Resume Screening With NLP

BY ASMA AL-SULAMI

TABLE OF CONTENTS

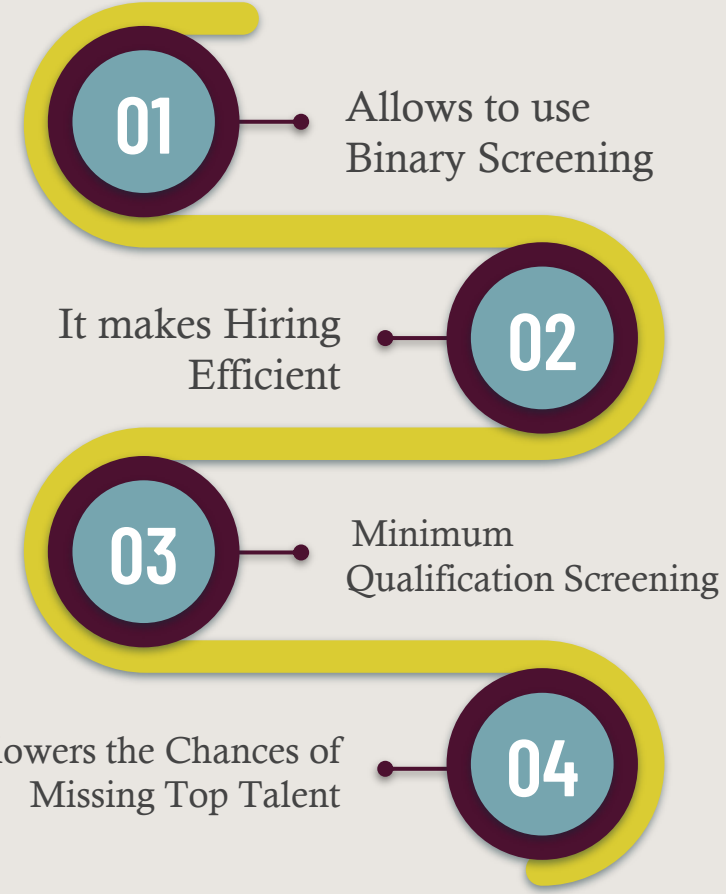
What is resume screening?

Resume screening is **the process of determining whether a candidate is qualified for a role based his or her education, experience, and other information captured on their resume**.

Why do we need Resume Screening?

Choosing the right people for the job is the biggest responsibility of every business since choosing the right set of people can accelerate business growth exponentially.Due to many big projects with big companies, their team does not have time to read resumes and choose the best resume according to their requirements.

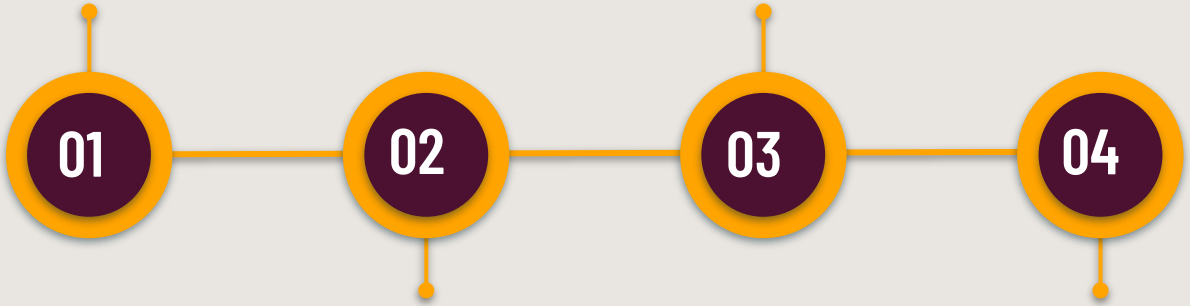
TARGET


01
Allows to use Binary Screening
02
It makes Hiring Efficient
03
Minimum Qualification Screening
04
It lowers the Chances of Missing Top Talent

PROJECT PERFORMANCE

Data Exploration

Model Building



Data Preprocessing

Topic Modeling

DATA SOURCE

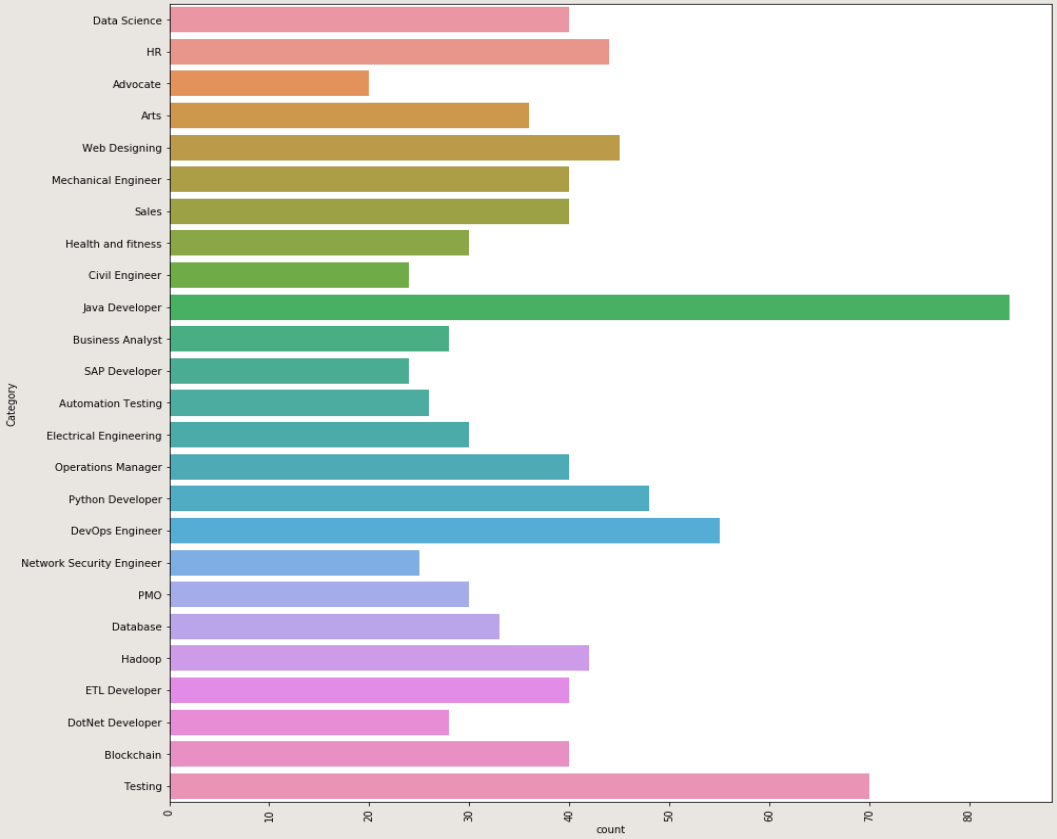
We have publically available data from Kaggle. https://www.kaggle.com/gauravdutakiit/resume-dataset

.

TOOLS

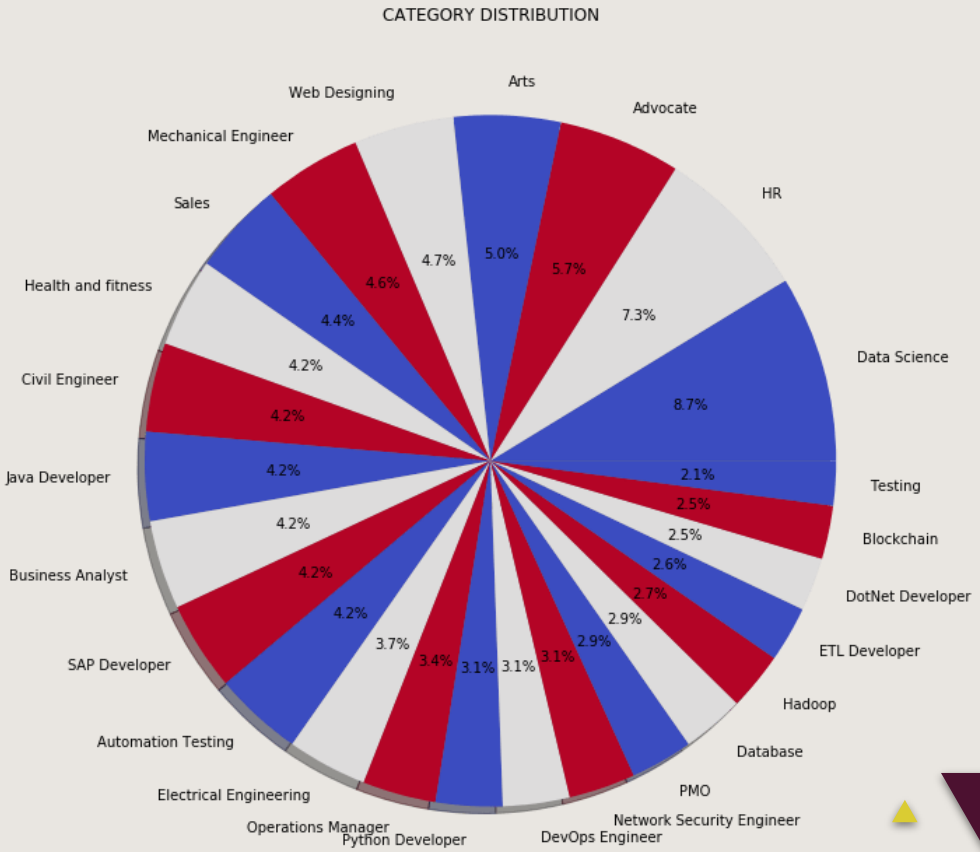
Warnings
naive_bayes
OneVsRestClassifier
CountVectorizer
TruncatedSVD, LSA
cosine_similarity
GridSpec
nltk
wordcloud
scipy

RESULTS



There are 25 different categories we have in the data. The top 3 job categories we have in the data are as follows.
Java developer, Testing, and DevOps Engineer.

RESULTS


CATEGORY DISTRIBUTION
Arts
Advocate
HR
Data Science
Testing
Blockchain
DotNet Developer
ETL Developer
Hadoop
Database
PMO
Network Security Engineer
DevOps Engineer
Python Developer
Operations Manager
Electrical Engineering
Automation Testing
SAP Developer
Business Analyst
Java Developer
Civil Engineer
Health and fitness
Sales
Mechanical Engineer
Web Designing
5.0%
5.7%
7.3%
8.7%
2.1%
2.5%
2.5%
2.6%
2.7%
2.9%
2.9%
3.1%
3.1%
3.4%
3.7%
4.2%
4.2%
4.2%
4.2%
4.2%
4.2%
4.4%
4.6%
4.7%

RESULTS

```
Accuracy of KNeighbors Classifier on training set: 0.99
```

```
Accuracy of KNeighbors Classifier on set set: 0.99
```

We can see that results are awesome. We are able to classify each Category of a given resume with 99% accuracy.

RESULTS

Topic Modeling Latent Semantic Analysis (LSA)

```
Topic  0
project, company, exprience, description, team, management, months, details, data, client

Topic  1
client, project, team, honeywell, fat, ges, activity, monitored, quality, korea

Topic  2
company, operations, management, services, team, international, india, inventory, customer, handling

Topic  3
test, testing, automation, business, qtp, involved, cases, integration, responsibilities, requirement

Topic  4
data, management, hadoop, project, hive, services, sap, billing, sla, report

Topic  5
data, company, hadoop, using, hive, ensuring, inventory, freight, shipments, etl
```

Conclusion

we learned how machine learning and Natural Language Processing can be applied to improve our day-to-day life through the example of Resume Screening. We just classified almost 1000 resumes in few minutes into their respective categories with 99% accuracy

THANKS!