**Milestone Report – 3**

**Capstone Project: News Article Classification**

* **Dayakar & Krupasindhu (24th 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**
4. **Install python mongodb connectors for connecting with Pyspark**

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

**\* !pip3 install Pyspark**

**\* !pip3 install mlflow**

**\* Load the data from mongodb**

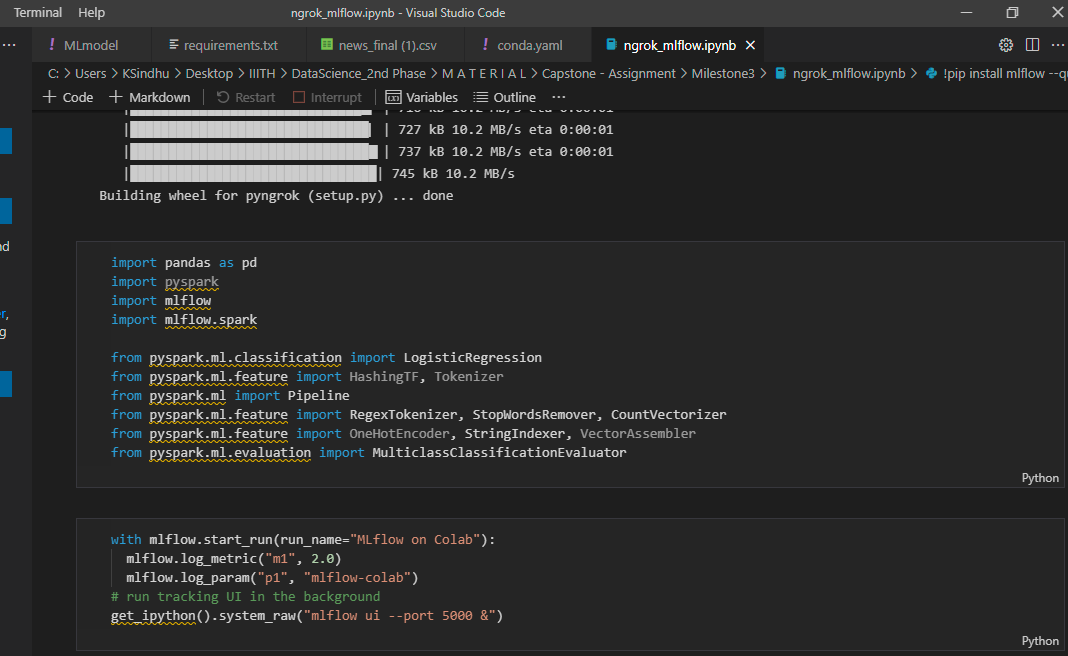
**\* Perform EDA to study the data**

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

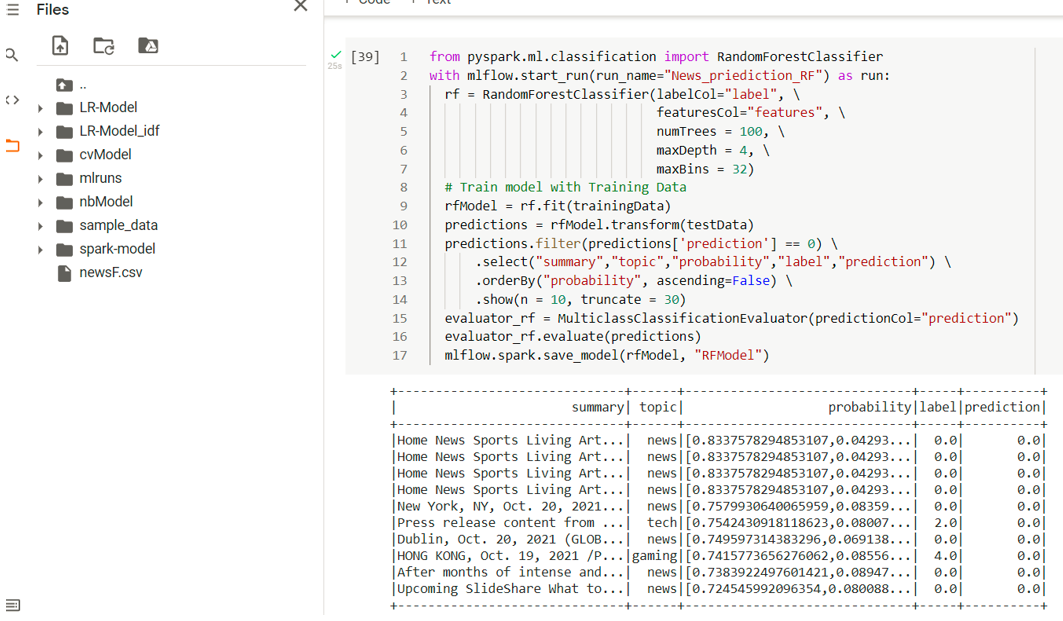
1. **Integrate Model with Model tracking APIs**

* **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**

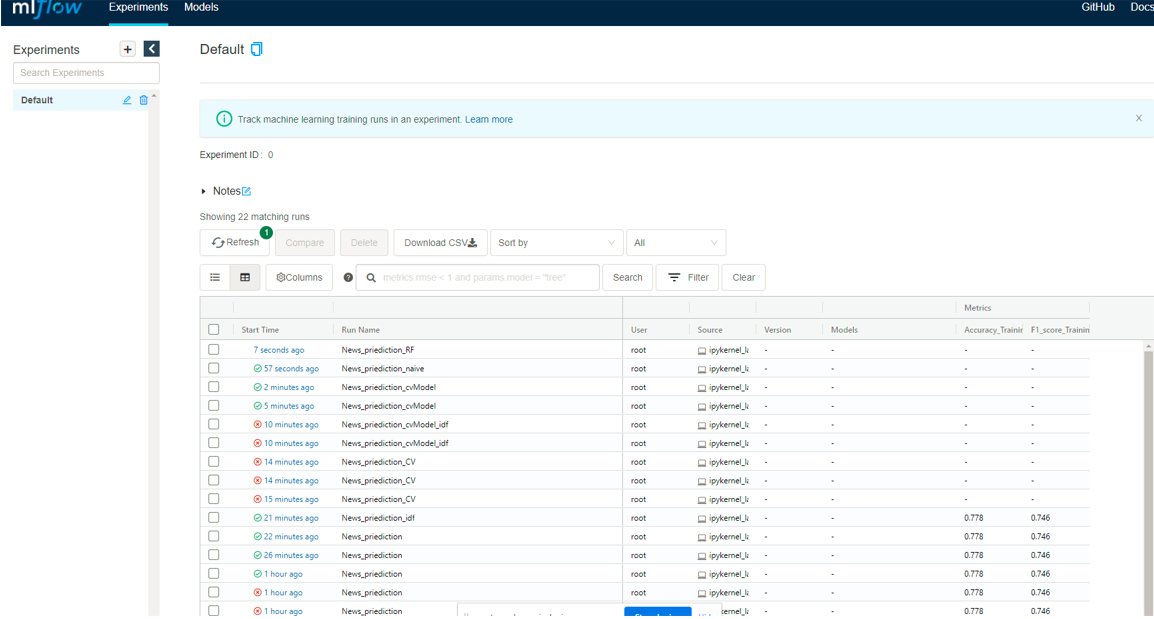
1. **MLFlow Libraries View:**



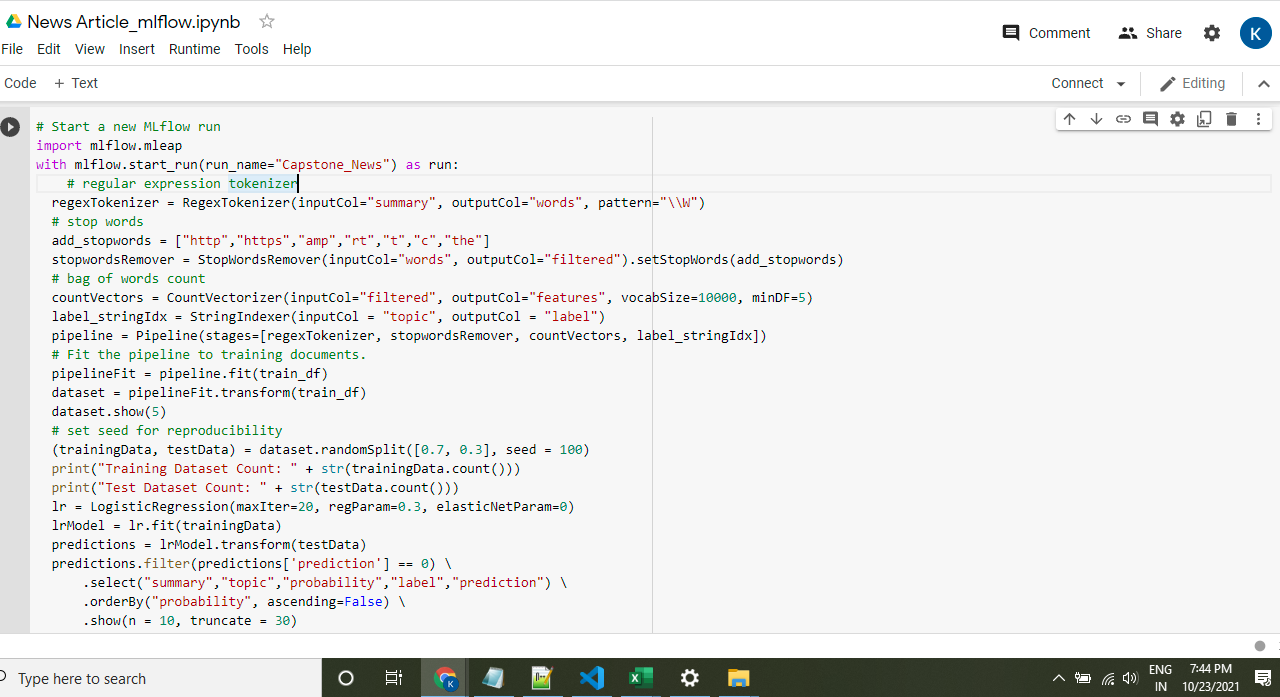
1. **Model View:**



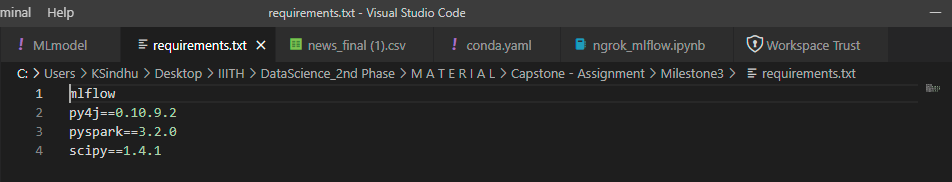
1. **Model Accuracy View:**



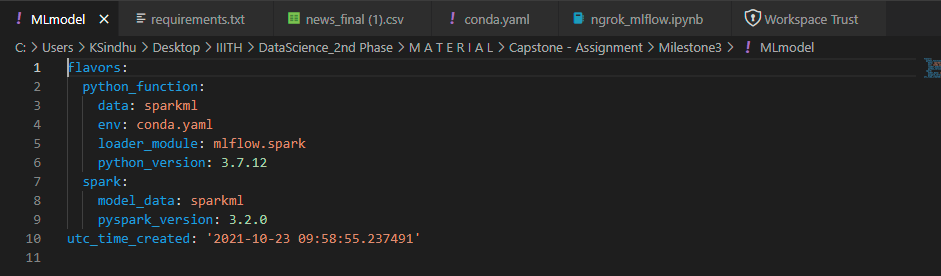
1. **ML Registry & Pipeline:**



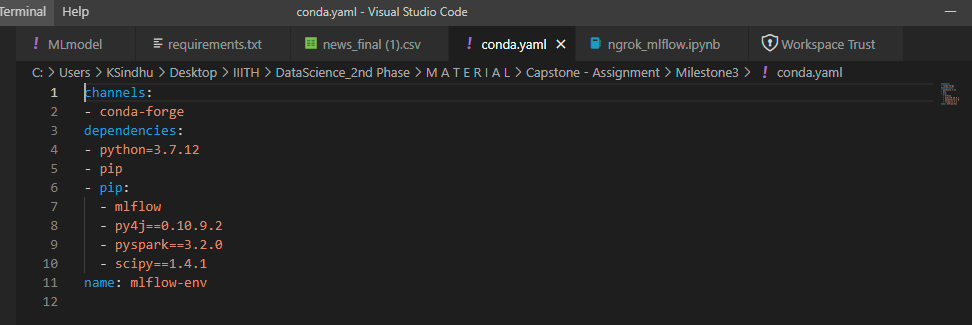
1. **Requirements.txt view:**



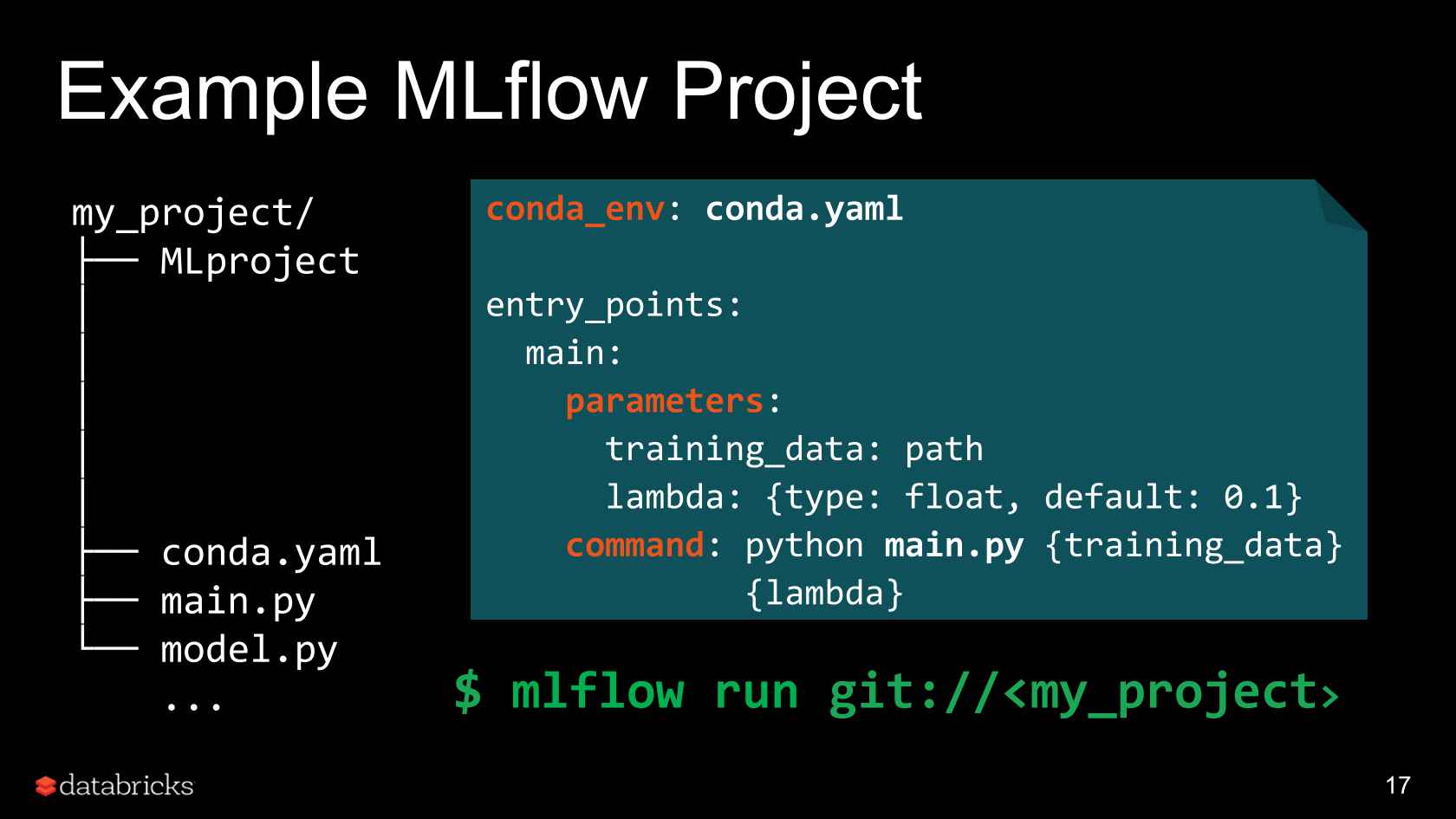
1. **ML Model view:**



1. **Conda.yaml view:**



1. **Deploy a ML Flow Model for real-time serving**



1. **GitHub Link / URL:**

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

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

1. **Challenges:**

ML results are difficult to reproduce