**Milestone Report – 4**

**Capstone Project: News Article Classification**

* **Dayakar & Krupasindhu (31st Oct 2021)**

1. **Project Architecture Design Completed** (Attached .PPT document for reference)
2. **ML Approach Documentation completed** (Attached .PPT document for reference)
3. **Developed Python script for Data preparation – (Completed as part of Milestone 1)**
4. **Install python mongodb connectors for connecting with Pyspark – (Completed as part of Milestone 1 &2)**

**\* !pip3 install pymongo[srv]**

**\* !pip3 install Pyspark**

**\* !pip3 install mlflow**

**\* Load the data from mongodb**

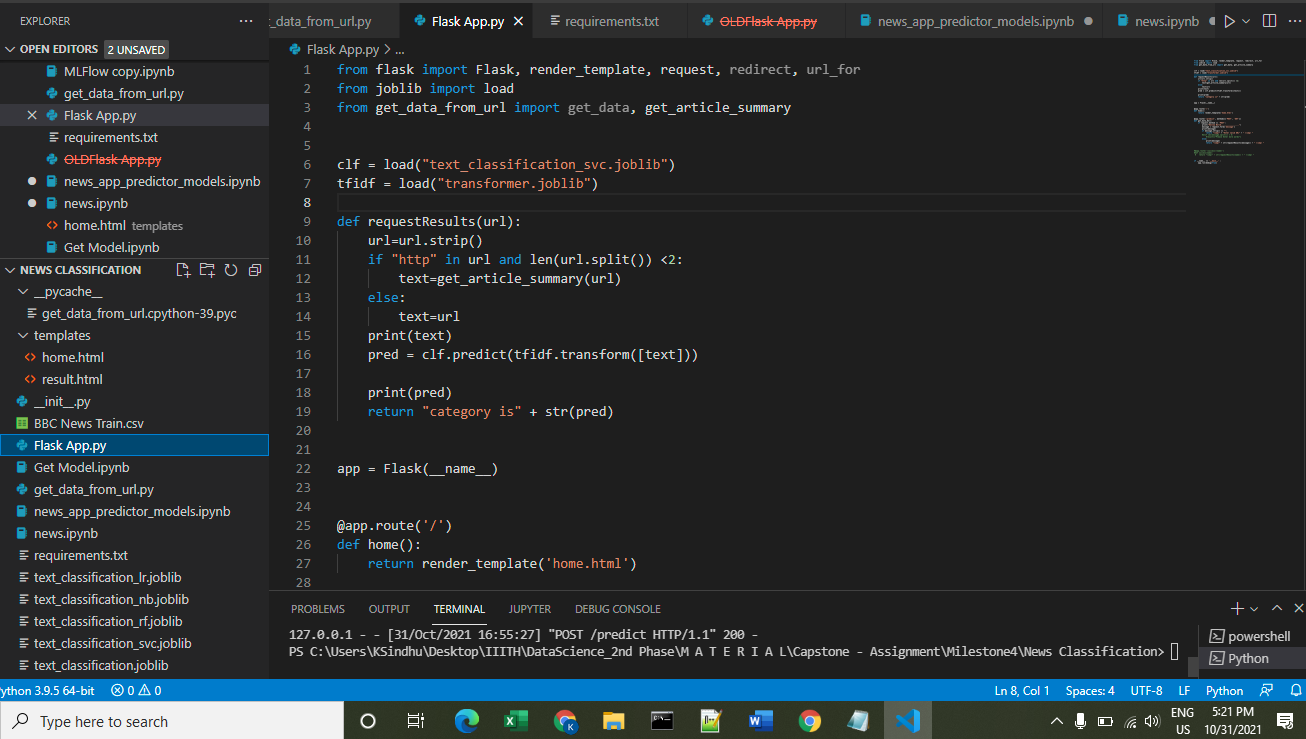
**\* Perform EDA to study the data**

**\* Select the relevant columns for modeling**

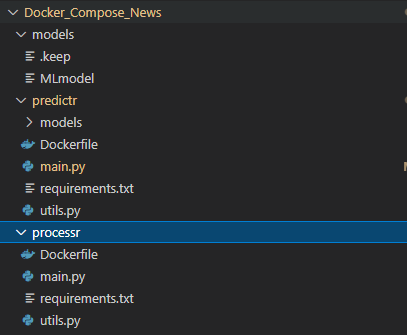
1. **Integrate Model with Model tracking APIs – (Completed as part of Milestone 1, 2 & 3)**

* **Split the data for train and test**
* **Model the data with different algorithms**
* **Evaluate the models and select the best performing and optimized model**
* **Track and log models with MLflow**
* **Register models with the Model Registry**
* **Describe models and make model version stage transitions**
* **Integrate registered models with production applications**
* **Search and discover models in the Model Registry**

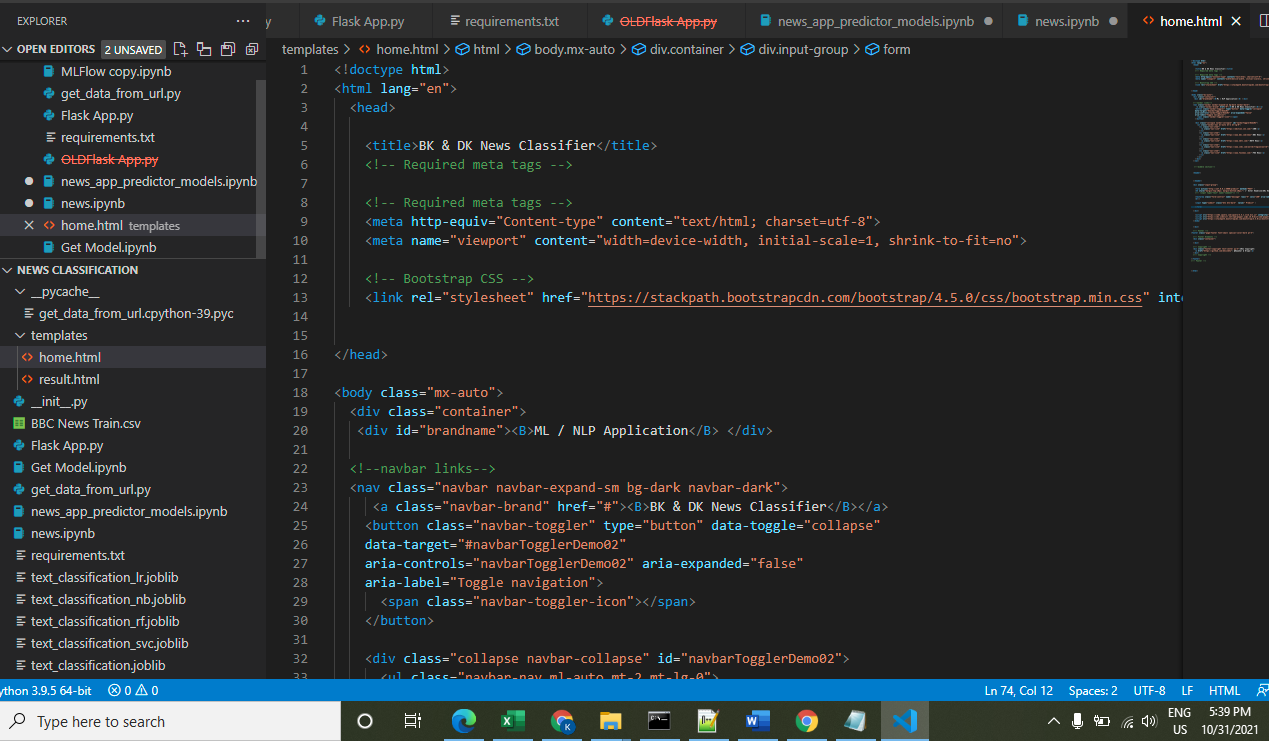
1. **MLFlow Libraries View – (Completed as part of Milestone 3)**
2. **Model View – (Completed as part of Milestone 3)**
3. **Model Accuracy View – (Completed as part of Milestone 3)**
4. **ML Registry & Pipeline – (Completed as part of Milestone 3)**
5. **Requirements.txt view – (Completed as part of Milestone 3)**
6. **ML Model view – (Completed as part of Milestone 3)**
7. **Conda.yaml view – (Completed as part of Milestone 3)**
8. **Deploy a ML Flow Model for real-time serving – (Completed as part of Milestone 3)**
9. **Expose model via model-prediction-service in the form of flask API.**



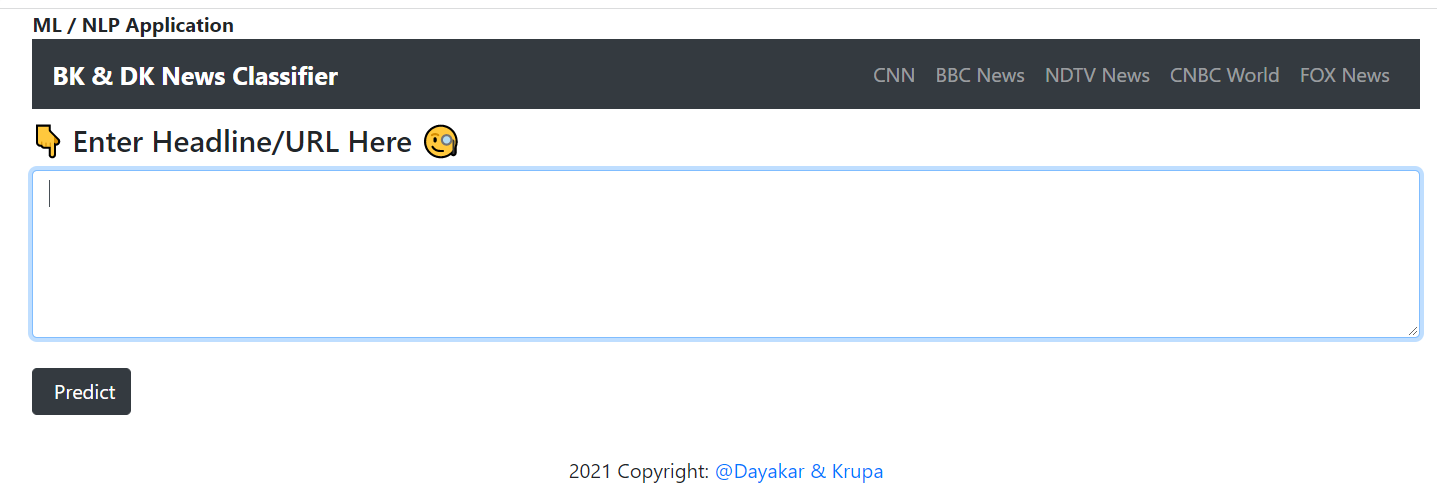
1. **Dockerized all the project by adding appropriate Dockerfile**



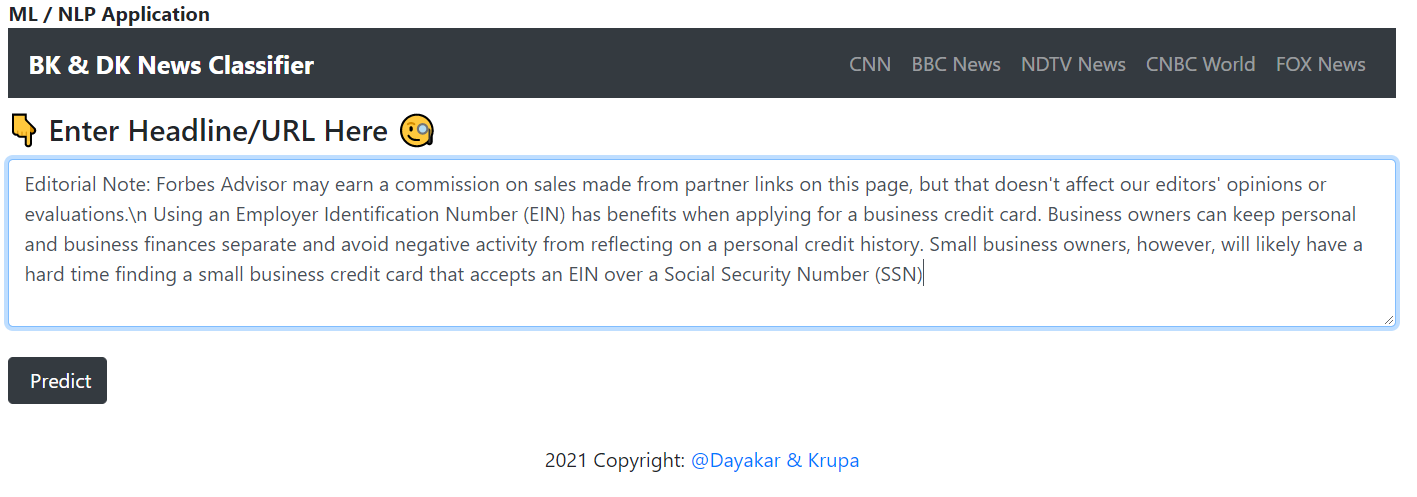
1. **Prepared the HTML page containing a form to take an article (Text or URL) as input and print the predicted category**



1. **Web Application View: It accepts either URL or New Headlines as input to predicts its category**



1. **Inputting News Headlines for prediction:**

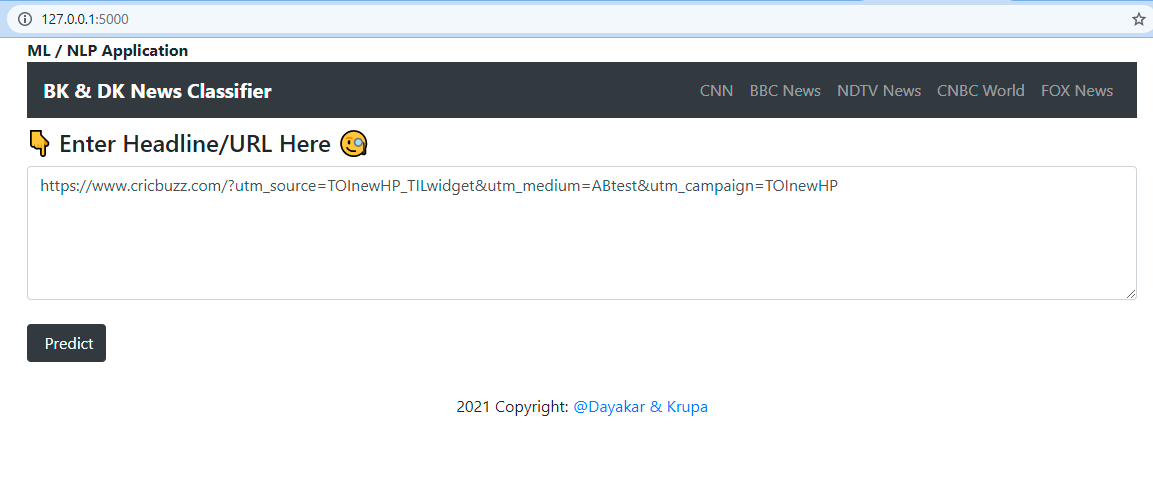


1. **Entered Headlines are predicted as ‘business’**

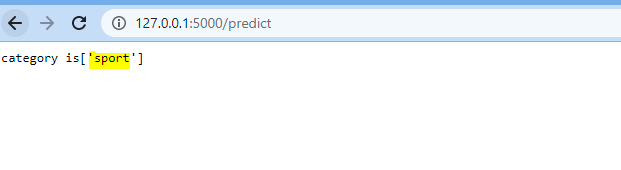
Predicted Category:



1. **Inputting URL as inputs for its category prediction:**



1. **Entered URL is predicted as “Sport”**



1. **GitHub Link / URL:**

<https://github.com/bksindhu/Capstone---News-articles-classifier/>

<https://github.com/DayakarKodirekka/Capstone---News-articles-classifier>

1. **Challenges:**

As limited data is rendering from RESTAPIs, the category/topic predictions are subject to the accuracy of the data available. Also, we used BBC News data to train the model.