

Fake Job Detection System

Data Analytics Project

By-

Ananya Ghosh

20MIC0063

BDS3002 - C2



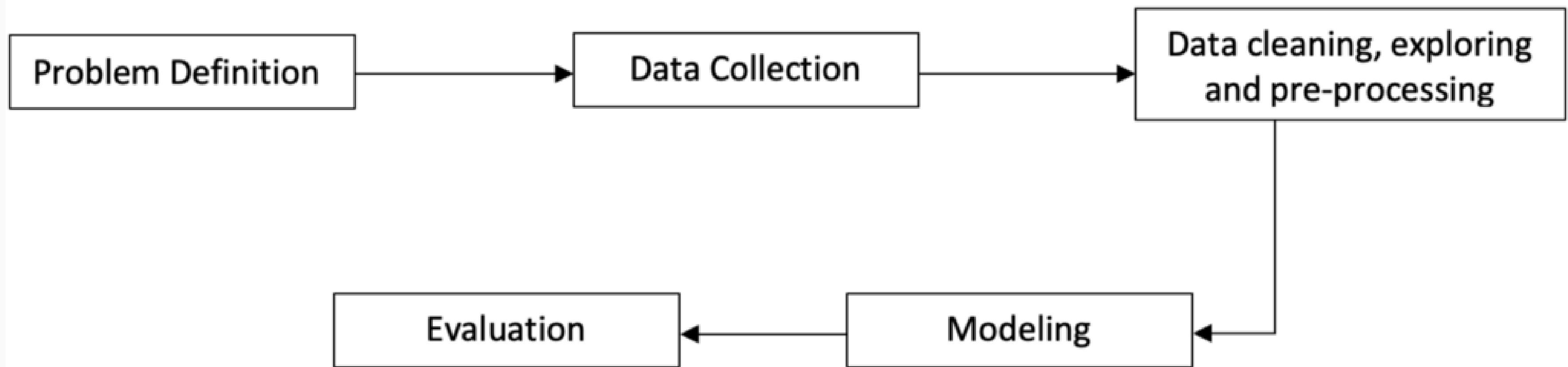


ABSTRACT

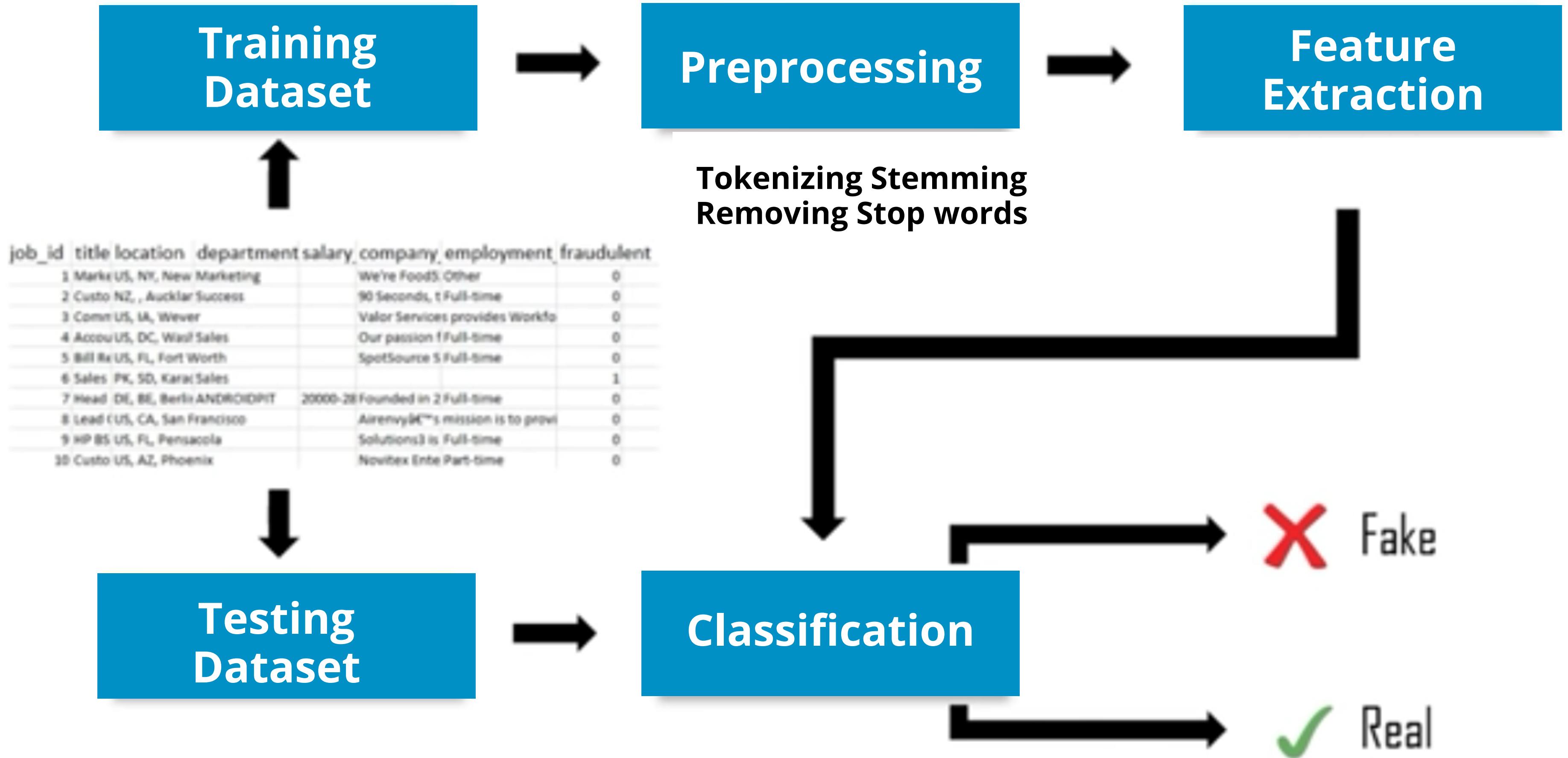


The Project aims to solve the problem of **rising instances of fraudulent job offers** make it a poignant time to learn how you could recognize and avoid becoming victims of fallacious job offers.. **Employment scams** are on the rise. According to CNBC, the number of employment scams doubled in 2018 as compared to 2017. The current market situation has led to high unemployment. Economic stress and the coronavirus's impact have significantly reduced job availability and job loss for many individuals. A case like this presents an appropriate opportunity for scammers. Many people are falling prey to these **scammers** using the desperation that is caused by an unprecedented incident. Most scammers do this to get personal information from the person they are **scamming**. **Personal information** can contain addresses, bank account details, social security numbers, etc. I am a university student, and I have received several such scam emails. The scammers provide users with a very **lucrative job opportunity and later ask for money** in return. Or they require investment from the job seeker with the promise of a job. This is a dangerous problem that can be addressed through Machine Learning techniques and Natural Language Processing (NLP).

Stages of Development



Flowchart

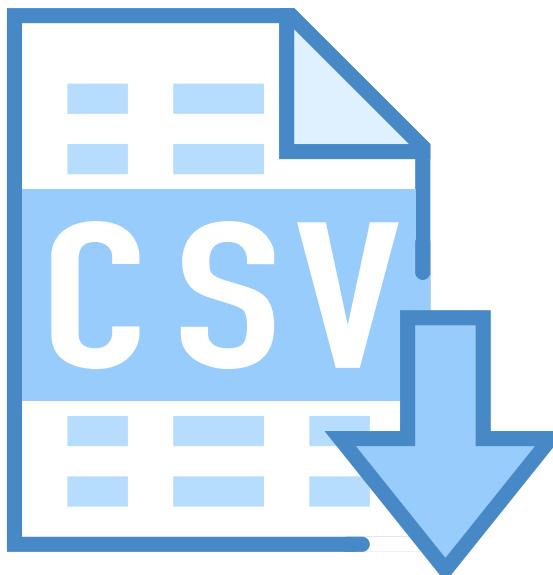


TECHNOLOGIES

DATASET

Kaggle Dataset

Real / Fake Job Posting Prediction
Dataset of real and fake job postings

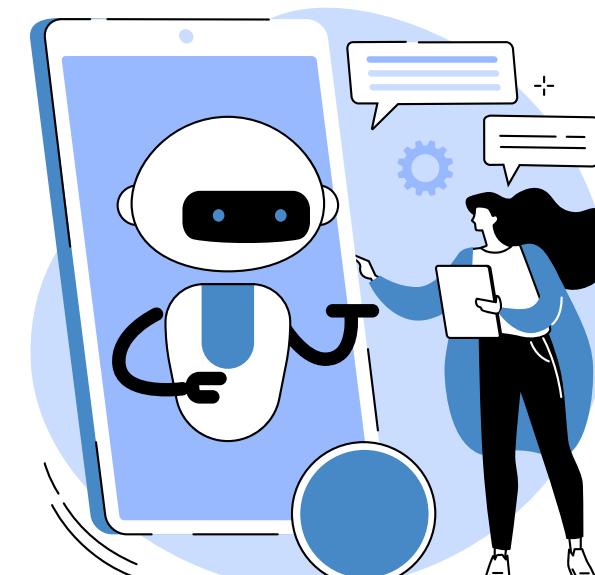


TECHNOLOGIES

Natural Language Processing NLP

NLP Pipelining

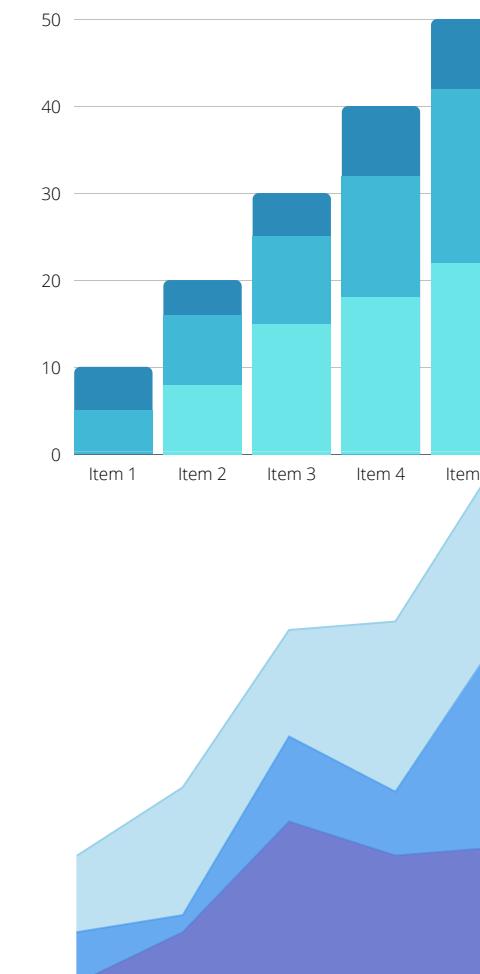
Numpy, ScikitLearn, Pandas, Spacy



ALGORITHM

Support Vector Machine SVM

Random Forest

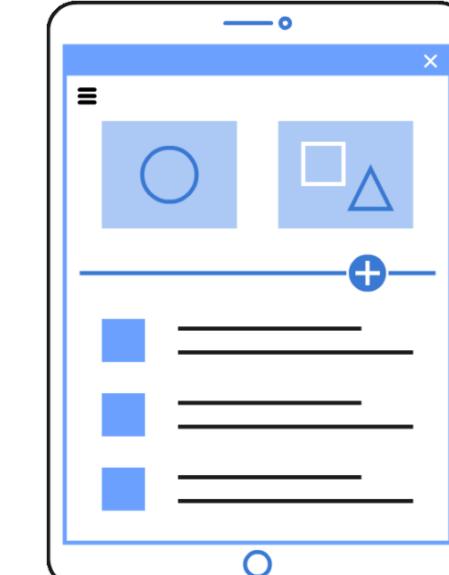


DEPLOYMENT

Frontend with HTML, CSS, JavaScript

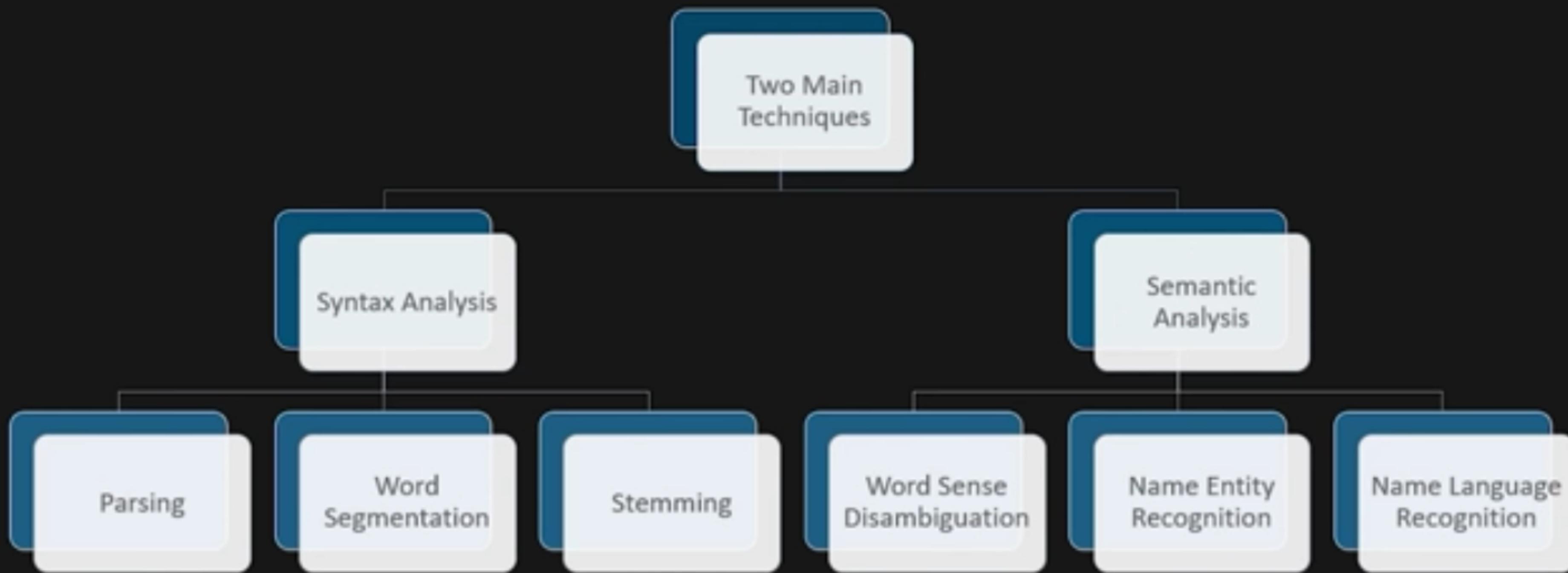
Backend Flask Python

ML Model in Jupyter Notebook Python

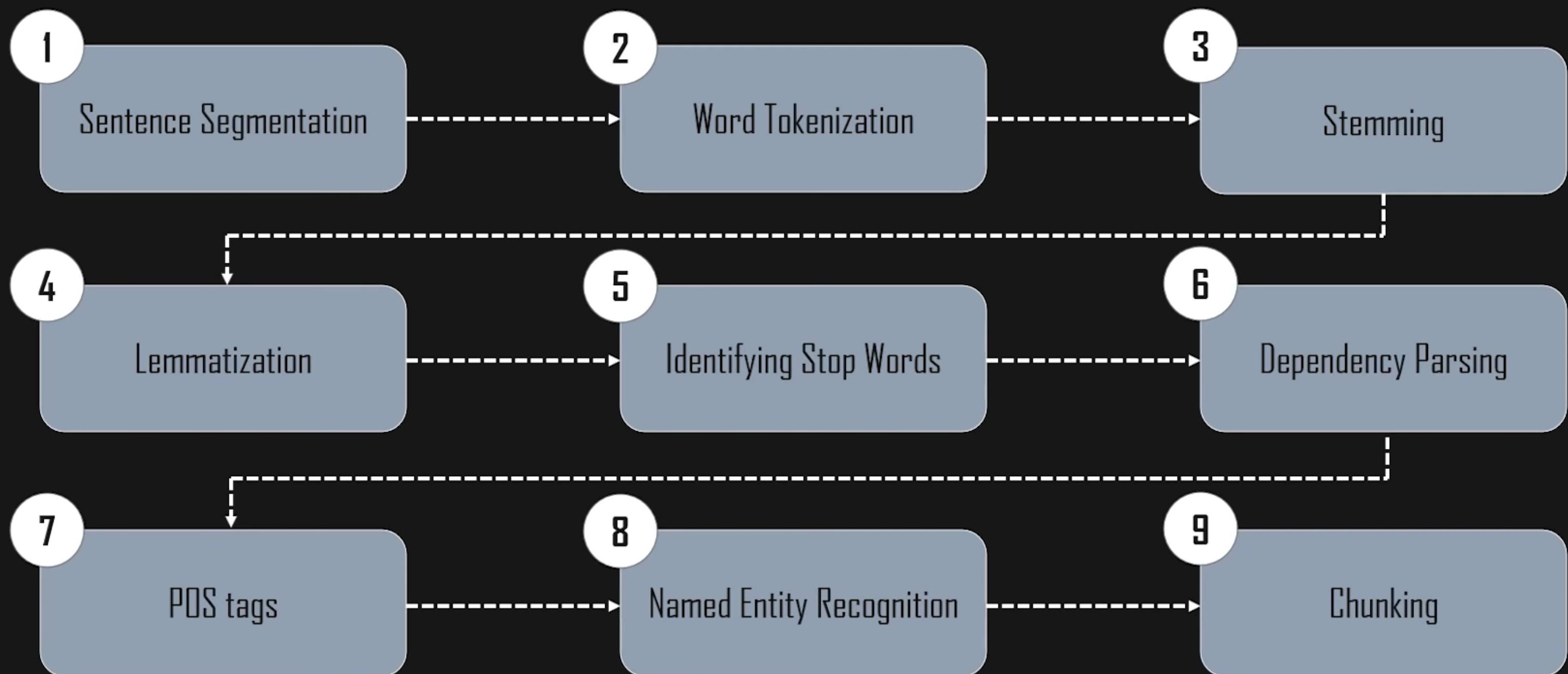


1

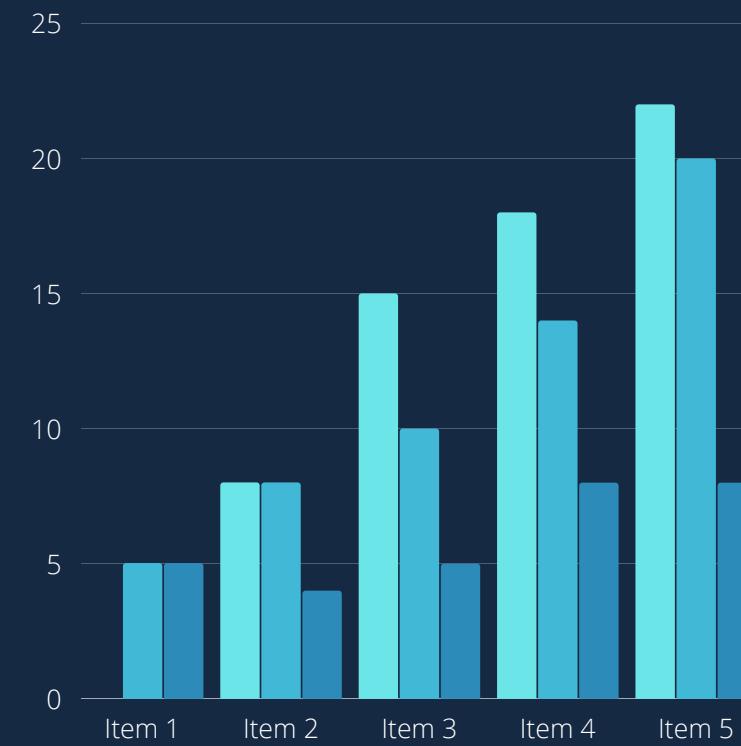
NLP- Natural Language Processing



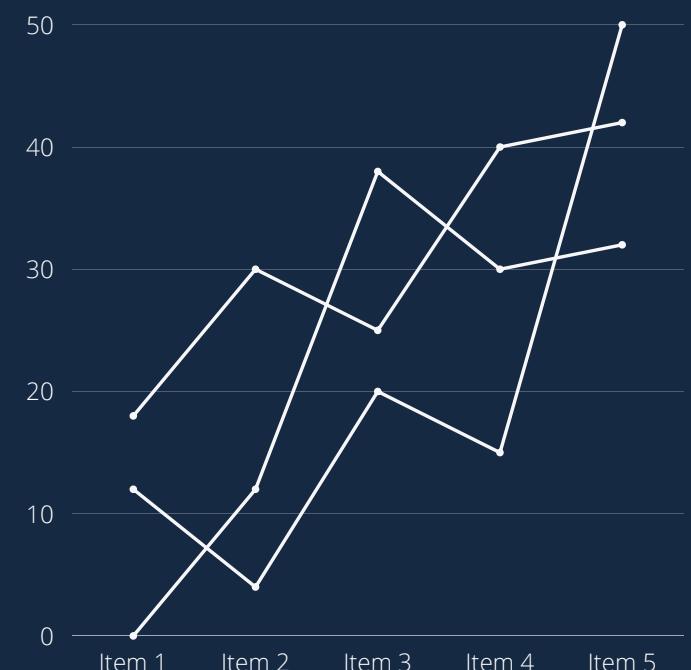
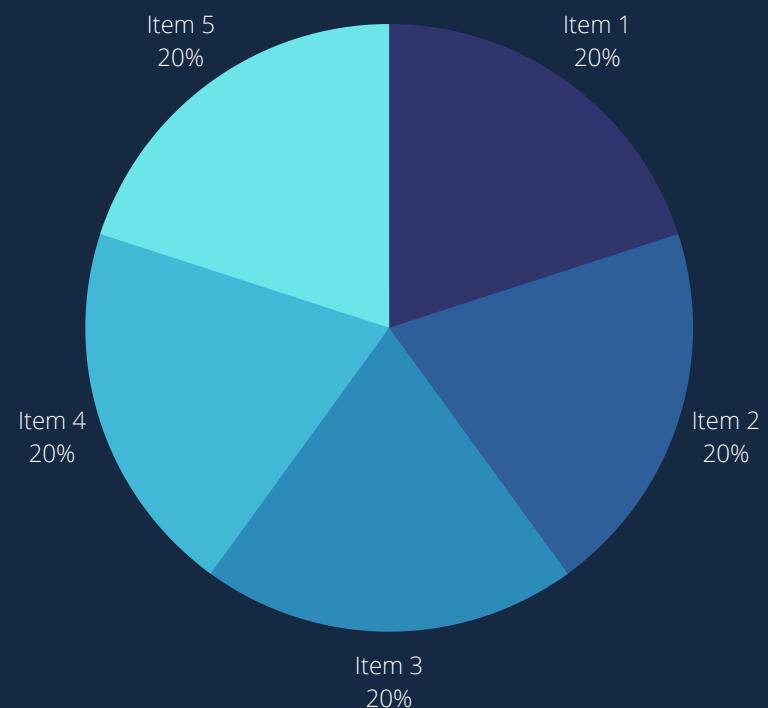
NLP Pipelining Steps



SVM

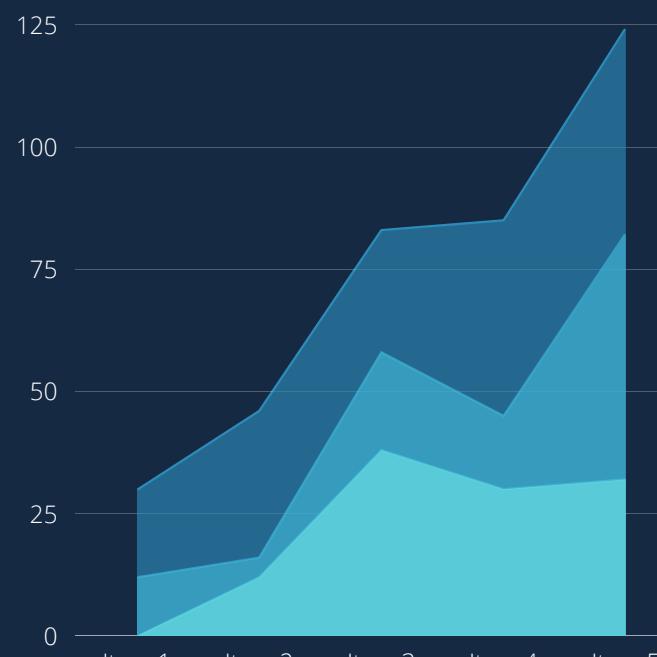
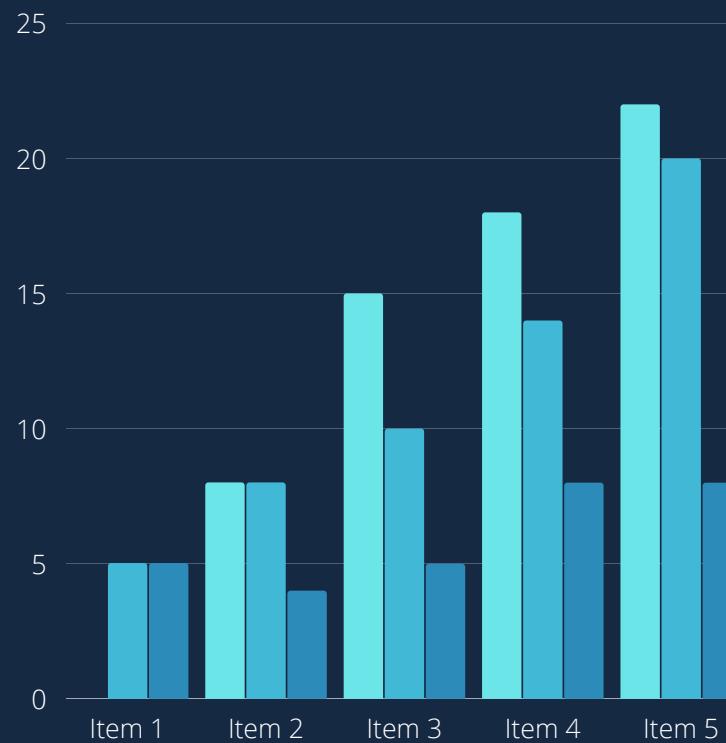


- Supervised Learning algorithm
- Classification as well as Regression problems
- The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.
- SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

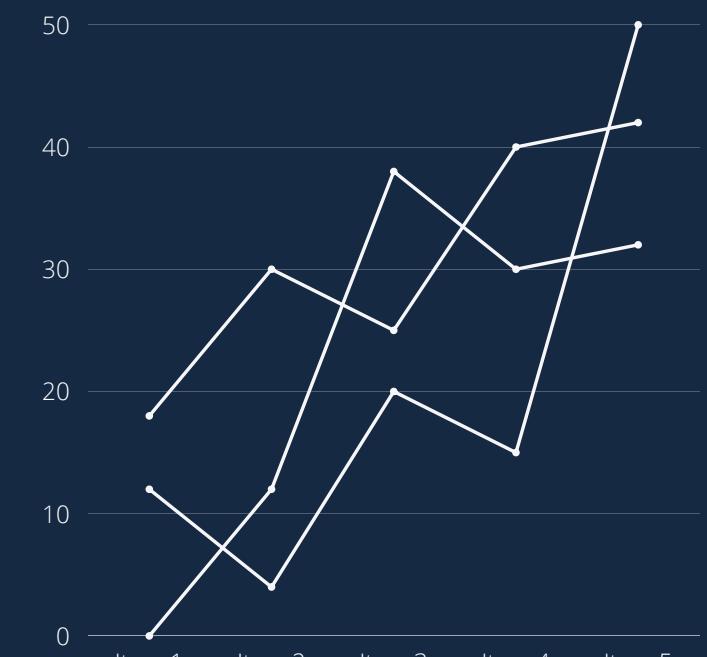
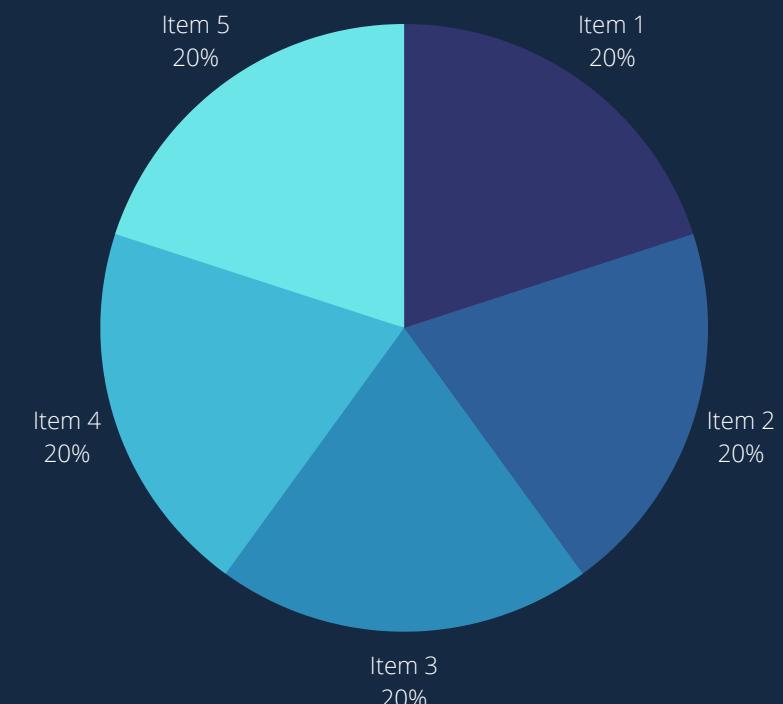


3

Random Forest

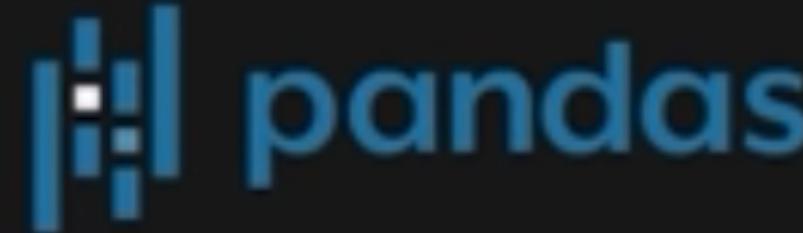


- Random forest is a supervised learning algorithm which is used for both classification as well as regression.
- Creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.



4

Python Libraries



NumPy



spaCy

5

Deployment



CSS



Flask



6

Data Visualization

**Data Viewed,
graphs, tables,
predictions in our
Website**

