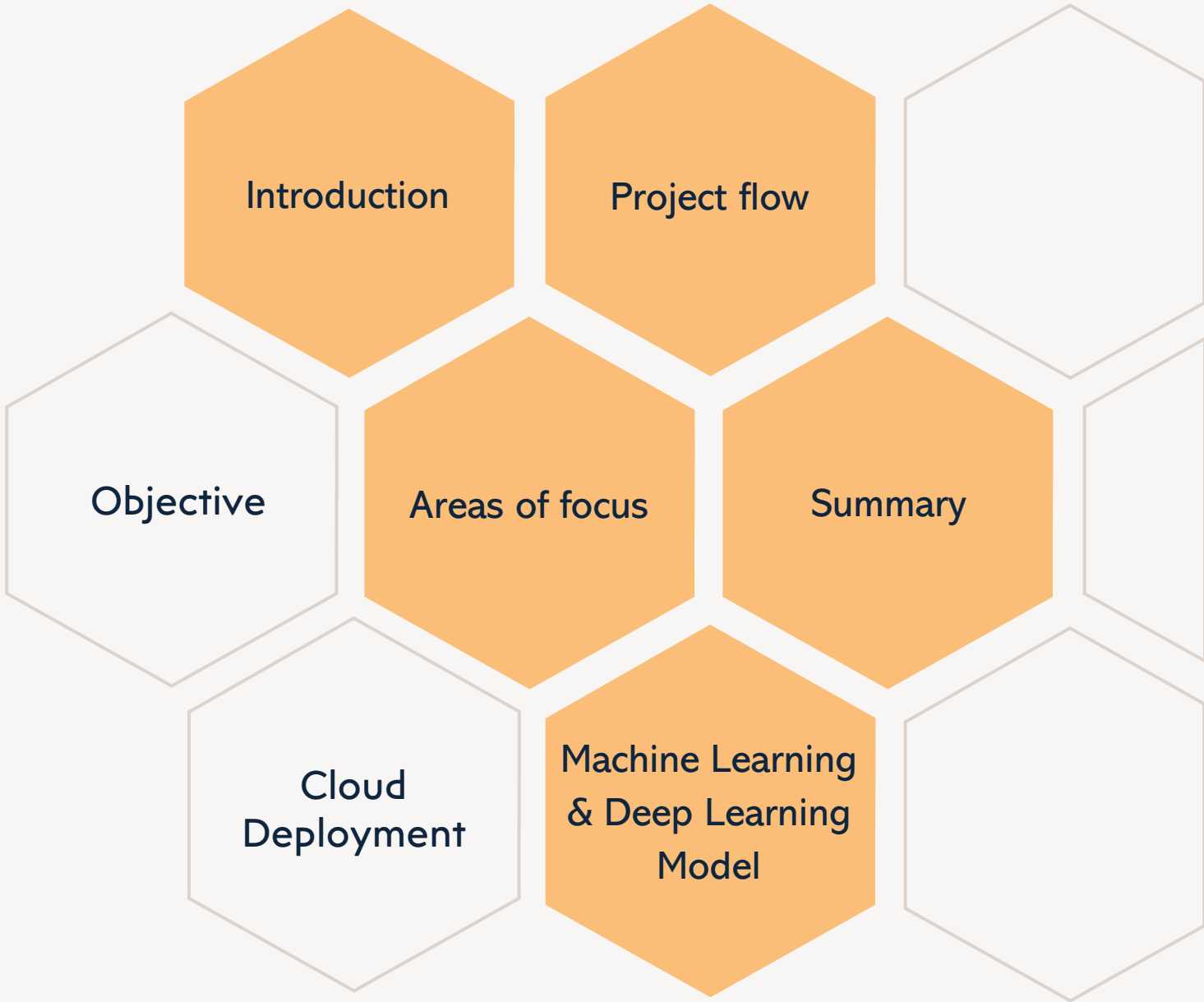
Times Job Post Sentiment Analysis

Capstone Project:
Identify Fake Job Post Vs Real Job Post using RNN(LSTM)

TRUPTI PARMAR

PGA-43

Agenda


Introduction
Project flow
Objective
Areas of focus
Summary
Cloud Deployment
Machine Learning & Deep Learning Model

Introduction

A **job posting** is an official advertisement of job openings published for job seekers. It is kind of a written announcement to inform candidates about available job opportunities.

Where are job postings published

- Career site
- Social media
- Job boards
- Paid ads
- Local media
- Newspapers
- Newsletter

Primary goal:

Detecting whether the jobs are real or fake on the basis of a suitable model and extracting useful information from the given dataset, i.e., getting useful insights from the job postings.

Motive:

- Now-a-days, there are a lot of job scams because of unemployment.
- There are a lot of websites which connect recruiter to a suitable candidate.
- Sometimes, fake recruiters post a job posting on the job portal with a motive to get money. This problem occurs with many job portals.
- Later, people shift to a new job portal in search of real job but the fake recruiters join this portal as well. So, it is important to detect real and fake jobs.

Plan/Flow for Project Development





Planning & Research Topic

Identify Problems regarding fake post

Data Collection

Using Web Scrapping from timesjobs-job portal

Data Preprocessing

Coordinate data cleaning

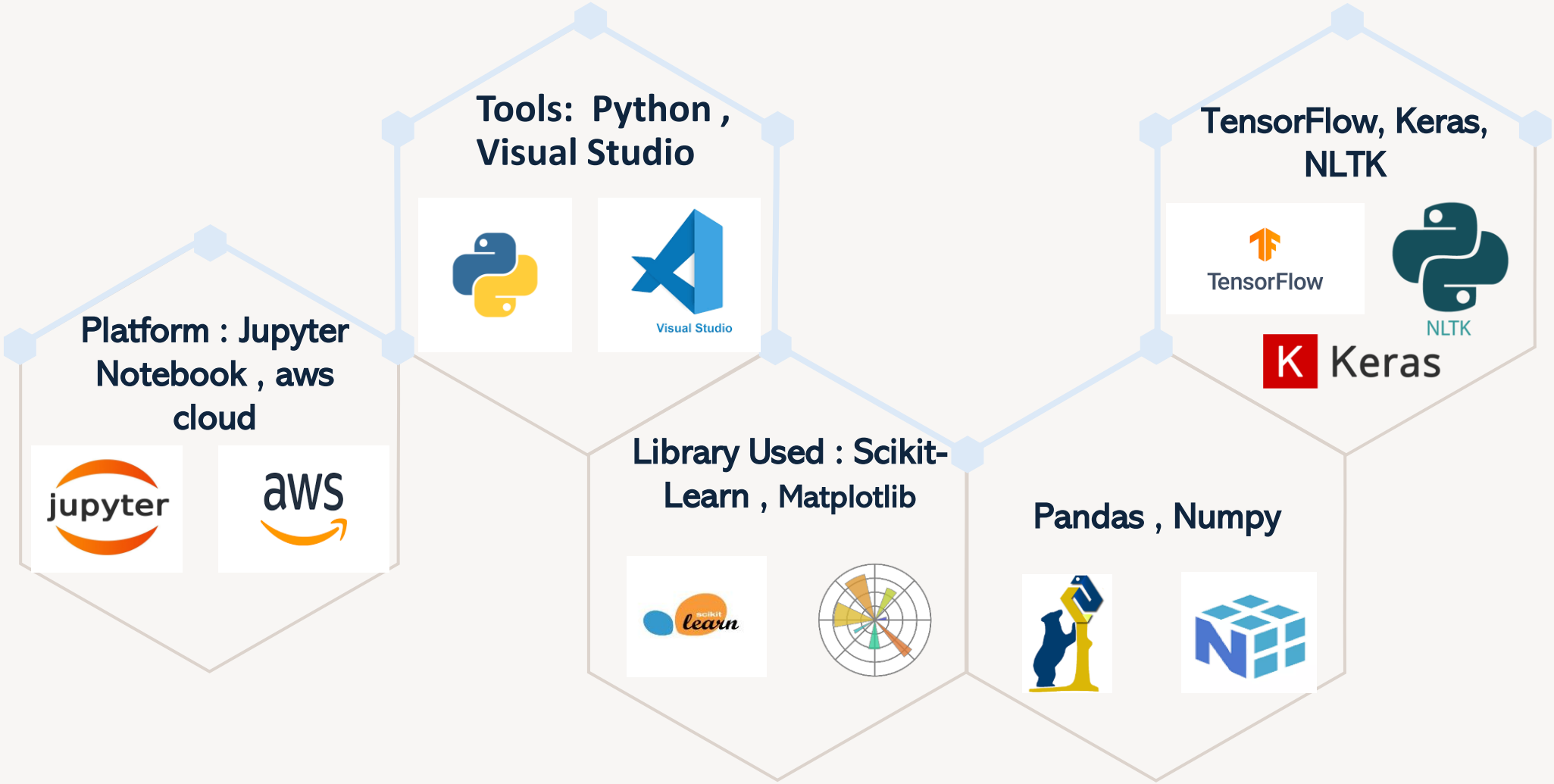
Machine learning model

Build Machine learning model and interpretation

Launch

Deploy Model on web application(localhost \cloud)

TOOLS AND PLATFORM USED



Flow for Machine Learning Model Development

Data Preprocess	Data Visualization/ NLP	Machine Learning & Deep Learning Model	Model Evolution	Web Development
Import data Data Cleaning	Word Cloud TF-TDF DTM Sentiment Analysis	Data Partition Train Model Test Model	Performance Metrics Accuracy Loss	Web Develop using Visual studio Code Deploy on AWS Cloud

Dataset Attributes

Attribute Name	Data type
Title	objective
Company	objective
Experience	objective
Salary	objective
Location	objective
Posted Date	Date&time
Key skills	objective
Description	objective
URL	objective

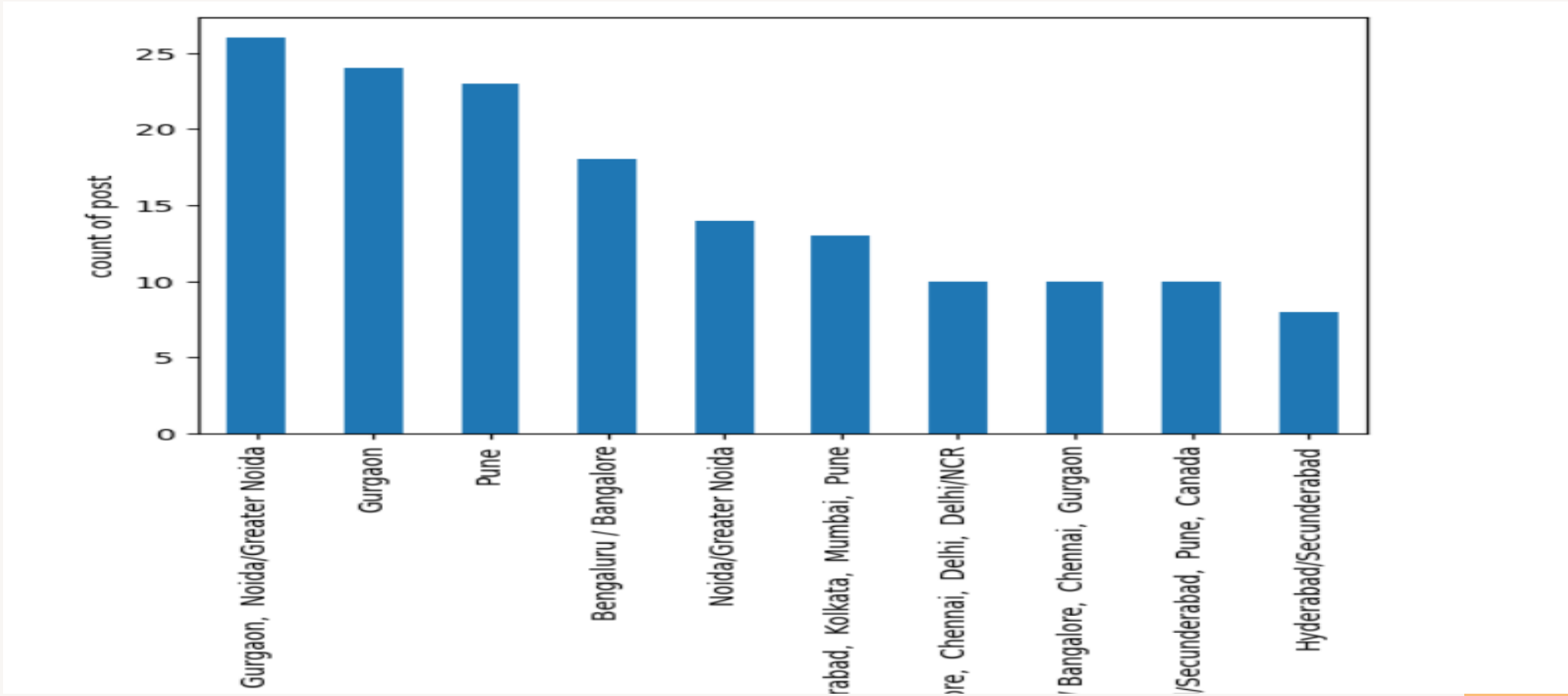
Data Preprocessing

Sentiment Analysis Using “Natural Language Processing”

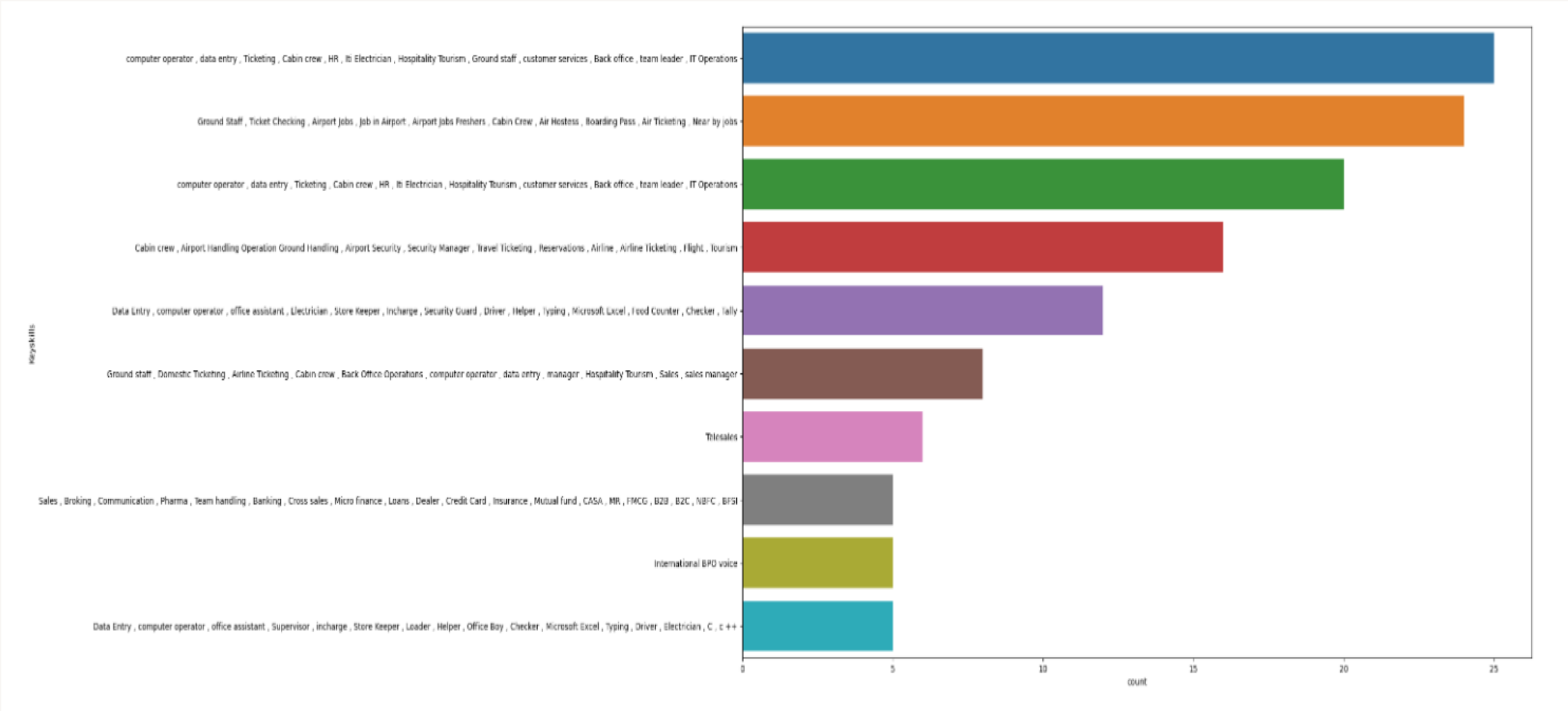
Text Mining/Data Cleaning

- Convert Lower Case
- Remove Numbers
- Remove Punctuation word
- Remove Stop Words
- Remove URLs
- Applying Lemmatization
- Remove White space

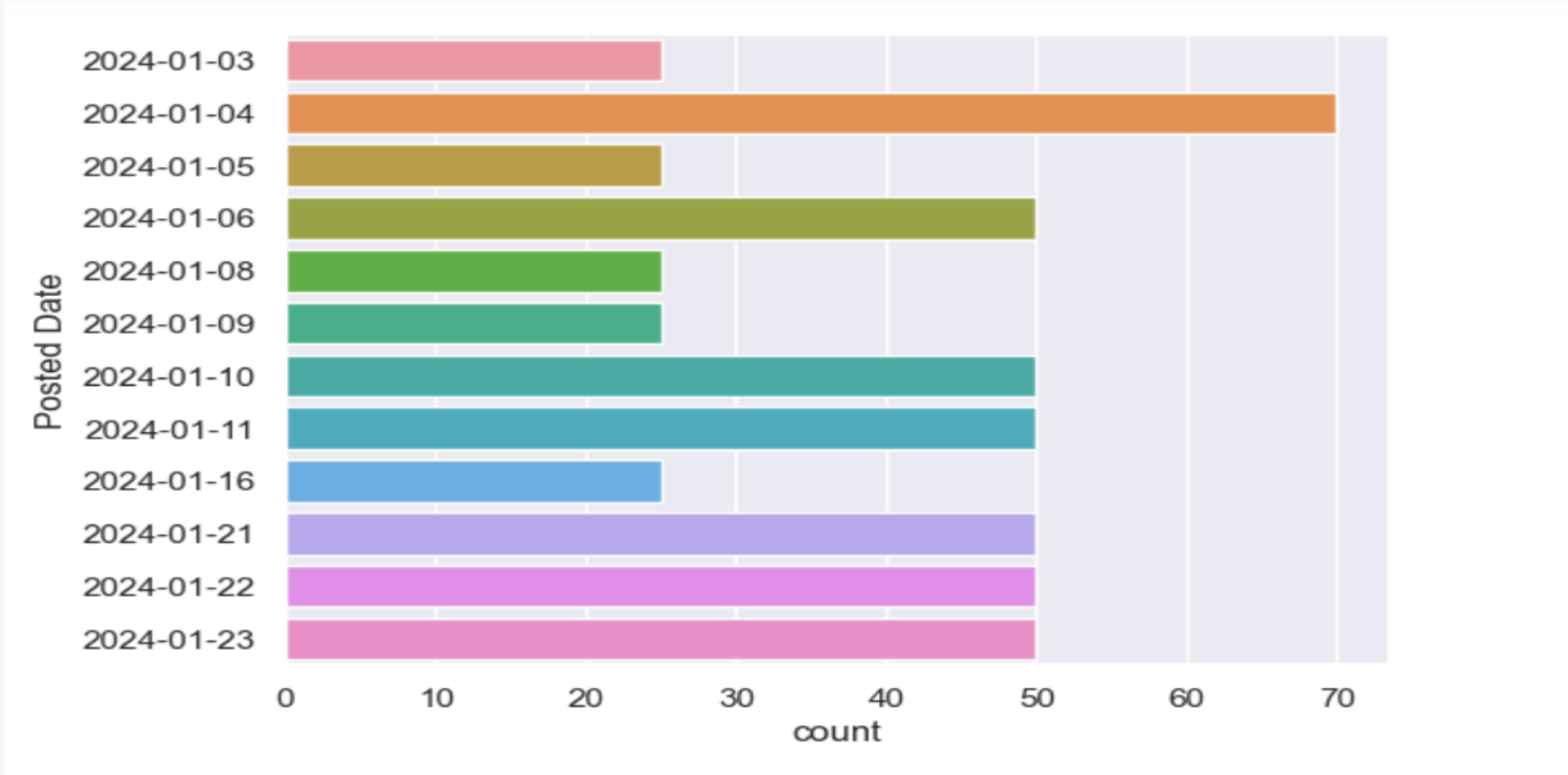
Location Visualization



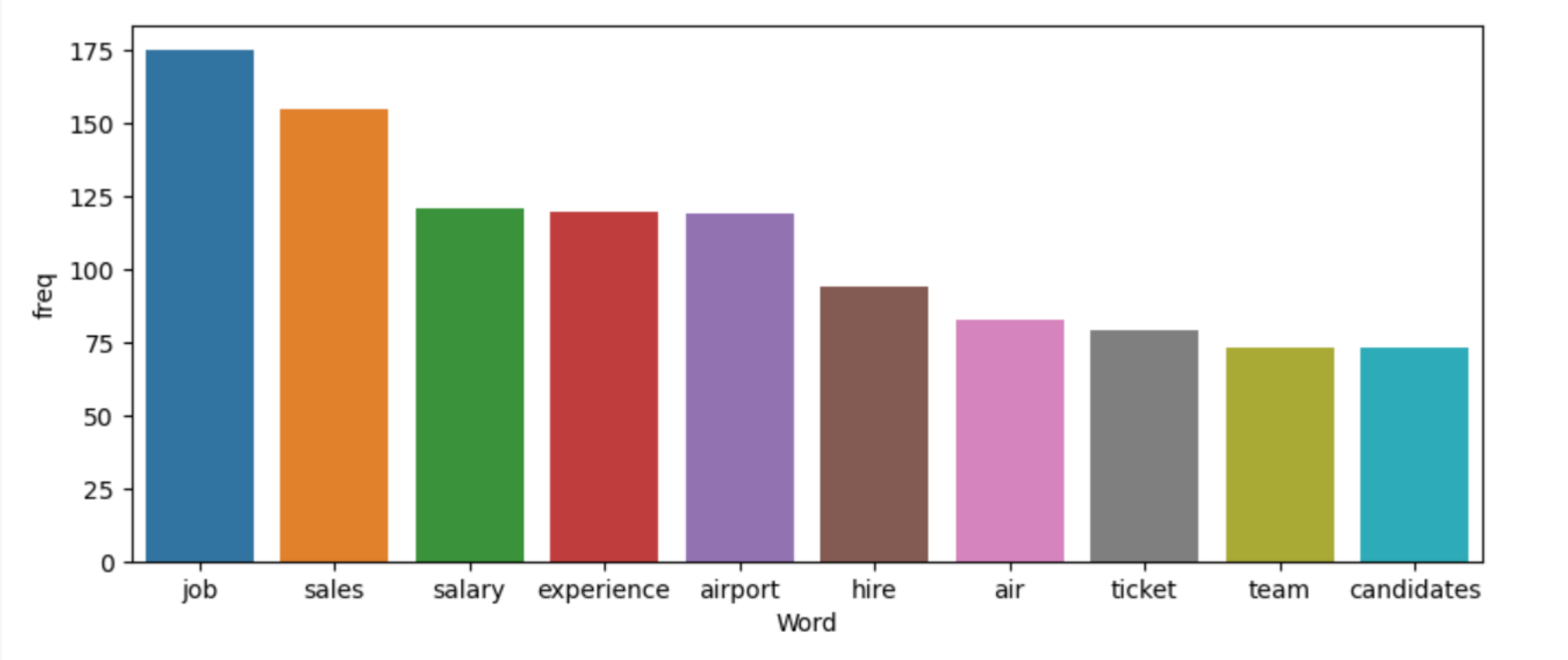
Key Skill Visualization



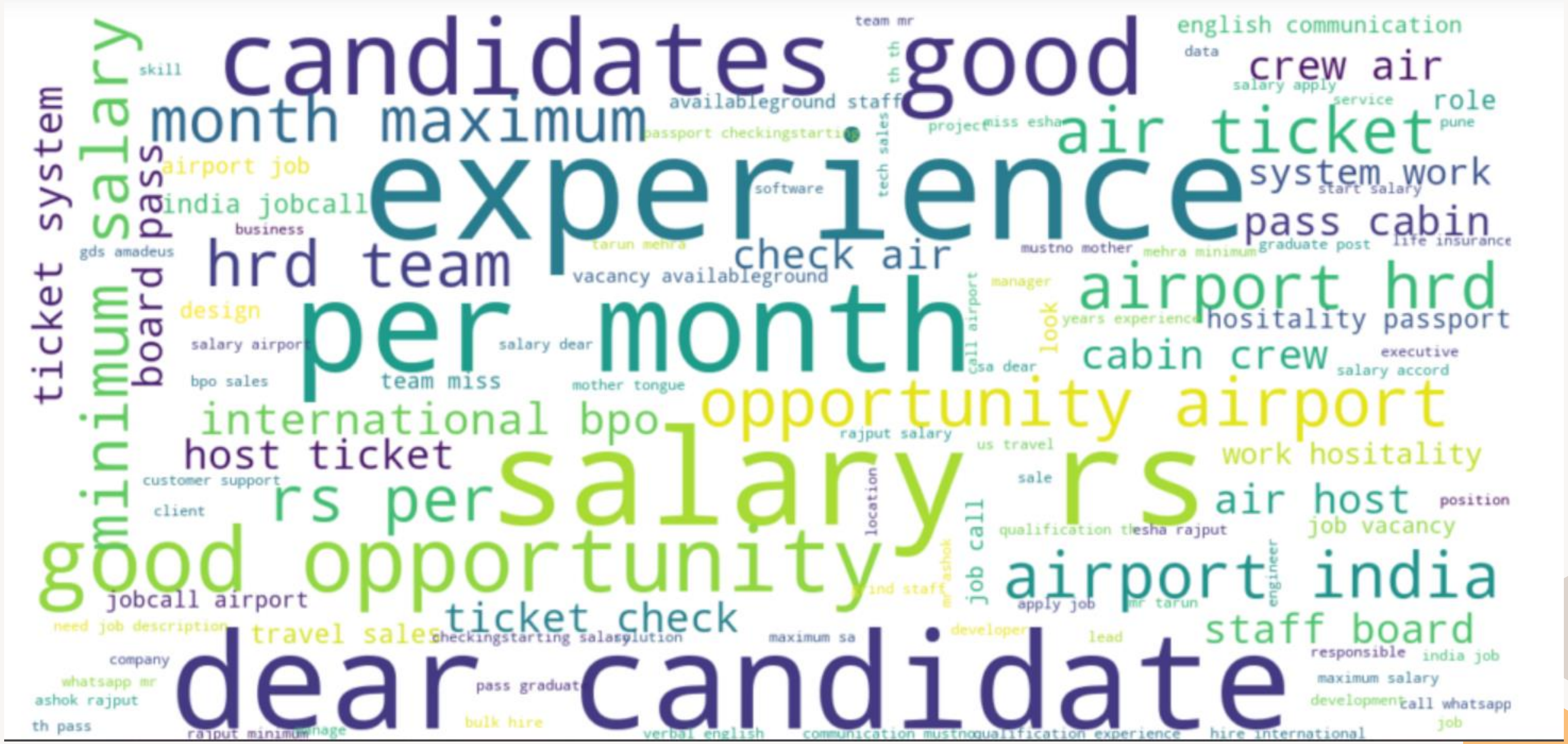
Posted Date Visualization



Most Important Word Frequency wise



Word Cloud

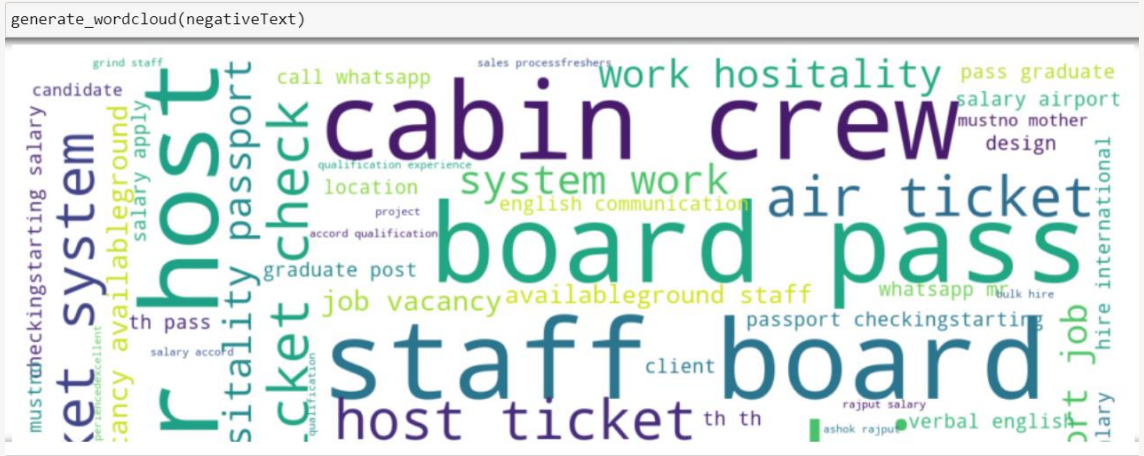

experience
candidates good
per month
salary rs
dear candidate
good opportunity
month maximum
minimum salary
ticket system
board pass
hrd team
air ticket
opportunity airport
airport hrd
international bpo
host ticket
rs per
airport india
cabin crew
crew air
pass cabin
system work
check air
ticket check
staff board
air host
work hositality
hositality passport
english communication
role
job vacancy
travel sales
jobcall airport
india jobcall
airport job
team miss
design
skill
business
gds amadeus
salary airport
bpo sales
customer support
client
availableground staff
vacancy availableground
salary dear
mother tongue
rajput salary
us travel
sale
qualification
esha rajput
apply job
maximum salary
india job
call whatsapp
responsible
executive
salary accord
position
graduate post
life insurance
years experience
software
business
team mr
data
salary apply
service
pune
start salary
mustno mother
mehra minimum
manager
job call
location
engineer
developer
lead
ashok rajput
th pass
whatsapp mr
company
need job description
pass graduate
bulk hire
verbal english
hire international
job

NEGATIVE, NEUTRAL & POSITIVE WORD-CLOUD

generate_wordcloud(PositiveText)

generate_wordcloud(neutralText)

generate_wordcloud(negativeText)



Sentiment Analysis

Presentation title

Cosine Similarity

```
def recommend(skills):
    index =data_cosine[data_cosine['Keyskills'] == skills].index[0]
    distances = sorted(list(enumerate(similarity[index])),reverse=True,key = lambda x: x[1])
    for i in distances[1:4]:
        print(data_cosine.iloc[i[0]].text)
```

```
recommend('Plant Engineer')
```

project manager (work from home)bengaluru / bangalore, chennai, gurgaonas a project manager in a saas startup, your primary responsibility is to plan, execute, and oversee various projects within the organization while coordinating with the team ...

canada project required piping engineerbengaluru / bangalore, chennai, hyderabad/secunderabad, pune, canadarespons ibilities:responsible for fel deliverables required to supply a quality engineering design package to the engineering design team.performing pipe stress / flexibility a...

canada project required oil & gas engineerbengaluru / bangalore, chennai, hyderabad/secunderabad, pune, canadarole expectations:this role requires the application of standard engineering techniques, procedures and criteria on job assignmen ts.we strive to be known for unmatched quality...

```
recommend('Senior Architect')
```

canada project required piping engineerbengaluru / bangalore, chennai, hyderabad/secunderabad, pune, canadarespons ibilities:responsible for fel deliverables required to supply a quality engineering design package to the engineering design team.performing pipe stress / flexibility a...

canada project required hvac engineerbengaluru / bangalore, chennai, hyderabad/secunderabad, pune, canadato suppor t our rapid growth, coolsys energy design is in search of an hvac design engineer to join our princeton team!in this role, t he engineer/designer will:preparation and ...

canada projects require electrical engineerbengaluru / bangalore, chennai, hyderabad/secunderabad, pune, canadadut ies & responsibilitiesdesigning, implementing, maintaining, and improving electrical design, products and systems.collaborat e with other engineers to design, and develo...

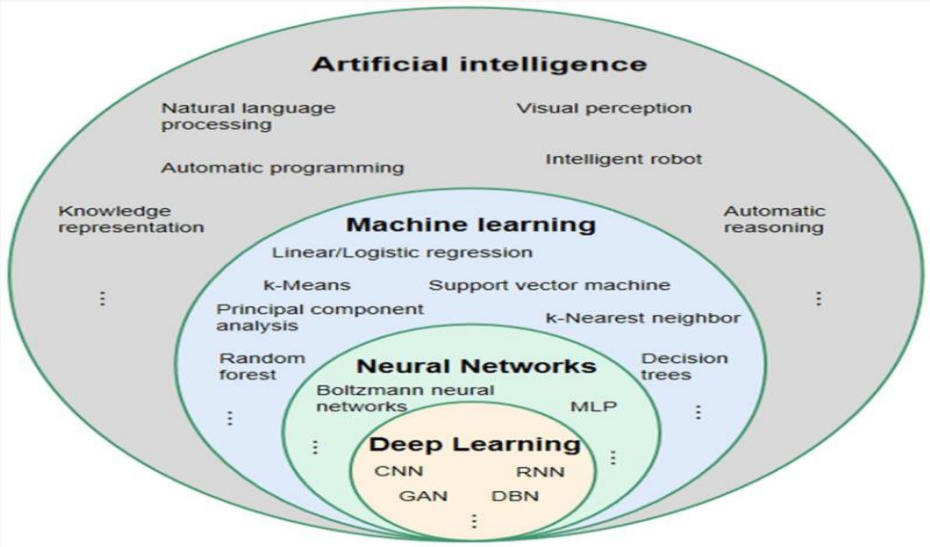
Sentiment Analysis

Sentiment Analyse using vaderSentiment Score

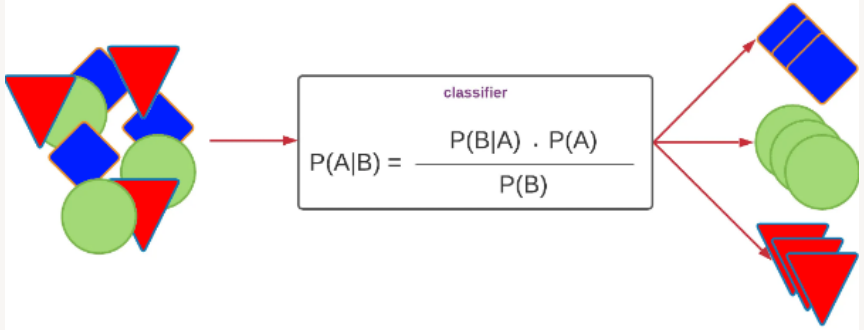
Number of Positive Sentiment : 171
Number of Negative Sentiment : 25
Number of Neutral Sentiment :299

	Description	compound	neg	neu	pos	Sentiment
0	consult patients understand health concernswri...	0.0000	0.000	1.000	0.000	Neutral
1	support rapid growth coolsys energy design sea...	0.8225	0.000	0.575	0.425	Positive
2	minimum requirements years progressive experie...	0.5106	0.000	0.820	0.180	Positive
3	job descriptionsenior architect lead design te...	0.0000	0.000	1.000	0.000	Neutral
4	report relationshipsthe successful candidate r...	0.4404	0.127	0.610	0.263	Neutral
5	job description responsibilitiesthe part distr...	0.0000	0.000	1.000	0.000	Neutral
6	job descriptionicg search senior quantity surv...	0.0000	0.000	1.000	0.000	Neutral
7	responsibilitiescoordinate task accord priorit...	0.0000	0.000	1.000	0.000	Neutral
8	role expectationsthis role require application...	-0.0772	0.080	0.920	0.000	Negative
9	job descriptionrenardet search lead mechanical...	0.0000	0.000	1.000	0.000	Neutral

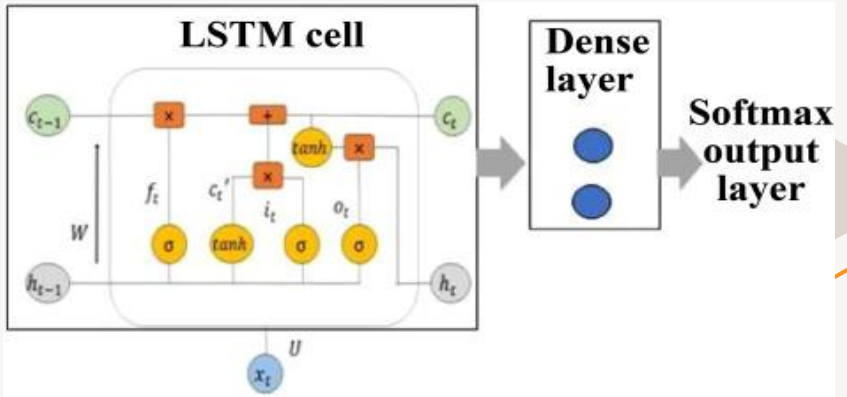
Machine Learning Model & Deep Learning Model



Naive Bayes Classifier



Long Short-Term Memory RNN Model



Machine Learning & Deep Learning Model

Naive Bayes Classifier Model

```
from sklearn.naive_bayes import MultinomialNB

Naivemodel = MultinomialNB(alpha=1)  # Laplace
Naivemodel.fit(X_train, y_train)
```

```
MultinomialNB
MultinomialNB(alpha=1)
```

LSTM Based Model

```
print(model.summary())
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 250, 100)          2500000

 spatial_dropout1d (Spatial  (None, 250, 100)          0
 Dropout1D)

 lstm (LSTM)                 (None, 100)               80400

 dense (Dense)               (None, 3)                 303

 activation (Activation)     (None, 3)                 0

=================================================================
Total params: 2580703 (9.84 MB)
Trainable params: 2580703 (9.84 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
None
```

```
# Optimizer
adam = optimizers.Adam(lr = 0.001)
# Compile the model
model.compile(loss = 'categorical_crossentropy', optimizer = adam, metrics = ['accuracy'])
```

Machine Learning Model

Naive Bayes Classifier Model

```
print("============================================")
print("Accuracy of Test Model : ",accuracy_score(train['Predicted'], train['Sentiment']))
print("============================================\n")
```

```
============================================
Accuracy of Test Model :  0.653179190751445
============================================
```

	precision	recall	f1-score	support
Negative	0.14	1.00	0.25	20
Neutral	0.99	0.99	0.99	209
Positive	0.00	0.00	0.00	117
accuracy			0.65	346
macro avg	0.38	0.66	0.41	346
weighted avg	0.61	0.65	0.61	346

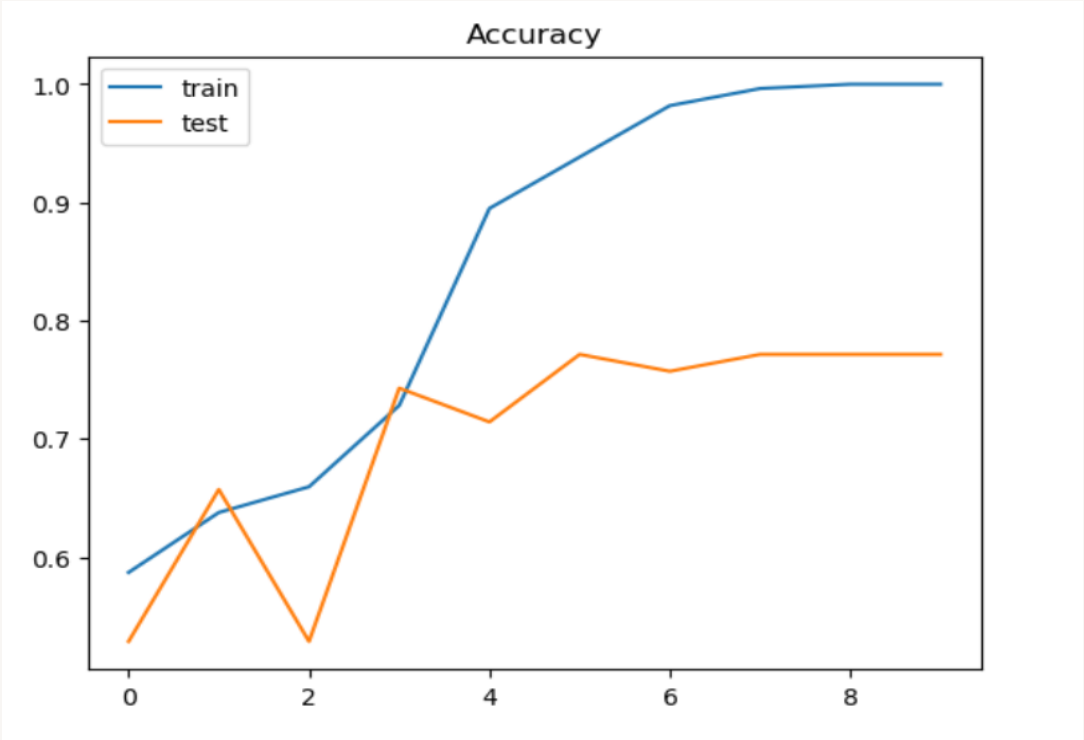
LSTM based Model

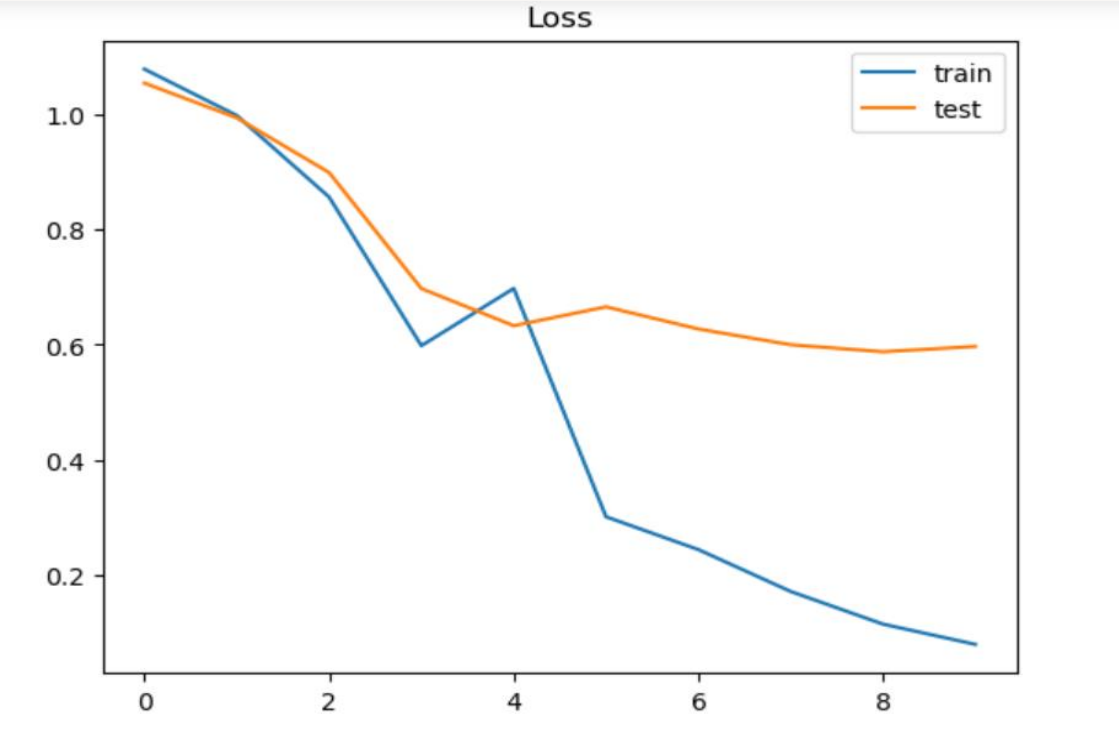
```
accuracy = model.evaluate(X_test,Y_test)
print('Test set\n  Loss: {:0.3f}\n  Accuracy: {:0.3f}'.format(accuracy [0],accuracy [1]))
```

```
5/5 [==============================] - 0s 38ms/step - loss: 0.4438 - accuracy: 0.8322
Test set
  Loss: 0.444
  Accuracy: 0.832
```

	precision	recall	f1-score	support
Negative	1.00	0.78	0.88	9
Neutral	0.84	0.89	0.86	88
Positive	0.80	0.75	0.77	52
accuracy			0.83	149
macro avg	0.88	0.80	0.84	149
weighted avg	0.83	0.83	0.83	149

RNN Model Performance(Accuracy and Loss) on each epoch



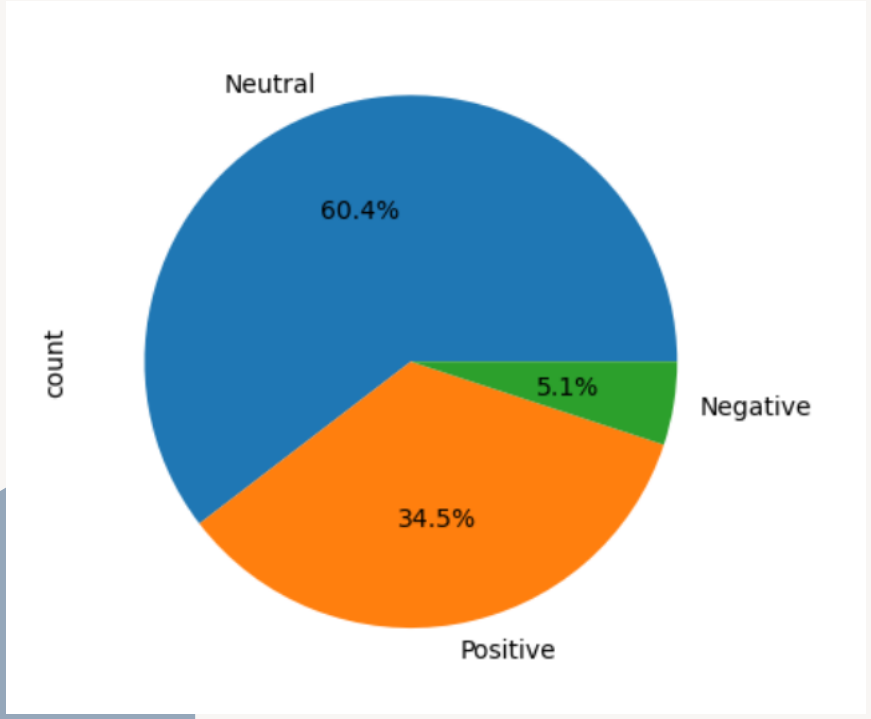


Summary

Recurrent neural networks (RNNs) are a type of neural network that are used for processing sequential data.

It gives good accuracy in test data ,for further deployment I used LSTM model for cloud development.

LSTM model have 83 % Accuracy which is good fit model



Areas of focus

B2B market scenarios

Visualize Job post made by scammers and can take action against them.

Develop winning strategies to keep ahead of the competition on job portals.

Increase **website traffic**

B2C scenarios

Help Job Seekers
by identify scammers, who can steal users identities and can hack users details

Help to identify Right job
Also help full to find job based on key skills

Help to identify The Spammers

References

Data

- https://www.timesjobs.com/candidate/job-search.html?searchType=personalizedSearch&from=submit&txtKeywords=&txtLocation=india

Code

- https://www.analyticsvidhya.com
- https://stackoverflow.com/
- https://www.tensorflow.org/guide/keras

cloud

- https://console.aws.amazon.com/console/home

Cloud Deployment

Welcome to TimesJob Post Sentiment Analyzer

Say Something: ....

Submit

The Sentiment of

' canada project required oil & gas engineerbengaluru / bangalore, chennai, hyderabad/secunderabad, pune, canadarole expectations:this role requires the application of standard engineering techniques, procedures and criteria on job assignments.we strive to be known for unmatched quality...'

is 46.0% positive !

Score table

SENTIMENT METRIC	SCORE
Positive	0.0
Neutral	0.954
Negative	0.046
Compound	0.46

Suggestion:

Can apply CNN model ,Using CNN can find fake vs Real by Uploading Images.

Other Factors to find Frauds

Six Common Signs That Job Isn't Real.

1. Contact can’t be found in a Google search
2. No company information
3. Grammatical errors & spelling mistakes
4. Money is involved immediately
5. Personal information required immediately
6. Sounds too good to be true

Learnings:

- Real world Data Collection(Web Scraping)
- Data Cleaning/Data Preprocessing
- Data Visualization
- Implements Machine Learning & Deep Learning Models
- Dealing with Error by Problem Solving Skills
- Cloud Deployment with AWS cloud
- Time Management Skills

Thank you

Trupti Parmar

trupti4focus@gmail.com

https://www.linkedin.com/in/trupti-Parmar/

https://github.com/Trupti2011