

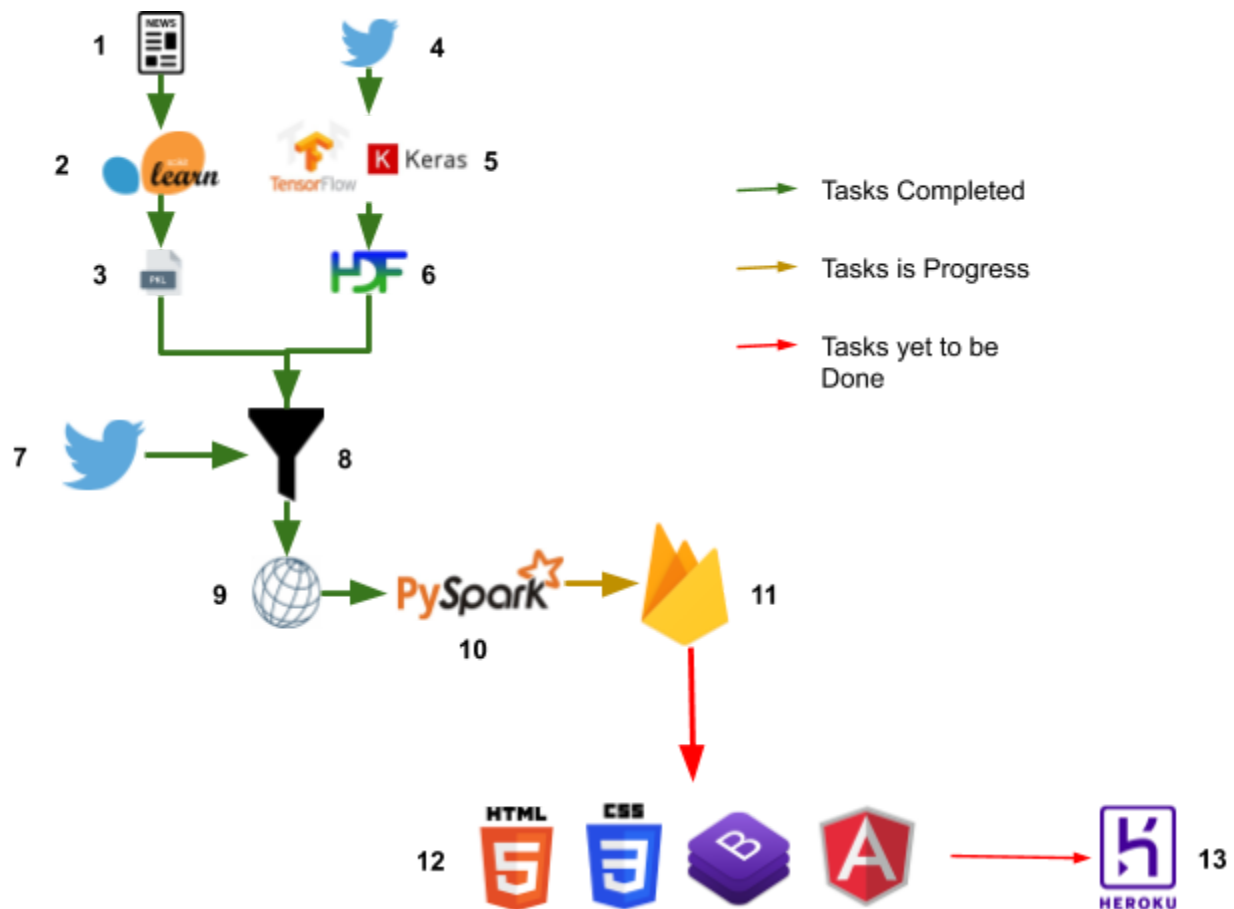
Tweeflix - Using Twitter To Detect And Investigate Disease Outbreaks

Kavin Kumar Arumugam and Shivani Sankar



1. What types of queries do you plan to develop?


We are planning to apply queries on each rdd from twitter streaming and push it to the firebase databases but we haven't planned properly yet.



2. What tools do you plan to use?








Tasks:

1. **News Data** - 
 - a. We have taken news data from.
<https://archive.ics.uci.edu/ml/datasets/News+Aggregator>
 - b. We also did some preprocessing on the News Data.
2. **Scikit Learn Library** - 

Used Naive Bayes Classifier to train the model from News Data.
3. **Content Classification Model** - 

Saved the trained model as .pkl
4. **Twitter Historical Data** - 
 - a. We have taken Twitter Sentiment Data from
 - b. We also did some preprocessing on Twitter Sentiment Data
5. **Tensorflow + Keras** - 

We used LSTM from Tensorflow and Keras to train the Twitter Historical Data.
6. **Sentiment Analysis Model** - 

Saved the trained data as .H5.
7. **Twitter Streaming** - 
8. **Filter Data from Twitter Streaming Data** - 
 - a. Filter data based on the content classification model.
 - b. Added sentiments to the twitter data using Sentiment Analysis Model.
 - c. Streamed the data to TCP.
9. **TCP** - 
10. **PySpark Streaming** - 

Used PySpark Streaming to stream the data from TCP.
11. **Firebase RealTime Database** -

We are planning to push the data to Firebase RealTime Database.
12. **Front End Using - HTML5 + CSS3 + Bootstrap + AngularJS** -

Create an HTML page from the Firebase Database.
13. **Heroku** -

We are also planning to deploy it to Heroku.

3. Division of work and timeline for completion:

So far we have divided the work equally and we planned to complete

1. Queries by tomorrow, and
2. Front end by the end of this week.

4. Plan for testing, debugging, and improvements:

1. Testing: We have planned to apply some UI testing on the front end.
2. Debugging and Improvements:: We haven't planned anything about it.