

Human Resources Automatic CV Classification

Master of Science in Artificial Intelligence and Data Engineering



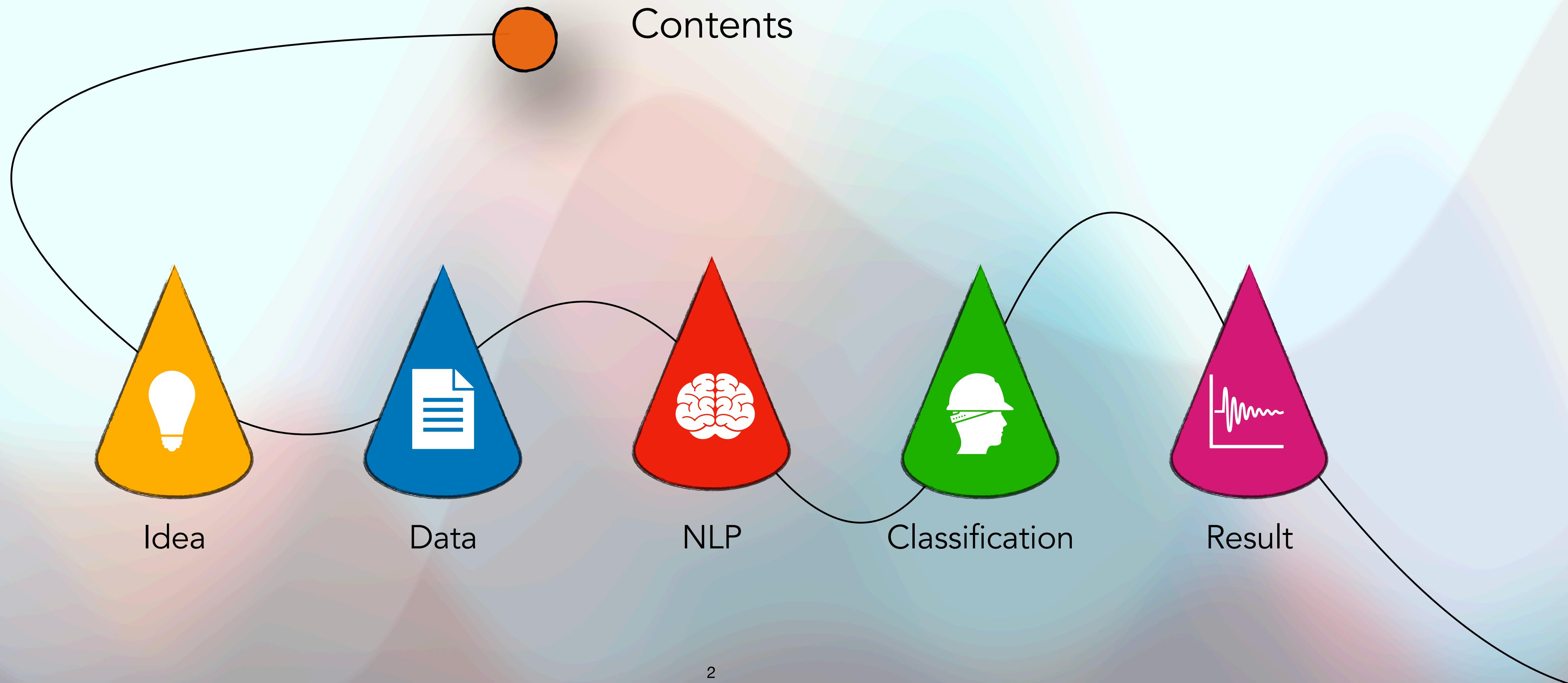
Work group:

Massimo Valentino Caroti

Simone Landi

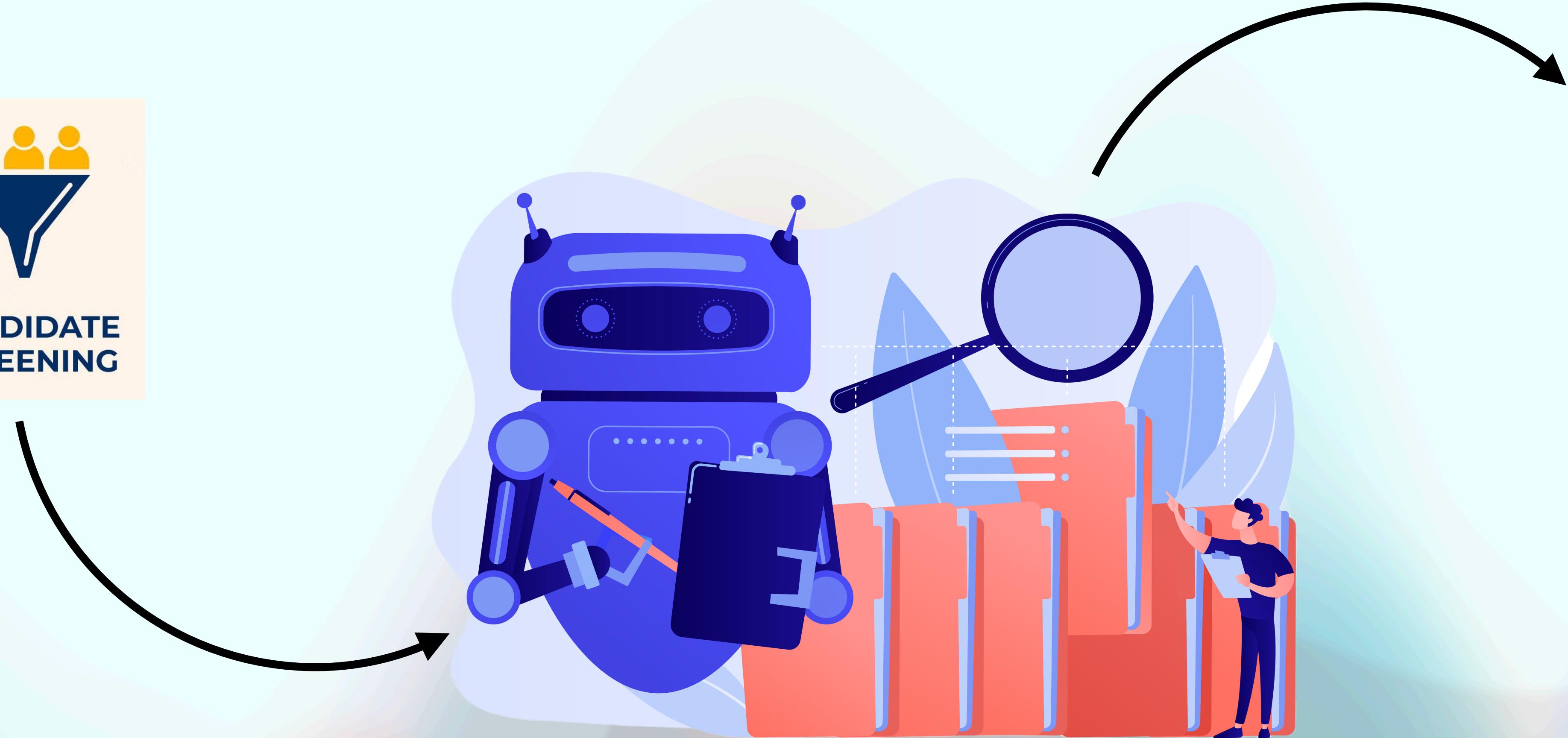
Jacopo Cecchetti

Human Resources Automatic CV Classification



HIRING PROCESS





Resume Classification System using the Natural Language Processing (NLP) and Machine Learning (ML) techniques could automate this tedious process. This project presents an automated system based on NLP and ML that classifies CVs according to job categories with performance guarantees.

Mathew Smith

UI DESIGNER

PROFILE

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Incidunt.

EDUCATION

ENTER YOUR MAJOR
Name Of Your University
2005-2009

ENTER YOUR MAJOR
Name Of Your University
2009-2011

EXPERTISE

- Photoshop
- Illustrator
- InDesign
- Word
- Power point

REFERENCE

Glenin M. Gregory Director Marix Media Ltd. Phone : +555 123 454	Jennifer S. Gavin Director Marix Media limited Phone : +555 123 454
--	---

INTERESTS

- Travel
- Music
- Writing
- Chess

Phone : +000 123 456 789
Email : username@gmail.com
Area : Your Street Address Here



For the project, we used a dataset in which there are more than 4000 complete resumes, for each CV there are: personal information, job for which he wants to apply, skills, description of past jobs and passions.

```
['hr' 'designer' 'information-technology' 'teacher' 'advocate'
 'business-development' 'health and fitness' 'agriculture' 'bpo' 'sales'
 'consultant' 'digital-media' 'automobile' 'chef' 'finance' 'apparel'
 'accountant' 'construction' 'public-relations' 'banking' 'arts'
 'aviation' 'data science' 'web designing' 'mechanical engineer'
 'civil engineer' 'java developer' 'business analyst' 'sap developer'
 'automation testing' 'electrical engineering' 'operations manager'
 'python developer' 'devops engineer' 'network security engineer' 'pmo'
 'database' 'hadoop' 'etl developer' 'dotnet developer' 'blockchain'
 'testing']
```



We used a word cloud associated with all resumes to visualize the most frequent words in the resumes. We also generated a word cloud for each job with the most frequent and important keywords to identify that job.

```
[ 'hr' 'designer' 'information-technology' 'teacher' 'advocate'  
'business-development' 'health and fitness' 'agriculture' 'bpo' 'sales'  
'consultant' 'digital-media' 'automobile' 'chef' 'finance' 'apparel'  
'accountant' 'construction' 'public-relations' 'banking' 'arts'  
'aviation' 'data science' 'web designing' 'mechanical engineer'  
'civil engineer' 'java developer' 'business analyst' 'sap developer'  
'automation testing' 'electrical engineering' 'operations manager'  
'python developer' 'devops engineer' 'network security engineer' 'pmo'  
'database' 'hadoop' 'etl developer' 'dotnet developer' 'blockchain'  
'testing' ]
```

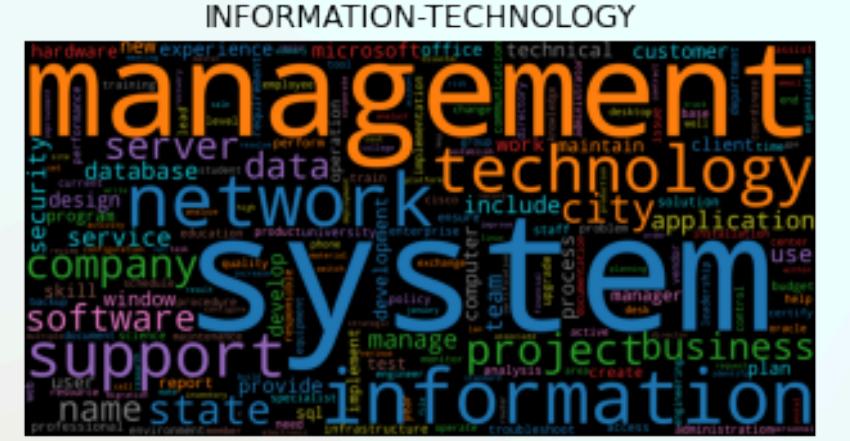




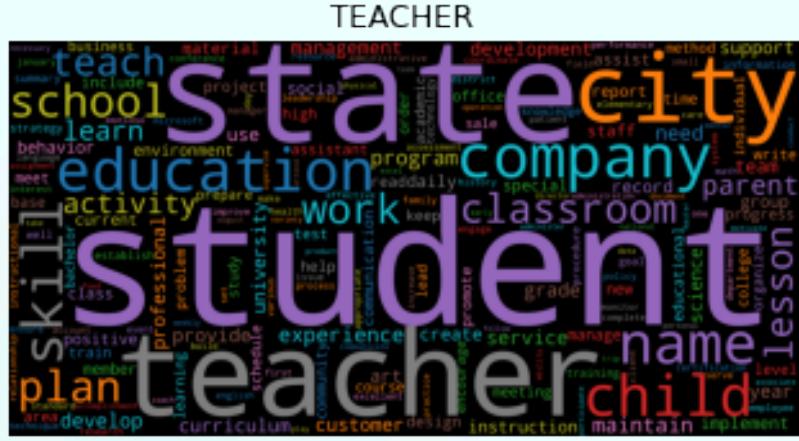
HR



DESIGNER



INFORMATION-TECHNOLOGY



TEACHER



ADVOCATE



BUSINESS-DEVELOPMENT



HEALTHCARE



FITNESS



AGRICULTURE



BPO



SALES



CONSULTANT



DIGITAL-MEDIA



AUTOMOBILE



CHEF



NLP Preprocessing

**“Roles and Responsibility:
Designed and developed the
UI of Dashboard.”**



NLP Preprocessing

● Data cleaning

**“roles and responsibility
designed and developed the
ui of dashboard”**



NLP Preprocessing

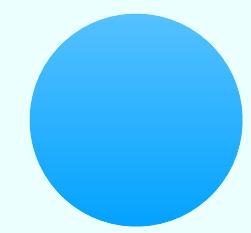
Tokenization

Data cleaning

‘roles’
‘and’
‘responsibility’
‘designed’
‘and’
‘developed’
‘the’
‘ui’
‘of’
‘dashboard’



NLP Preprocessing



Filtering

Tokenization

Data cleaning

‘roles’

‘responsibility’

‘designed’

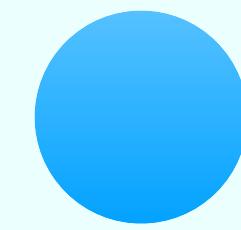
‘developed’

‘ui’

‘dashboard’



NLP Preprocessing



Part-of-speech tagging

Filtering

Tokenization

Data cleaning

('roles', 'n')

('responsibility', 'n')

('designed', 'v')

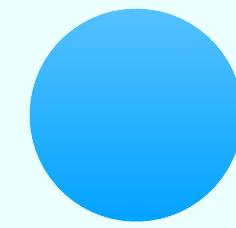
('developed', 'v')

('ui', 'a')

('dashboard', 'n')



NLP Preprocessing



Lemmatization

'role'

Part-speech-tagging

'responsibility'

Filtering

'design'

Tokenization

'develop'

Data cleaning

'ui'

'dashboard'



Text Vectorization

Term Frequency - Inverse Document Frequency (TFIDF)

dashboard	design	develop	homepage	responsibility	role	ui
0.368954	0.316166	0.43701	0.000000	0.437010	0.437010	0.43701
0.000000	0.483802	0.00000	0.564579	0.000000	0.000000	0.66872
0.000000	0.000000	0.00000	0.000000	0.707107	0.707107	0.00000
0.759339	0.650696	0.00000	0.000000	0.000000	0.000000	0.00000
0.000000	0.483802	0.66872	0.564579	0.000000	0.000000	0.00000
0.707107	0.000000	0.00000	0.707107	0.000000	0.000000	0.00000

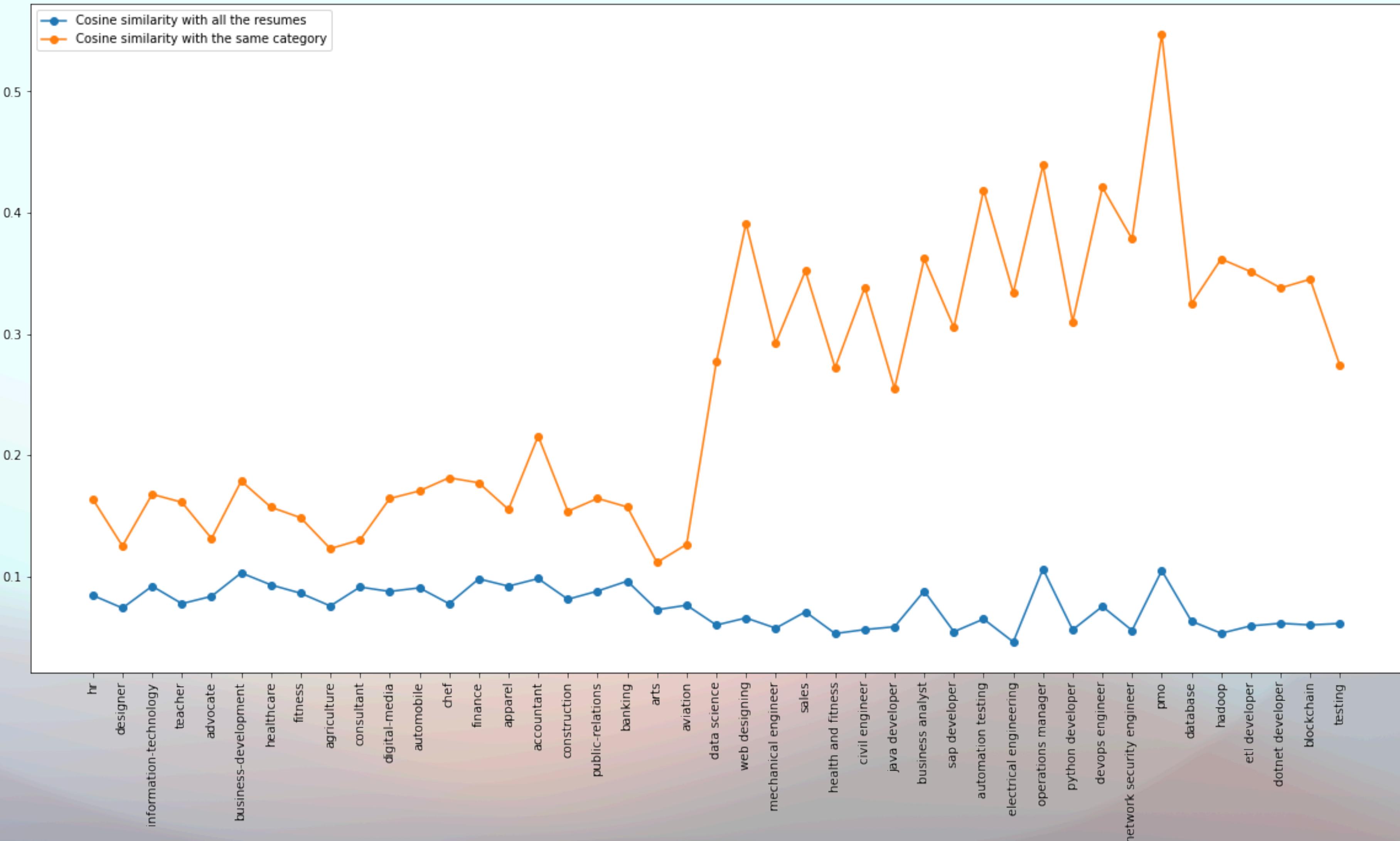
$$tf_{i,j} = \frac{\text{occurrences of } w \text{ in document } d}{\text{total number of words in document } d}$$

$$idf_i = \ln \frac{\text{total number of documents}}{\text{number of documents containing } w}$$

$$(tf-idf)_{i,j} = tf_{i,j} \times idf_i$$

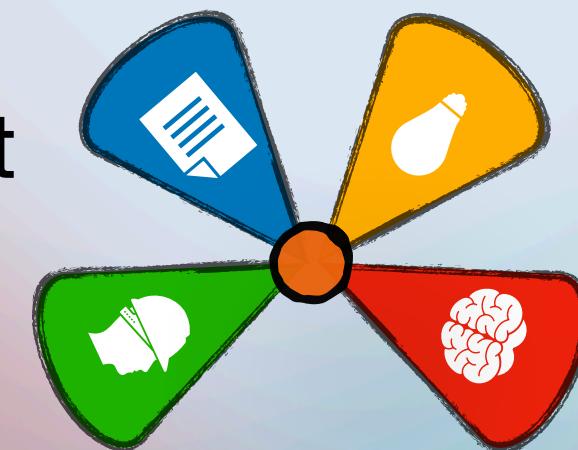
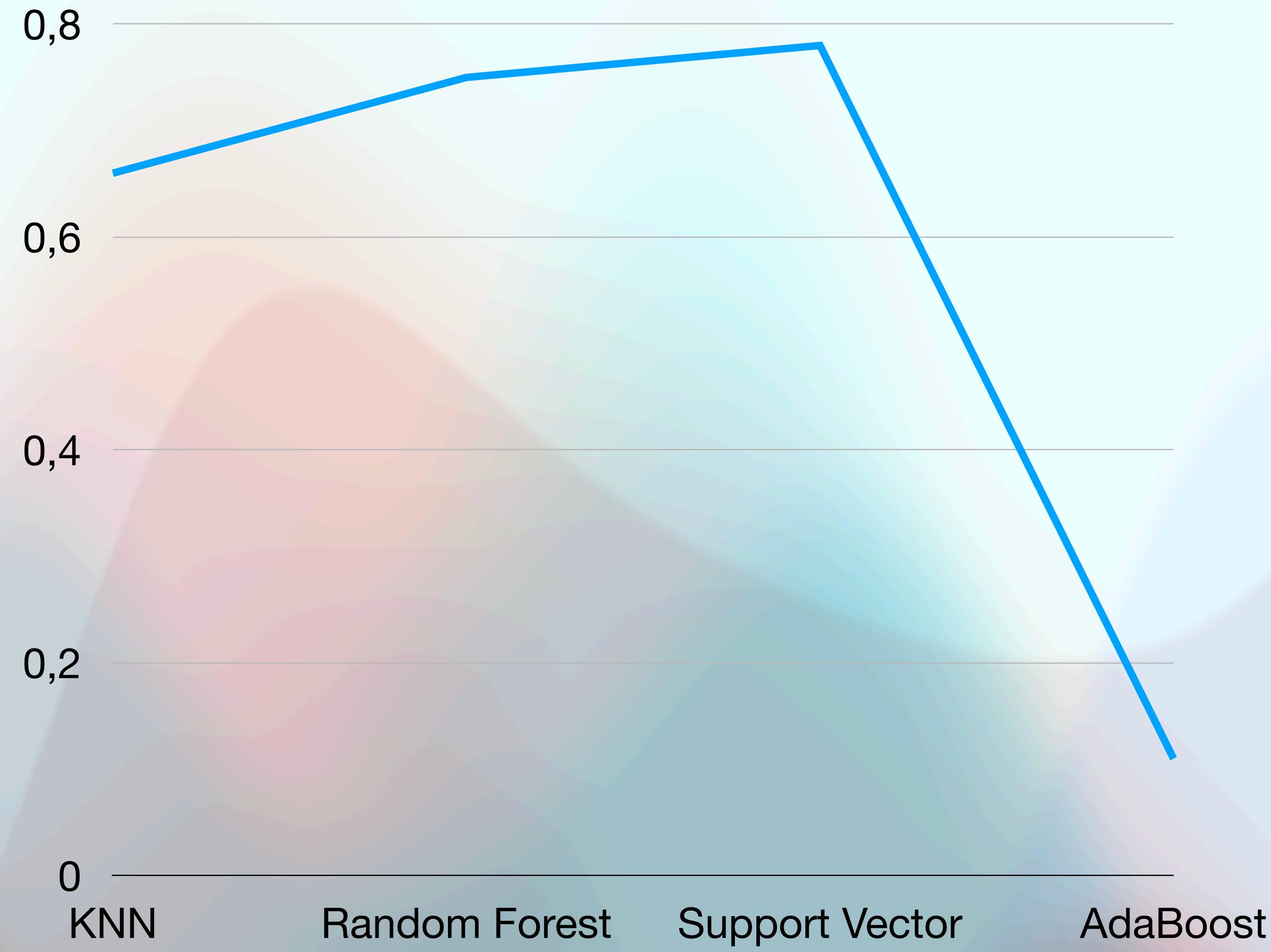


Document similarity



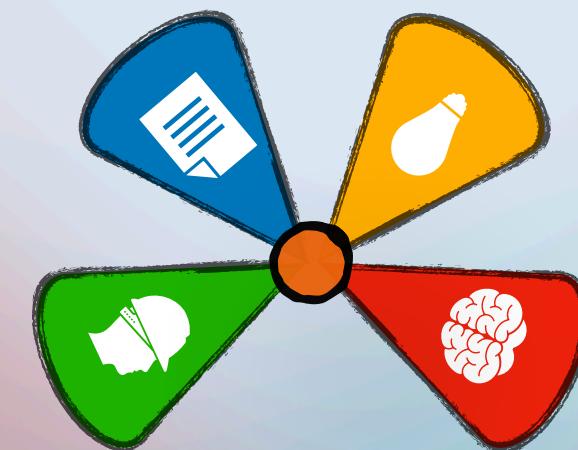
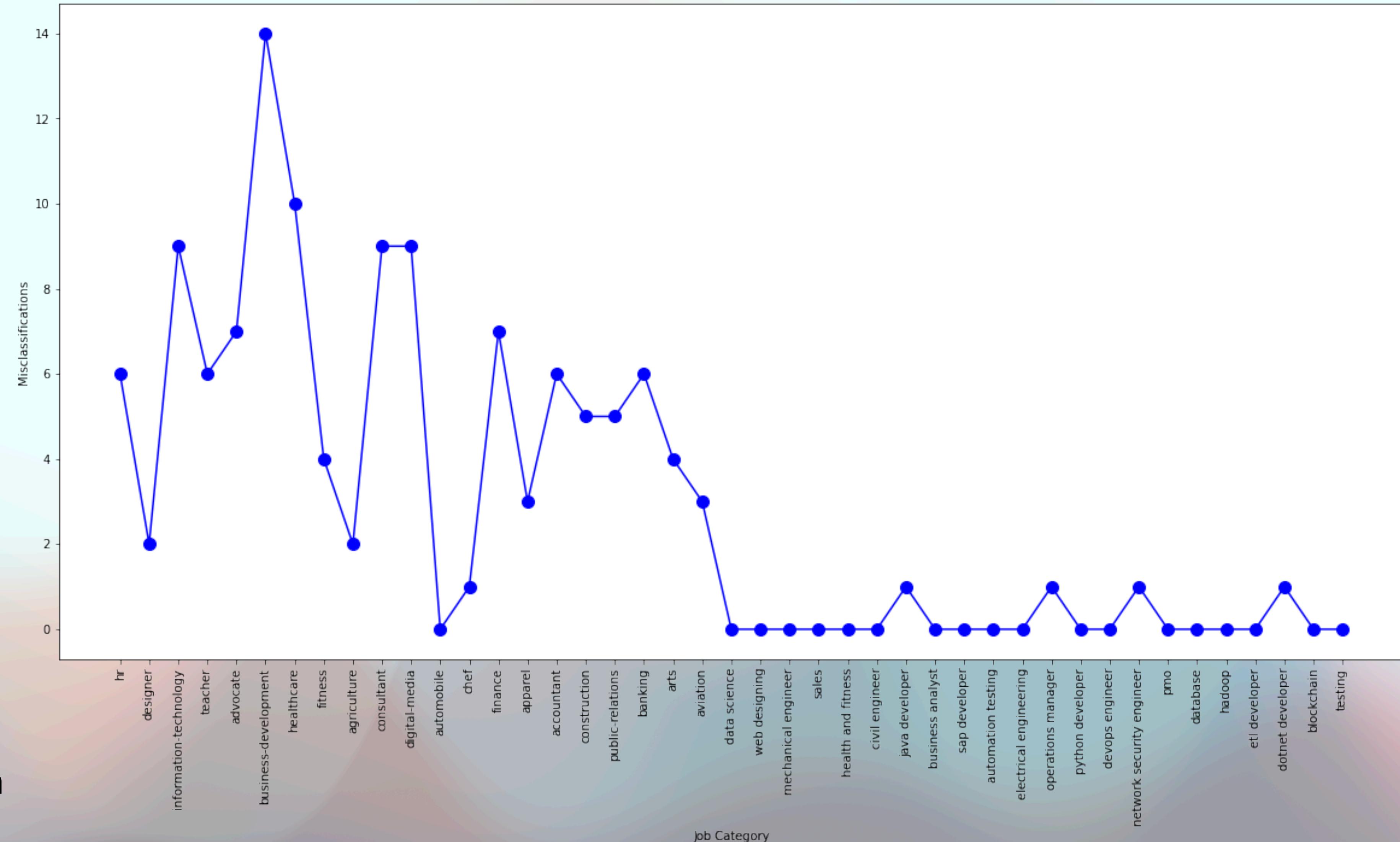
Classification Accuracy

Classification



Misclassified resumes

Classification



Results

Based on the curriculum and category, the model was designed to classify the curriculum into the right category



Results

Classification

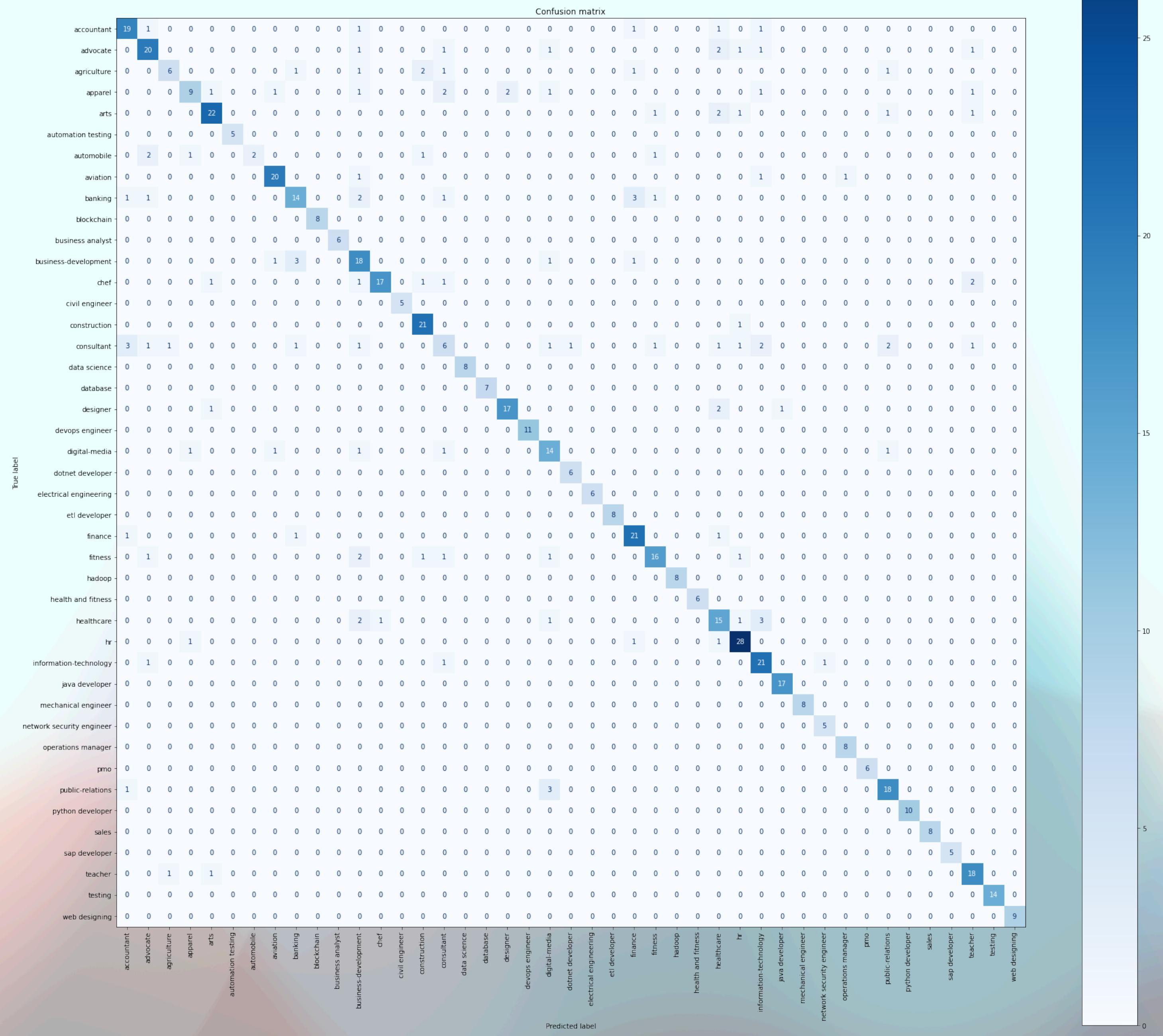
Linear Support Vector model: a supervised learning models with associated learning algorithms that analyze data for classification and regression analysis

accuracy of 0.78 on 10-fold cross-validation



Confusion Matrix

Result



Potential implications

- speed up the selection of profiles
- reduce the resources used in identifying the right talent
- the model can be further improved to target specific industries

Result



Limitations

- accepts CVs in CSV format
- loss of information due to text compression

Classification

Results

Result





Classification



Results



Result

Solutions

- libraries that can read different file formats
- having more data for training the model



Conclusion

- The proposed model classifies the curriculum into different categories automatically
- 0.78 accuracy with the Linear Support Vector classifier
- Involving domain experts such as HR professionals would help to build a more accurate model

