**Milestone Report : Week 4**

1. **Environment Details: -**

* **OS : Ubuntu 18.05 bionic**
* **Data Sources: -**
  + **MYSQL DB**
* **Data Processing: - PySpark, Python 3**
* **ML Modelling :- PySpark**
* **Tokenizer : tfidfvectorizer**
* **Model :- naive\_bayes.MultinomialNB , LogisticRegression**
* **Containerization : Docker**
* **Webpage : Flask and Html CSS**

1. **What goes in an input.**

* **All individual component such as Frontend , model training, model prediction are containerized as docker .**

1. **How the input is being processed**

* **Using docker compose and common network , the docker containers are setup in same network.**

1. **What comes out as an output**

* **News Classifier Webpage**
* **Containerized Component – Predictor, Trainer , front end , kafka , kafka-connect, mongodb**

**GitHub link - https://github.com/behers10/CapstoneProject.git**

**Output Architecture Diagram.**

Diagram

Description automatically generated