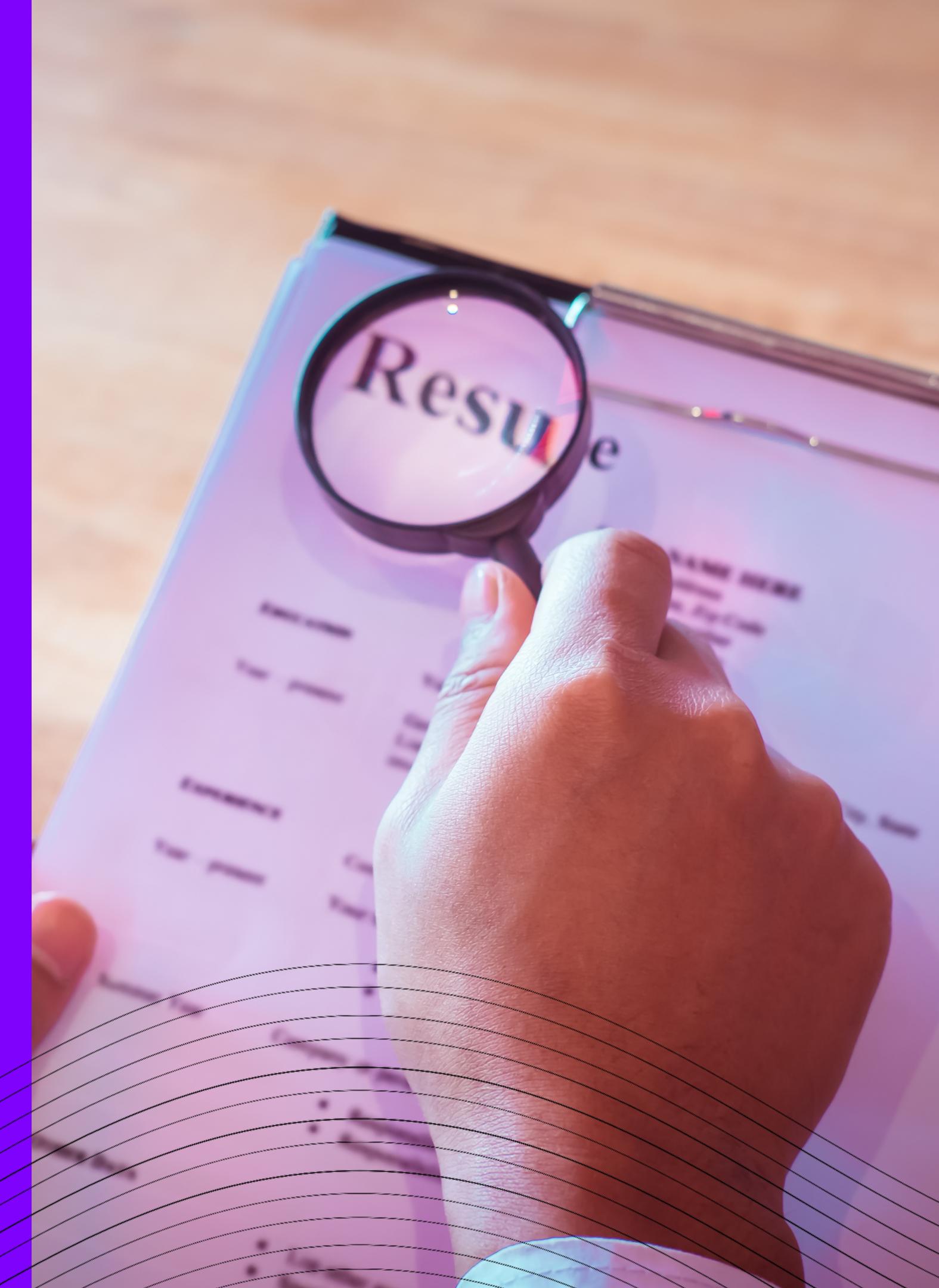
Agenda

- Introduction
- Project Goal
- Data structure
- NLP Preprocessing
- Modeling
- Conclusions
- Tools

Introduction

Companies often receive thousands of resumes for each job posting and employ dedicated screening officers to screen qualified candidates.

Hiring the right talent is a challenge for all businesses. This challenge is magnified by the high volume of applicants if the business is labor-intensive, growing, and facing high attrition rates.





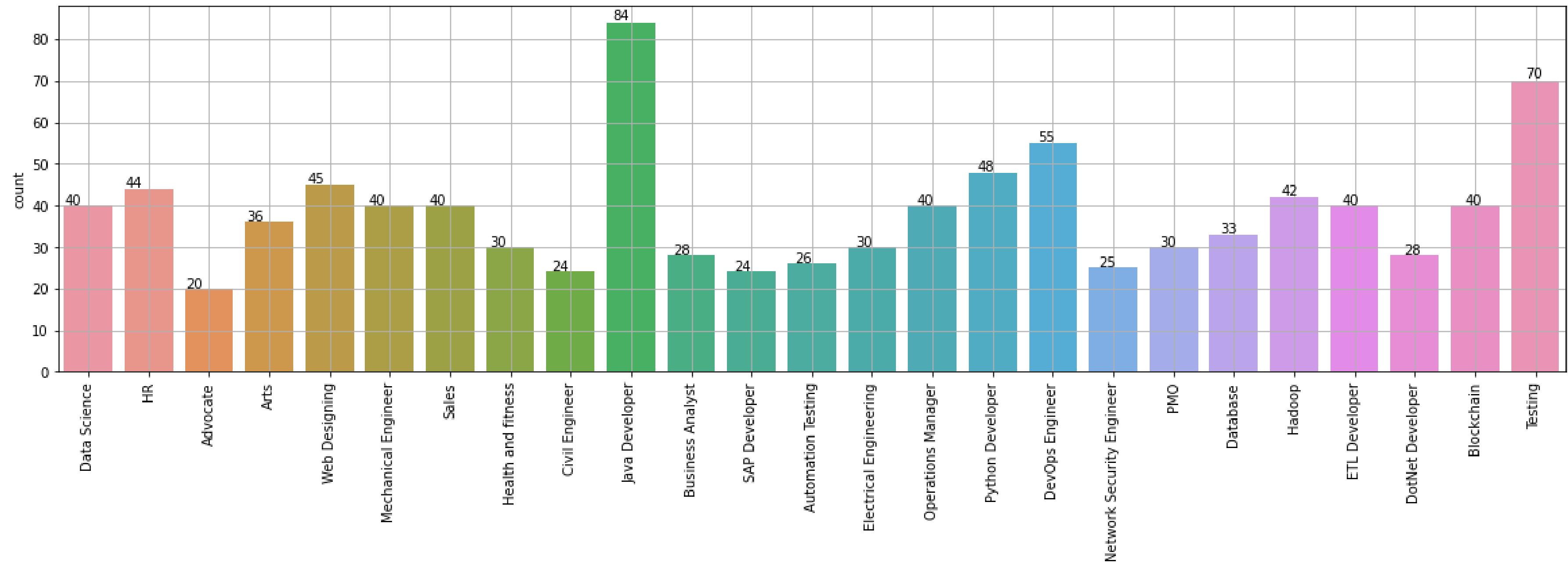
Project Goal

Usually, large companies do not have enough time to open every resume, so we will use machine learning algorithms for the job of checking resumes.

Data structure

- Data Source: Kaggle
- Data set contains :
 - 962 rows
 - 2 columns

Data structure

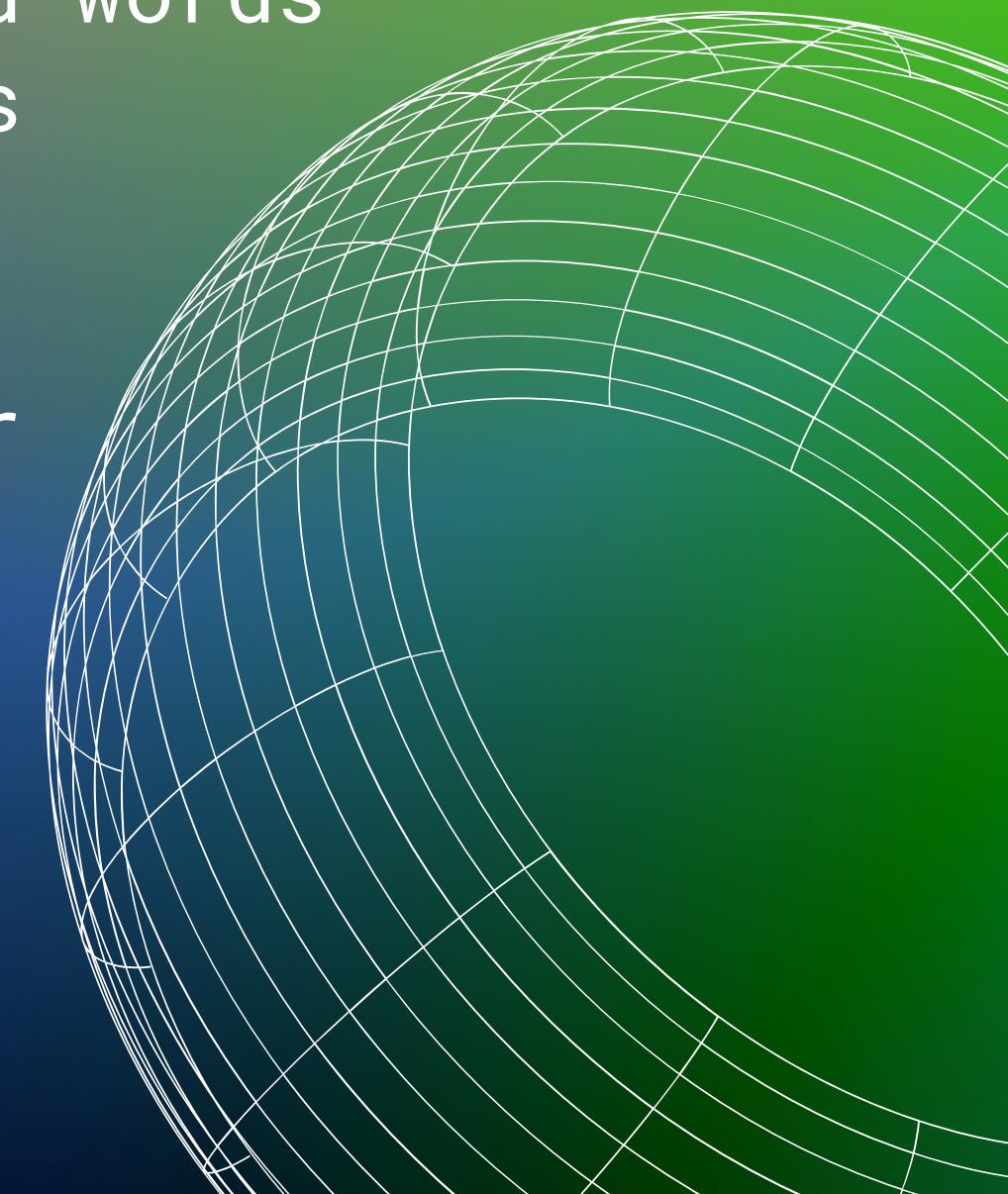




NLP Preprocessing

Clean data

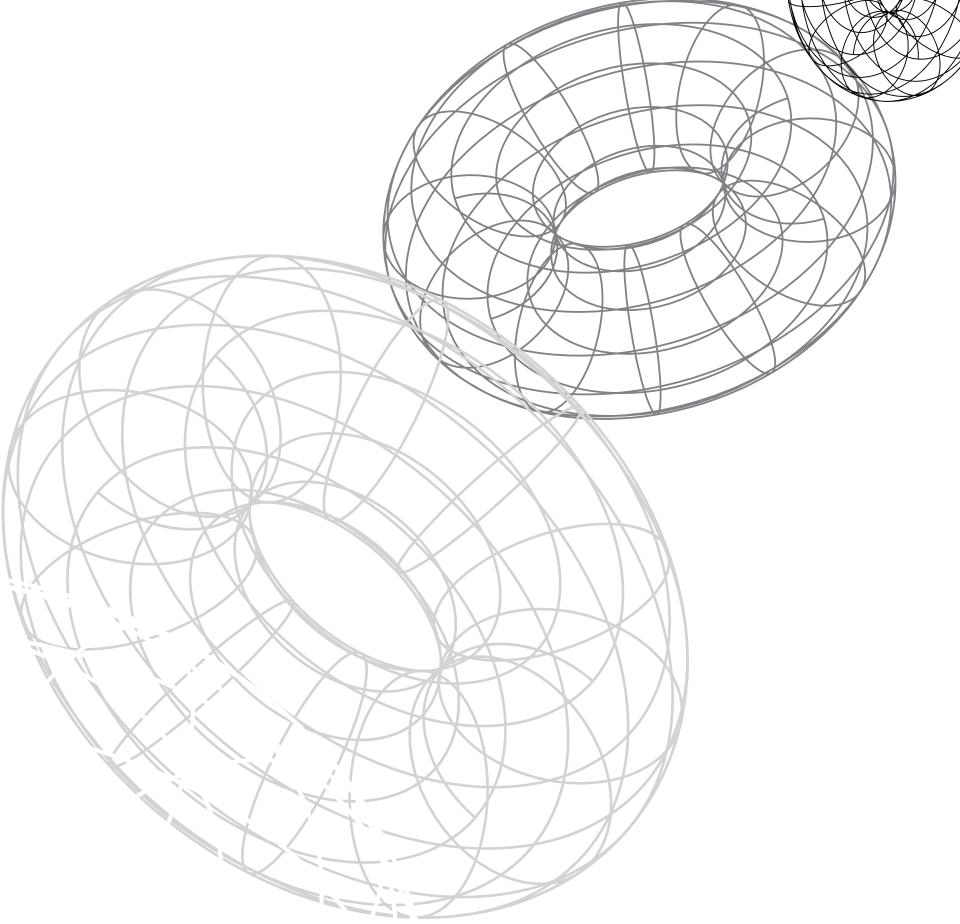
- remove some characters
- remove punctuation
- remove frequented words
- remove stop words
- Stemming
- lemmatization
- TF-IDF Vectorizer
- Topic modeling :
LDA
LSA
NMF



Modeling

- Unsupervised
Topic modeling
 - LDA
 - LSA
 - NMF

- Supervised
 - KNNNeighbors Classifier



Modeling

● LDA lemmatization

```
Top 4 words for topic #0  
['test', 'fitness', 'sales', 'java']
```

```
Top 4 words for topic #1  
['python', 'pune', 'test', 'testing']
```

```
Top 4 words for topic #2  
['net', 'hive', 'developer', 'hadoop']
```

```
Top 4 words for topic #3  
['mechanical', 'design', 'engineering', 'java']
```

● LDA Stemming

```
Top 4 words for topic #0  
['use', 'databas', 'test', 'develop']
```

```
Top 4 words for topic #1  
['client', 'oper', 'manag', 'report']
```

```
Top 4 words for topic #2  
['technolog', 'java', 'test', 'develop']
```

```
Top 4 words for topic #3  
['sale', 'engin', 'develop', 'work']
```

Modeling

● LSA lemmatization

Top 4 words for topic #0

```
['pune', 'developer', 'testing', 'java']
```

Top 4 words for topic #1

```
['jquery', 'ajax', 'developer', 'java']
```

Top 4 words for topic #2

```
['servers', 'shell', 'hive', 'hadoop']
```

Top 4 words for topic #3

```
['roles', 'photoshop', 'application', 'ui']
```

● LSA Stemming

Top 4 words for topic #0

```
['use', 'java', 'develop', 'test']
```

Top 4 words for topic #1

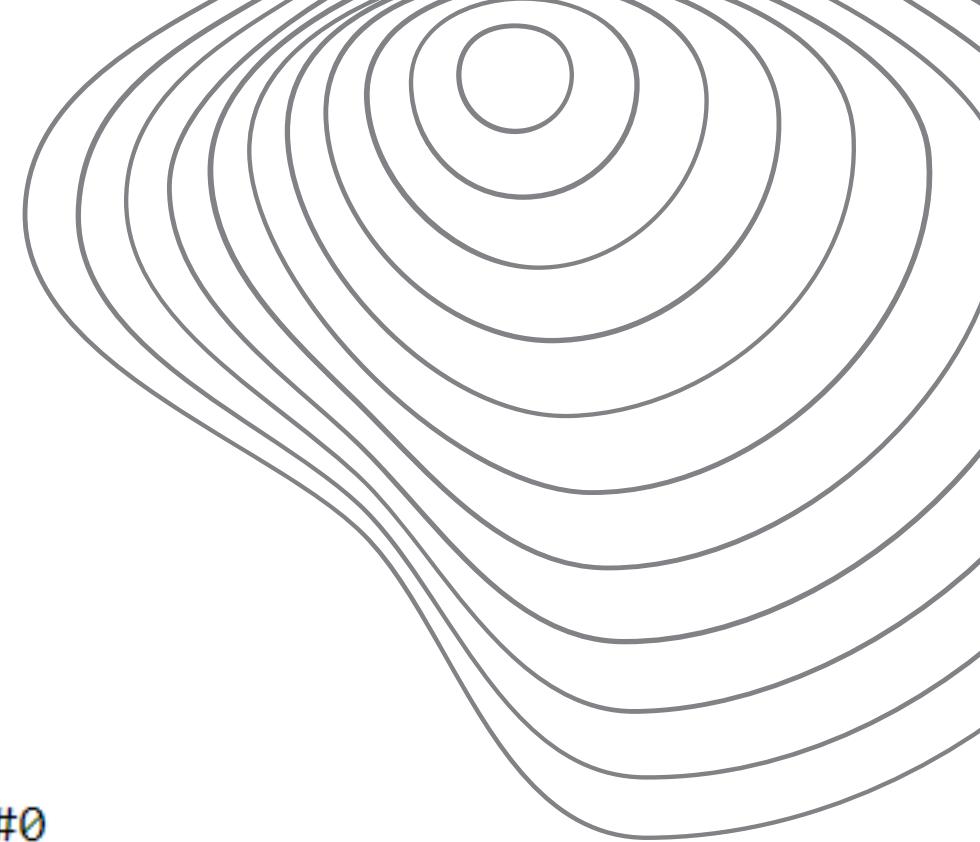
```
['javascript', 'jqueri', 'ajax', 'java']
```

Top 4 words for topic #2

```
['script', 'server', 'hive', 'hadoop']
```

Top 4 words for topic #3

```
['window', 'transform', 'autom', 'test']
```



Modeling

● NMF lemmatization

```
Top 4 words for topic #0  
['test', 'business', 'operations', 'sales']
```

```
Top 4 words for topic #1  
['python', 'ajax', 'developer', 'java']
```

```
Top 4 words for topic #2  
['database', 'hive', 'shell', 'hadoop']
```

```
Top 4 words for topic #3  
['responsibility', 'photoshop', 'ui', 'application']
```

● NMF Stemming

```
Top 4 words for topic #0  
['busi', 'report', 'manag', 'sale']
```

```
Top 4 words for topic #1  
['jqueri', 'python', 'develop', 'java']
```

```
Top 4 words for topic #2  
['databas', 'script', 'server', 'hadoop']
```

```
Top 4 words for topic #3  
['autom', 'transform', 'engin', 'test']
```

Modeling

Word Clouds for NMF model with stemming

Topic one

Mangement



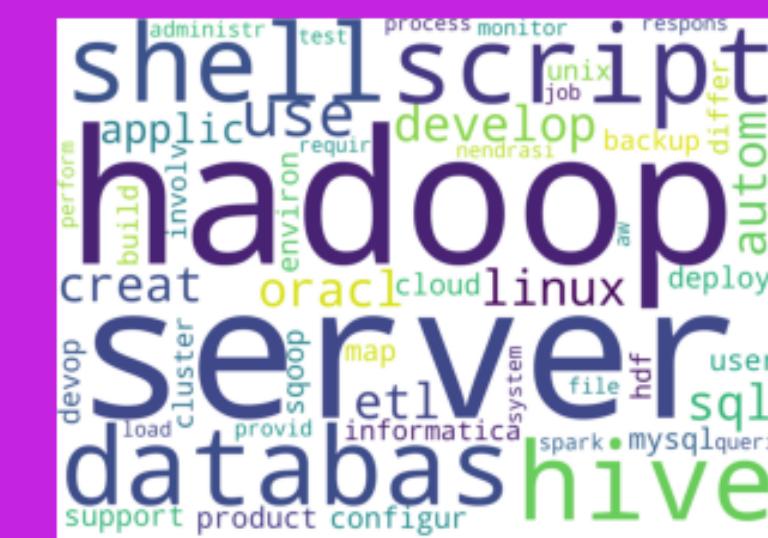
Topic two

Development



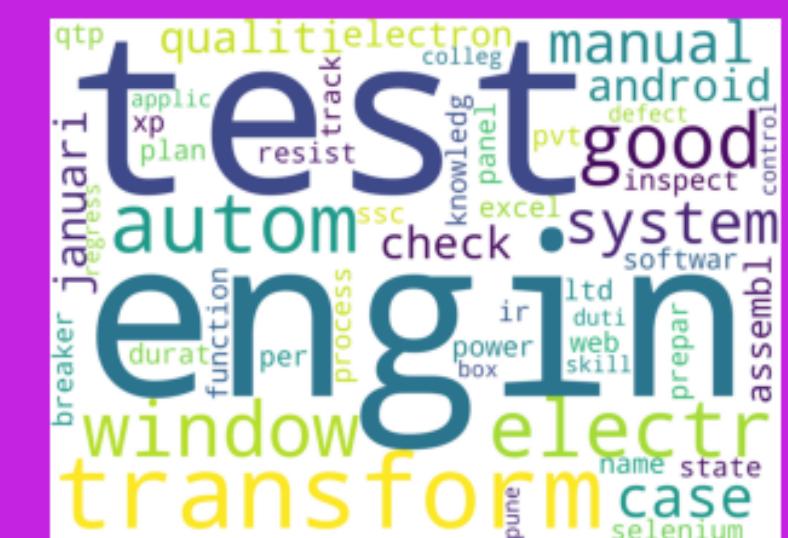
Topic Three

Server Administration



Topic Four

Engineering



Modeling

Our best model: NMF with stemming

Resume

NMF

skill program languag python panda numpy scipi...

Development

educ may may b e uit rgpv scientist scientist ...

Server Administration

area interest deep learn control system design...

Development

skill r python sap hana tableau sap hana sql s...

Server Administration

educ mca ymcaust faridabad haryana scienc inte...

Development

KNeighbors Classifier

- The Accuracy of KNeighbors Classifier on training set: 0.99
- The Accuracy of KNeighbors Classifier on test set: 0.98

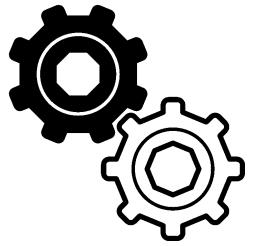
Conclusions

- Filtering Resumes easily.
- Automating the process of resumes selection.
- Ability to handle a larger number of resumes in a short time.
- The NMF model dealt with the topics better than the other models.

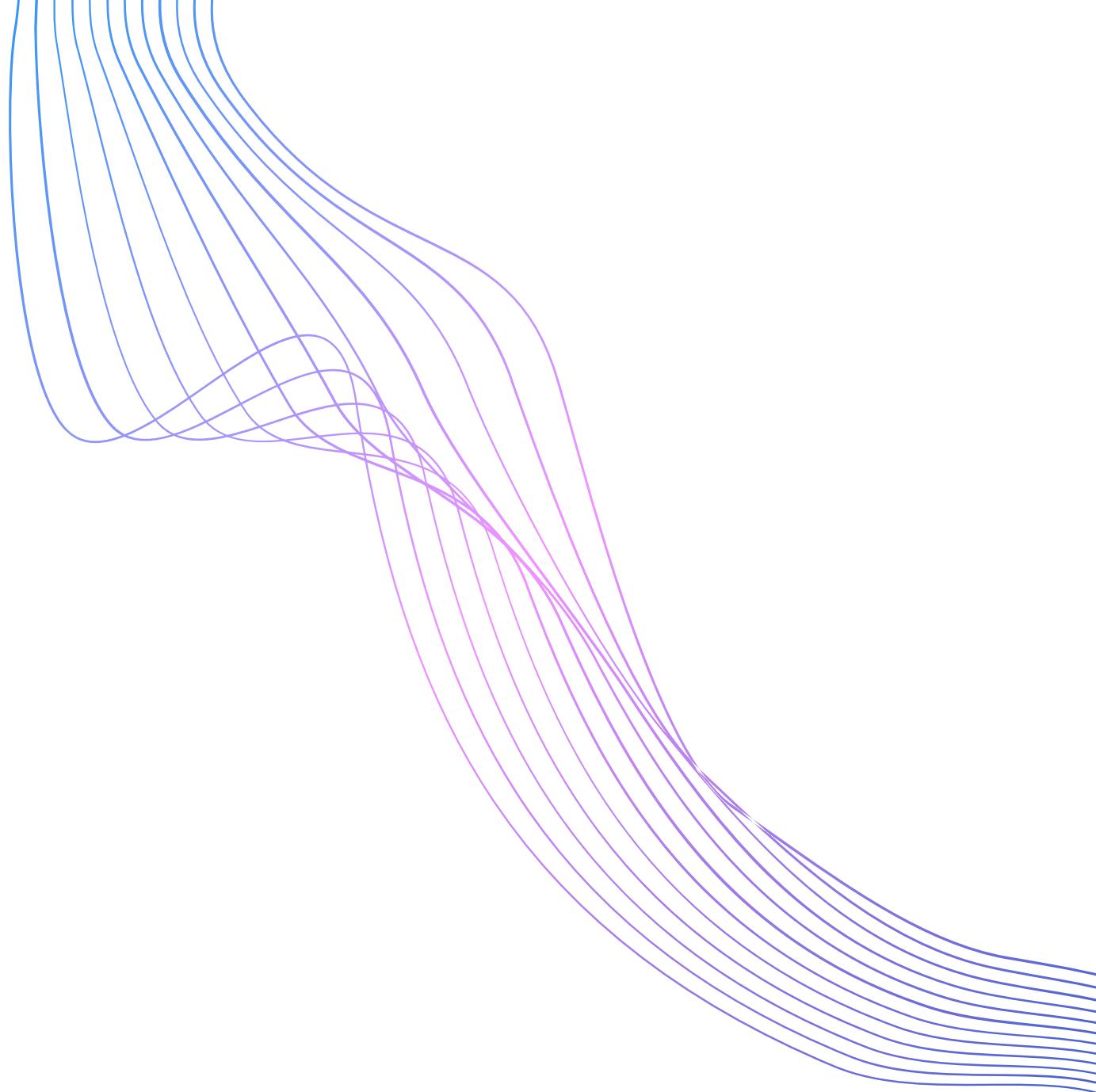
Future Work

- Merge another dataset that has a job description and another detail.
- Doing a recommendation system when the job description is available.
- Deploy a recommendation system.

Tools



- Python
- Google Colab
- pandas
- NumPy
- Sklearn
- Seaborn
- Nltk
- re
- wordcloud
- warnings





Thank you