Ideas for Capstone Project – III

Please find below list of Ideas and their details, I found interesting.

1. NLP | Text Classification | Coronavirus tweets: Sentiment Analysis

**Source:** <https://www.kaggle.com/datatattle/covid-19-nlp-text-classification>

Rows: 45K

Columns: 6

The problem is based on Text Classification of Twitter data. Data has been sourced from Kaggle. The tweets in the dataset have been pulled from Twitter and manual tagging has been done to the labels. We need to analyze the tweets from different users and classify the sentiment behind the tweets on corona virus perspective. Classification has to be done in multiple classes as below:

* Extremely positive
* Positive
* Neutral
* Negative
* Extremely Negative

1. NLP | Text Classification | Categorize News Dataset

**Source:** <https://www.kaggle.com/rmisra/news-category-dataset>

Rows: 202K

Columns: 6

The problem is based on Text Classification of the News articles. Data has been sourced from Kaggle. Dataset contains news headlines and short description about each headline from the year 2012 to 2018. Each news headline has a corresponding category. The model trained on this dataset could be used to label news articles for different categories or to identify the type of language used in different news articles. News has to be categorized into multiple classes, some of the examples of labels are as below:

* Politics
* Sports
* Entertainment
* Business
* Crime
* <https://www.kaggle.com/semram/machinelearning-group21>
* <https://www.kaggle.com/derinrobert/newsclassification-using-lstm>

1. NLP | Text Classification | News Validation: Fake or Real News

**Source:** <https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset>

Rows: 44.9 K

Columns: 4

The problem is based on Text Classification of the News articles. Data has been sourced from Kaggle. Two different datasets are available, separate dataset for Fake News (23.5k Rows) and Real News (21.4 k Rows). Dataset contains title, text and category of news articles as dependent features through which we need to classify either the news is Fake or Real.