



iNeuron

NEWS ARTICLE SORTING



Presented by Krishna Chandra Yadav

Objective- Problem Statement

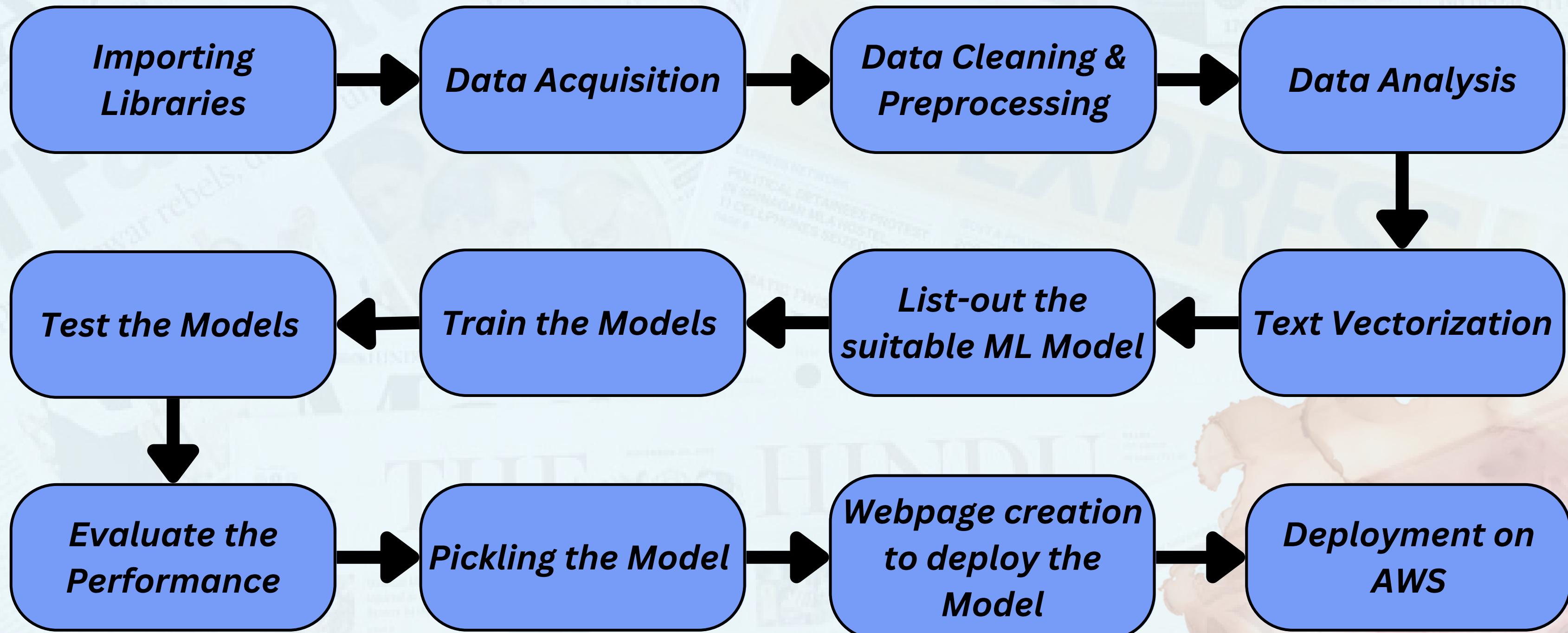
- In today's data-driven world, where news companies store terabytes of data, the pursuit of valuable insights is paramount. Among numerous applications of analytics, news article classification stands out as a prime example.
- On the internet, vast daily news volumes are generated. With increasing user demand for information, efficient news classification is crucial. Machine learning models can automate this process, helping users quickly access relevant news and offer personalized suggestions based on their interests.

Benefits /Uses

Classifying news articles means sorting them into different groups, like sports, politics, business or entertainment. This helps us find articles we like faster. News websites use it to make our experience better.

- News article classification organizes articles into categories, making it easier for readers to find what they're interested in.
- It helps advertisers show relevant ads to the right audience, improving the effectiveness of advertising.
- News websites use classification to enhance user experience and provide better content recommendations.

Pipeline



Data Acquisition

Data has been taken from Kaggle : <https://www.kaggle.com/c/learn-ai-bbc/data>



Dataset Description

Files:

- BBC News Train.csv - the training set of 1490 records
- BBC News Test.csv - the test set of 736 records
- BBC News Sample Solution.csv - a sample submission file in the correct format

Data Field:

- ArticleId - Article id unique no. given to the record
- Article - text of the header and article
- Category - category of the article (tech, business, sport, entertainment, politics.)

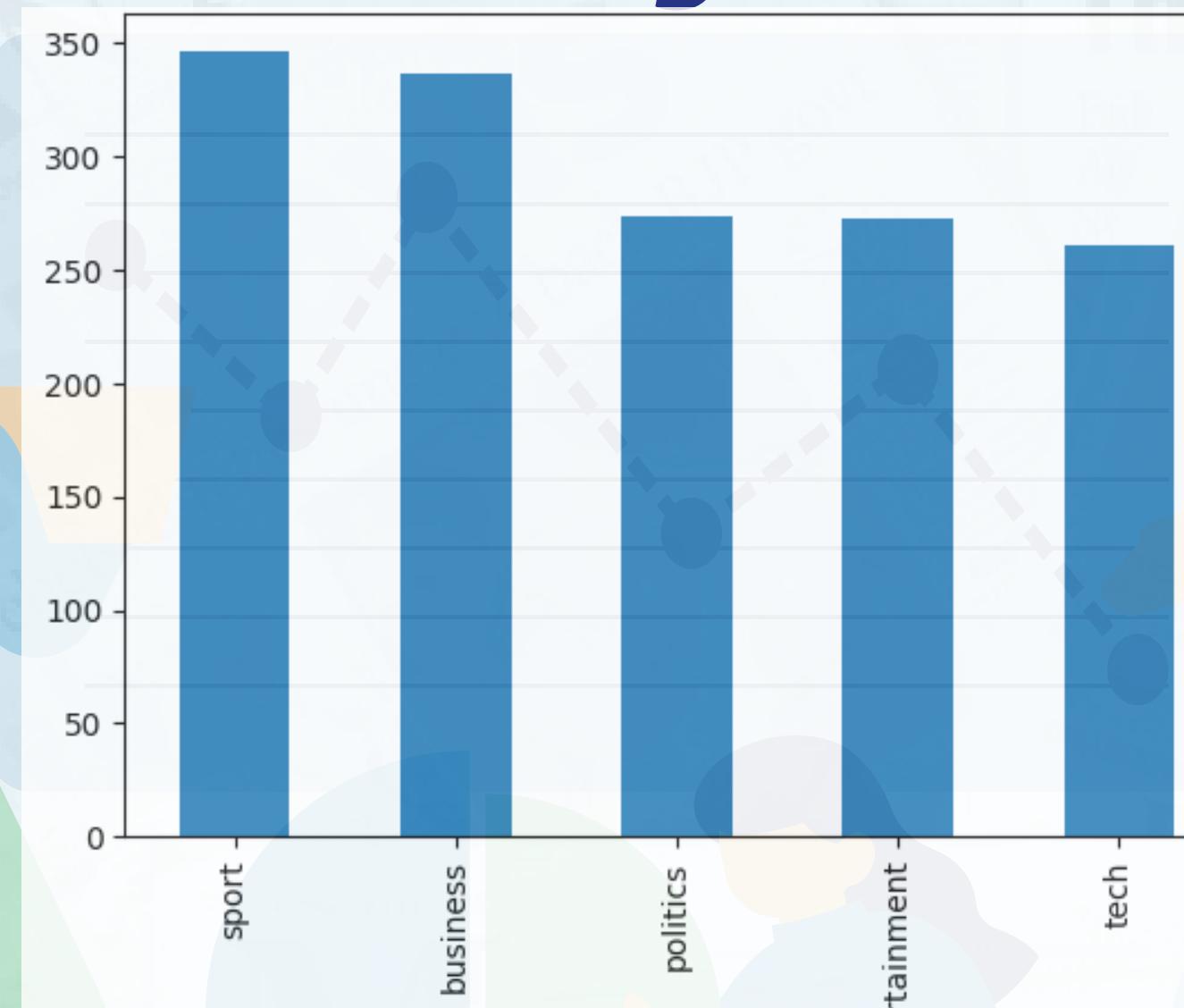
Data Cleaning & Preprocessing

- Lower casing of all Documents.
- Punctuation Removal.
- Stop-words Removal.
- Tokenization.
- Stemming.
- Label Encoding.



Data Analysis

Here, we can see that the number of articles per classes (Sport, Business, Politics and Tech) are roughly balanced.



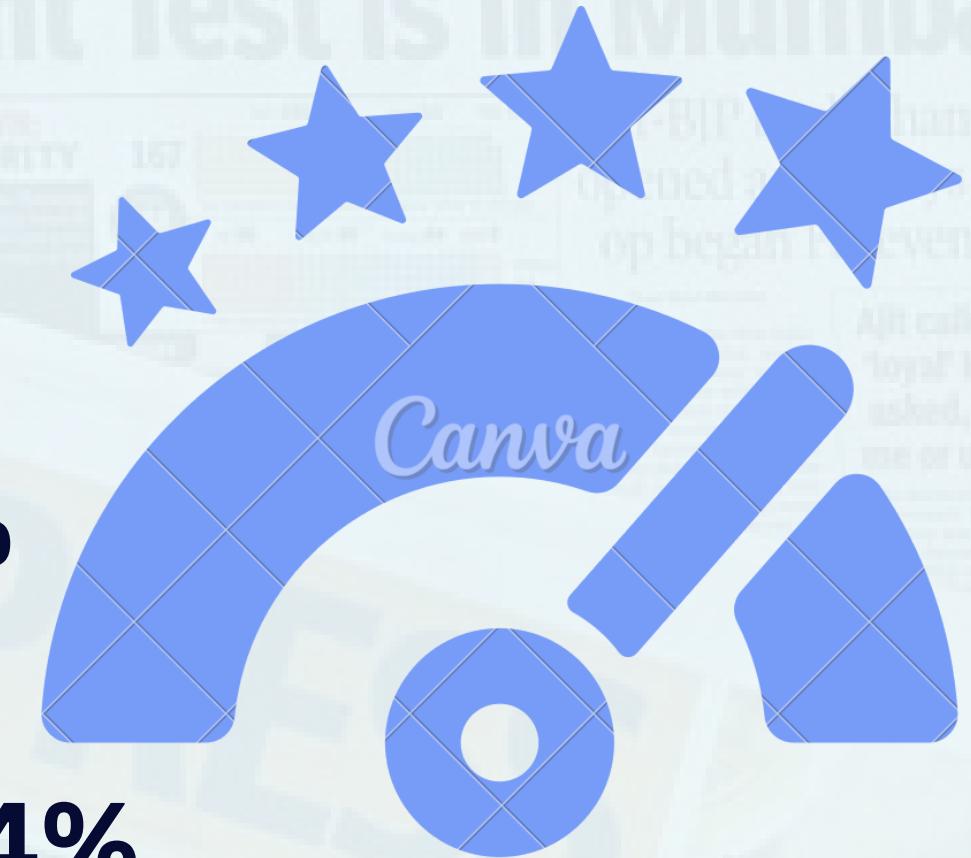
Text Vectorization & Models

- TF-IDF is used for text vectorization.
- Model used for experiments:
 - Support Vector Machine
 - Random Forest
 - Gradient Boosting Classifier
 - Ada-Boost Classifier



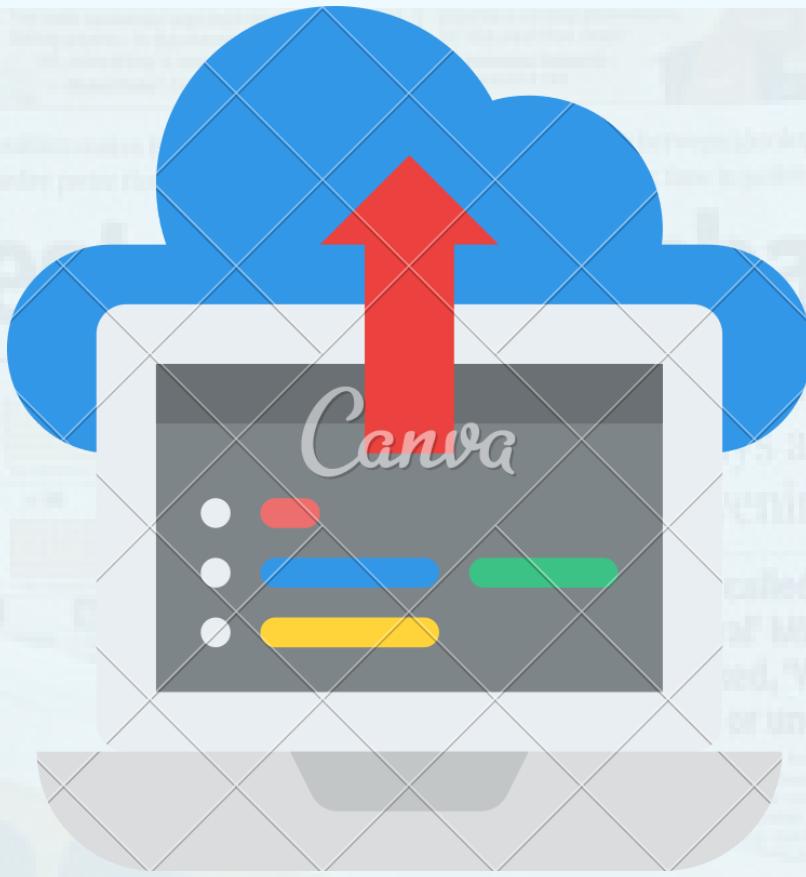
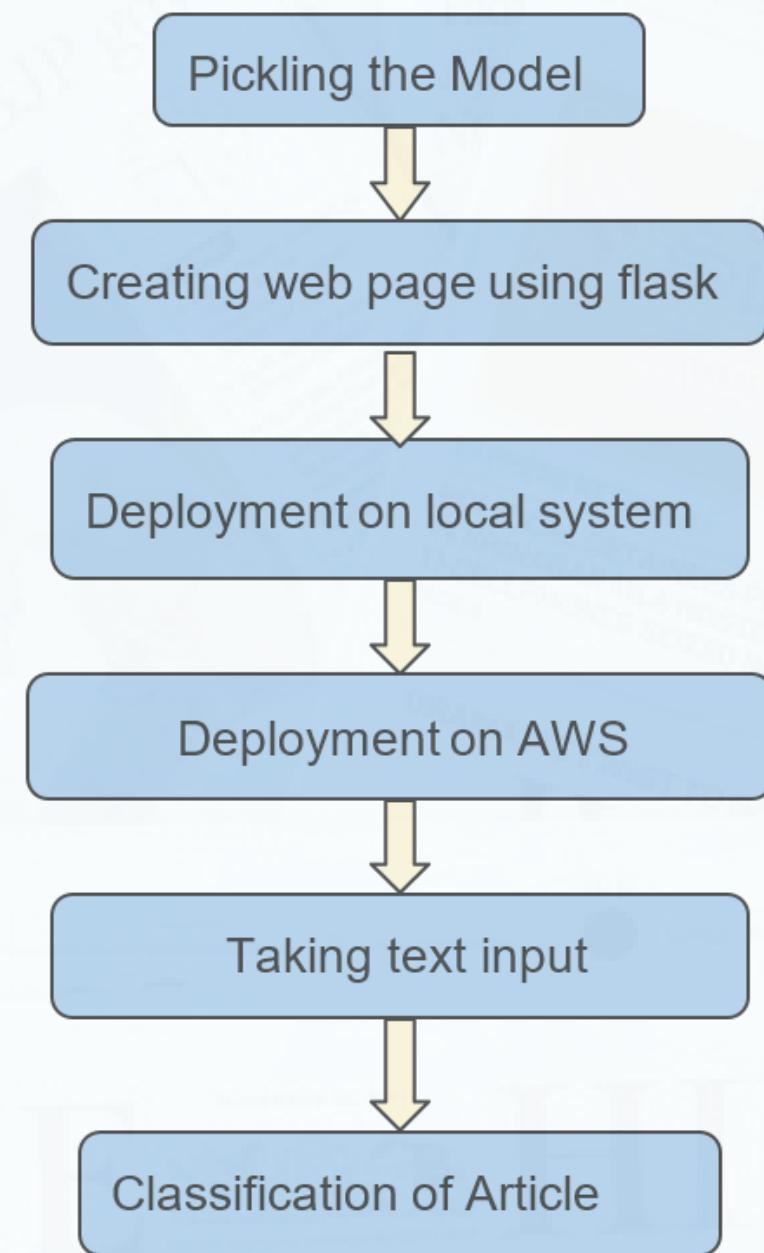
Model training & Evaluation

- Accuracy of Algorithms:
 - Support Vector Machine : **96.1393 %**
 - Random Forest : **95.3029%**
 - Gradient Boosting Classifier : **93.3754%**
 - Ada-Boost Classifier : **70.9753%**



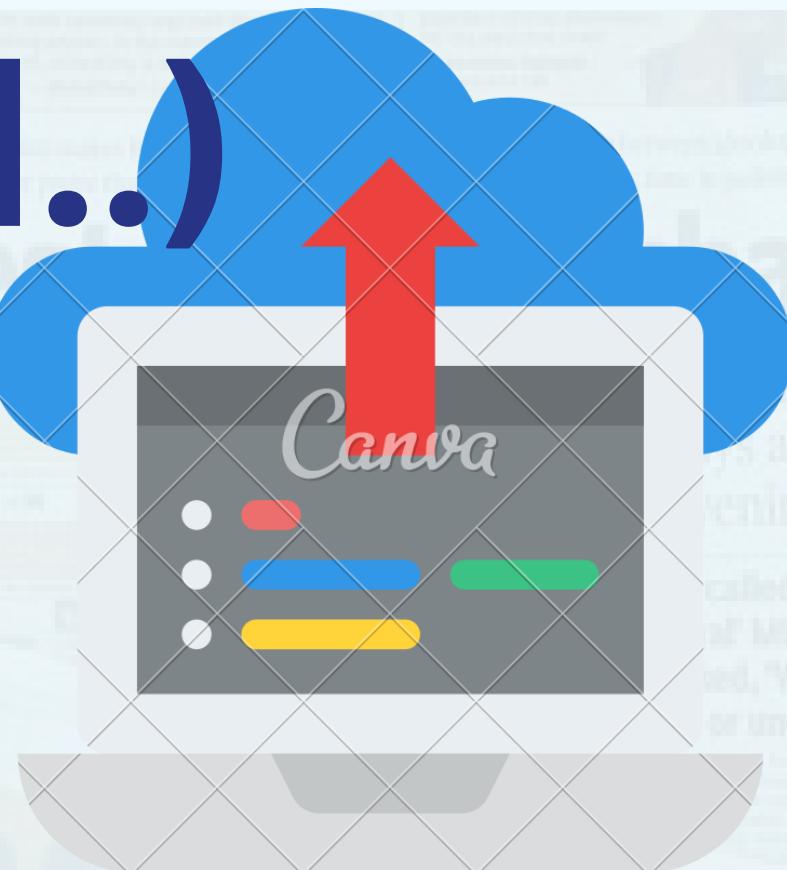
We observe that SVM is giving better performance than others.

Model Deployment



Model Deployment (Contd..)

- AWS cloud platform is used to deploy the selected Model:



```
AWS Services Search [Alt+S] Sydney Krishna Chandra Yadav

8 additional security updates can be applied with ESM Apps.
Learn more about enabling ESM Apps service at https://ubuntu.com/esm

Last login: Tue Sep  5 12:40:02 2023 from 13.239.158.4
ubuntu@ip-172-31-8-109:~$ ls
'Article Classification - AWS'  nltk_data
ubuntu@ip-172-31-8-109:~$ cd Article\ Classification -\ AWS/
ubuntu@ip-172-31-8-109:~/Article Classification - AWS$ ls
Article.py          'BBC News Test.csv'   TfIdf_Vec.pkl    model_svm.pkl  test1.py
'BBC News Sample Solution.csv' 'BBC News Train.csv' final-bbc-article-classification.ipynb templates
ubuntu@ip-172-31-8-109:~/Article Classification - AWS$ python3 Article.py
 * Serving Flask app 'Article'
 * Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on all addresses (0.0.0.0)
 * Running on http://127.0.0.1:8080
 * Running on http://172.31.8.109:8080
Press CTRL+C to quit
```

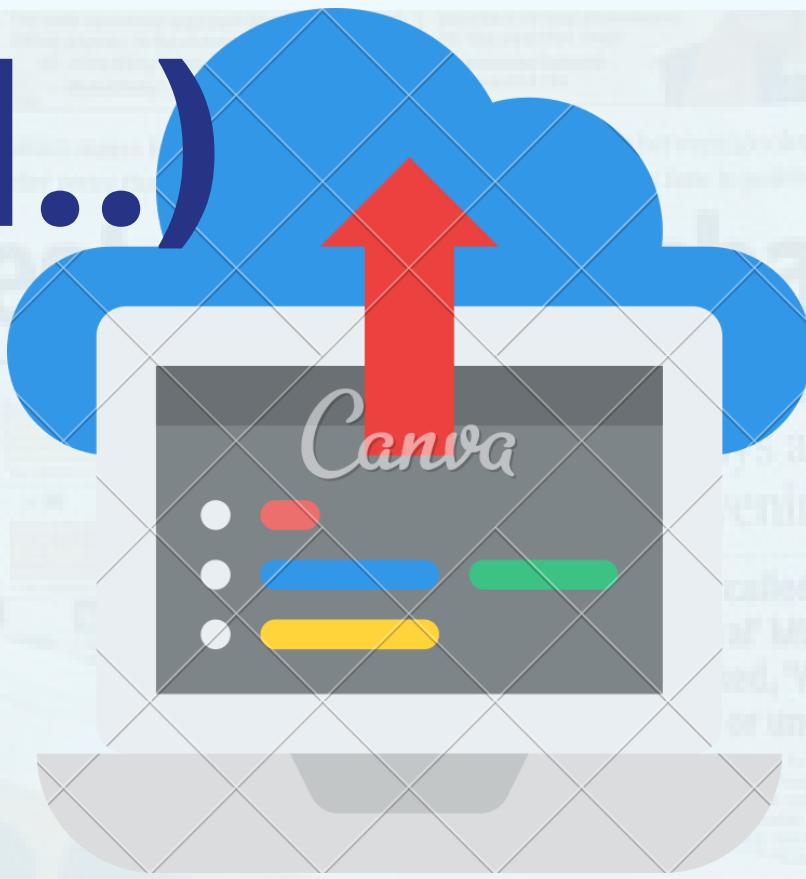
Model Deployment (Contd..)

- AWS cloud platform is used to deploy the selected Model:

News Article Classification

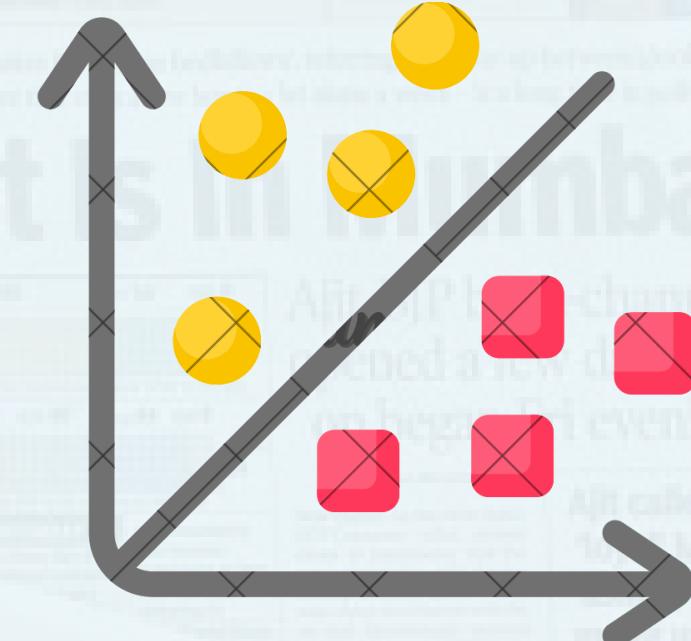
Enter an article:

Classify



Article Classification

- Here classification is going to perform on any random latest political news :



September 10, 2023 06:56 am | Updated 09:01 pm IST

THE HINDU BUREAU

COMMENTS SHARE READ LATER

Arrested Chandrababu Naidu in ACB court | Photo Credit: Special Arrangement

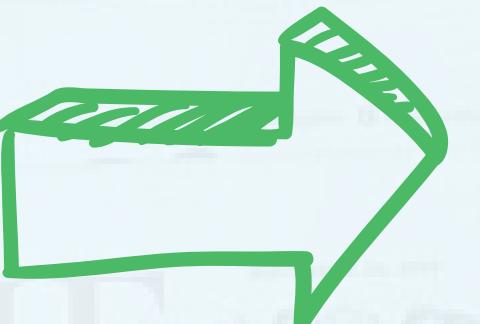
Telugu Desam Party president N. Chandrababu Naidu was sent to 14-day judicial custody till September 23 by an Anti-Corruption Bureau court in Vijayawada on September 10 in a case of alleged corruption.

Meanwhile, keeping the law the order issue in view, the authorities concerned imposed Section 144 of the Cr.P.C. across the State, and people were advised not to move in groups and carry any lethal weapons.

Earlier on September 10 morning, the TDP leader was produced in the court where Justice Hima Bindu heard the arguments by legal counsels from Government and the ones representing the TDP chief. Meanwhile, the NTR Commissionerate Police tightened security at the ACB Special Court. The security system was reviewed Police Commissioner Kanthi Rana Tata.

In the early hours of Sunday, Mr. Naidu was shifted to the SIT office after being taken to a hospital in Vijayawada for medical tests following a 10-hour-long interrogation, a police official said.

Also Read | [Chandrababu Naidu mired in corruption, alleges Andhra Pradesh IT Minister](#)



News Article Classification

Enter an article:

Telugu Desam Party president N. Chandrababu Naidu was sent to 14-day judicial custody till September 23 by an Anti-Corruption Bureau court in Vijayawada on September 10 in a case of alleged corruption.

Meanwhile, keeping the law the order issue in view, the authorities concerned imposed Section 144 of the Cr.P.C. across the State, and people were advised not to move in groups and carry any lethal weapons.

Classify

Classification Result:

Politics

Thank you



THE HINDU

on Sunday

Padnavis sworn in as C