

# Prediction of Cryptocurrency Behavior Using Sentiment Analysis on Twitter and News Data

Aram Saponjyan, Charles E Fryer, Farhad Sedaghati, Mandeep Mundy

## Summary

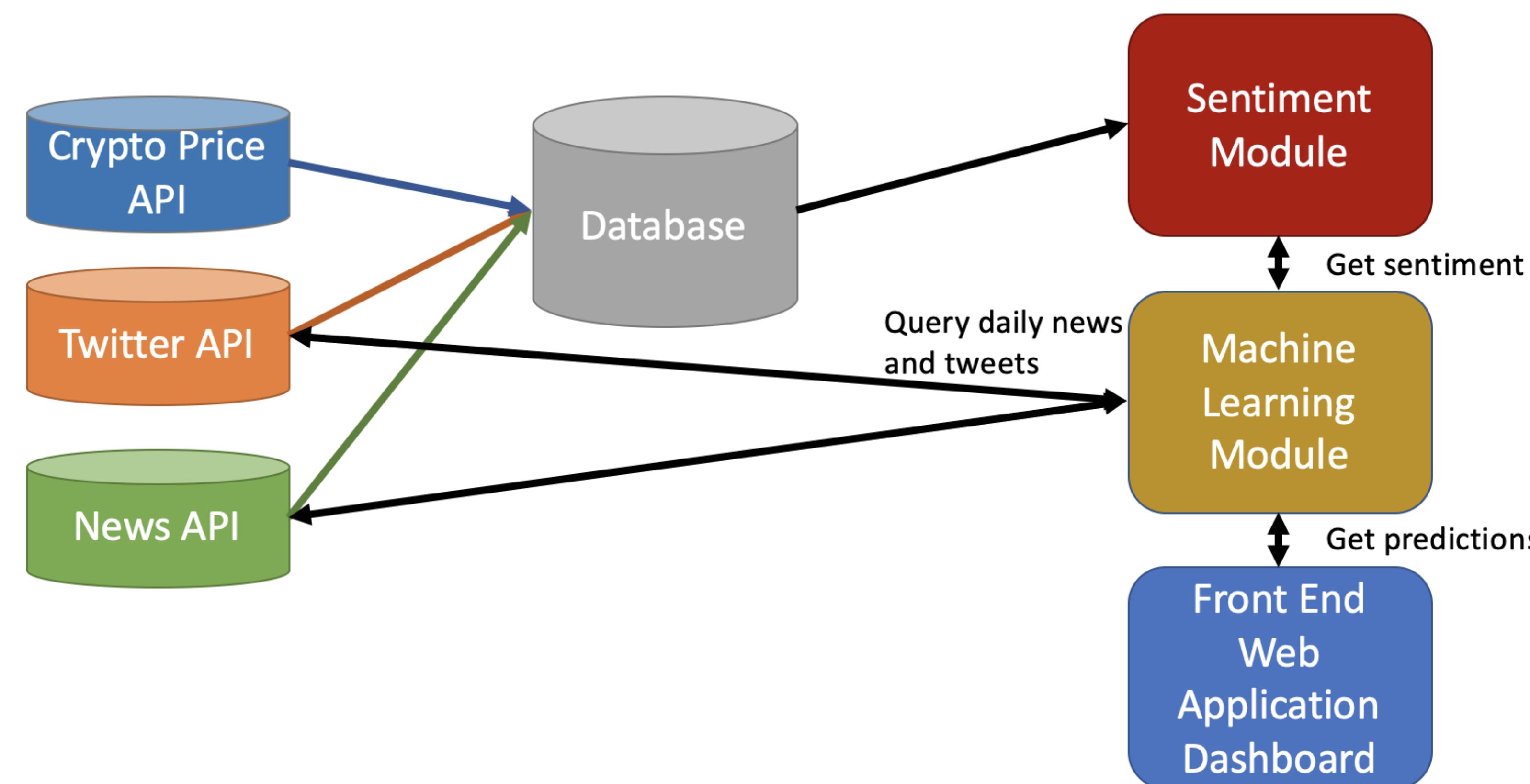
We study the impacts of business news and Tweet data on cryptocurrency price with sentiment analysis. Cryptocurrency prices are volatile and subject to significant changes. Expert traders use patterns that exist in the real-world market behavior to predict if price will go up or down. Various parameters can affect the price of a given stock. A major factor shown to have a strong correlation with the price movement is sentiment in business news and Twitter data.

## Motivation

This application seeks to provide users with an interactive platform that enables them to visually explore crypto trends based on news and twitter sentiment. Users will also be able to day trade with recommendations from the model on individual coins.

## Approach

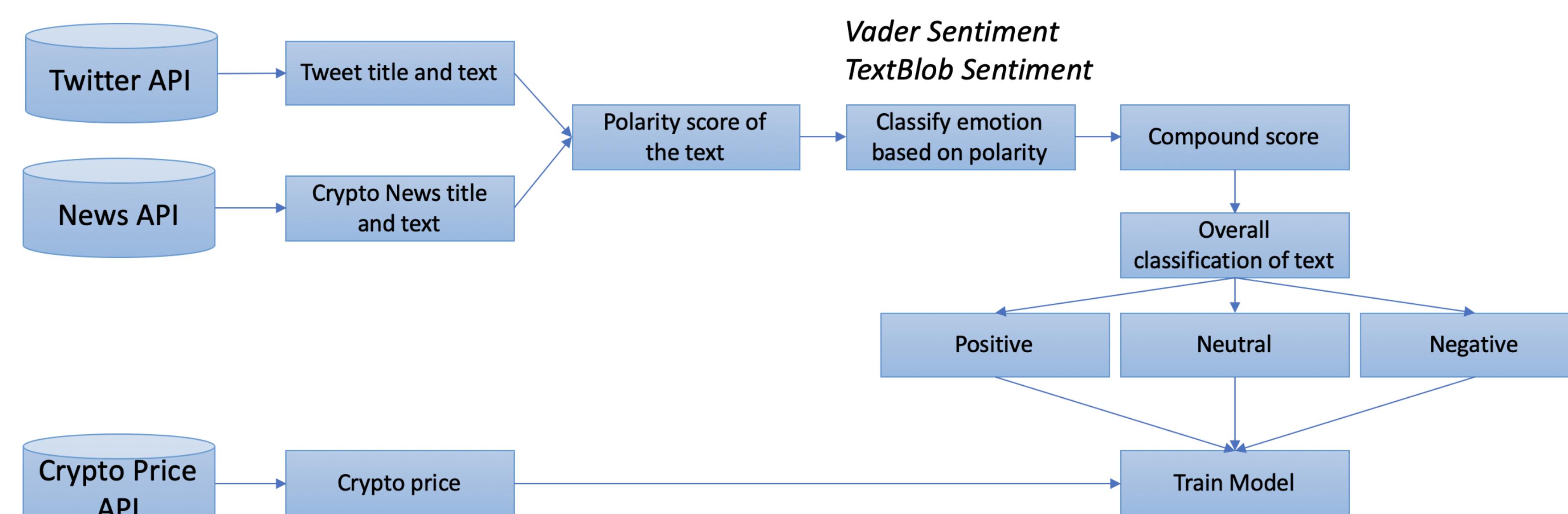
The trading recommendations and visualizations are powered by a machine learning model that considers multiple social media and news outlet sentiment signals.



## Sentiment + Machine Learning

Sentiment is determined with VADER and TextBlob on the full tweet and news article samples. The outputs then train on the LDA model, buy / sell labels are restricted to +/- 3 percent daily price movements. The web application is hosted in AWS and leverages the CoinGecko API to retrieve cryptocurrency data, the Twitter API for social media data, and the News API to retrieve data from reputable cryptocurrency news outlets.

## Application Architecture



## Ensemble The Sentiment

We approach the model features with an ensemble point of view. Using the VADER compound score sentiment for twitter data and TextBlob for news sentiment as 2 sets of independent variables which are input into the LDA model. This final data set is what the model trains on.

## Data Combination

Current research technique use either a single social media or news outlet to perform sentiment analysis. Our approach built upon the techniques used in these studies by utilizing multiple social media and financial news sentiment sources into an overall predictive model.

## Linear Discriminant Model

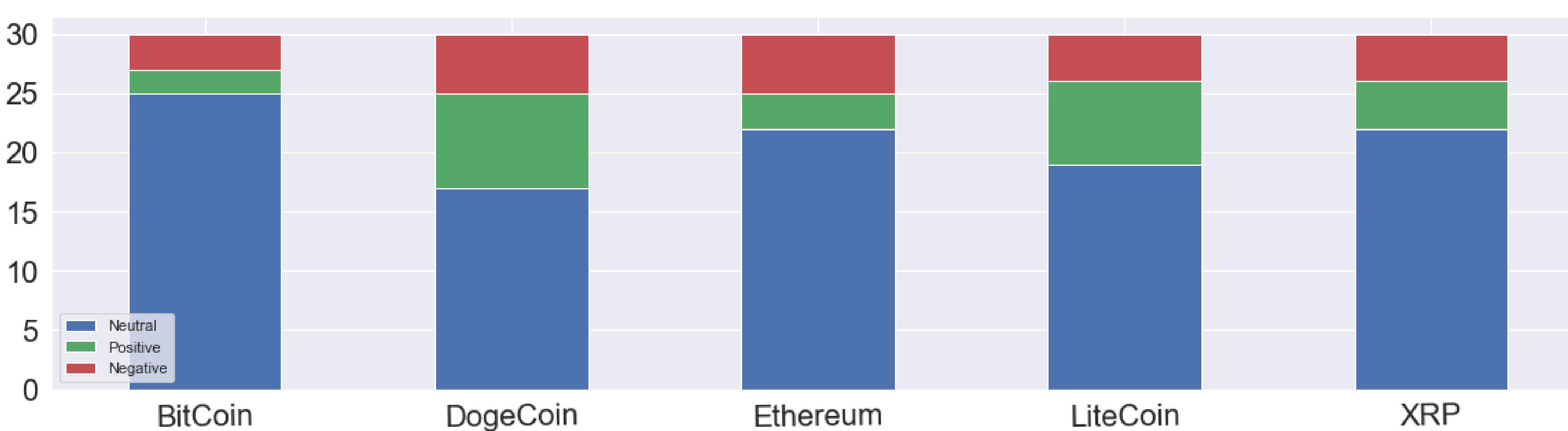
The goal of LDA is to project the features in higher dimensional space onto a lower-dimensional space in order to reduce the dimensionality and resource costs. It reduces the features in a particular dataset while retaining most of the data.

## Application Load Balance

We use AWS Elastic Load Balancing and EC2 Auto Scaling which will automatically distribute our incoming traffic across multiple EC2 instances.

## Data

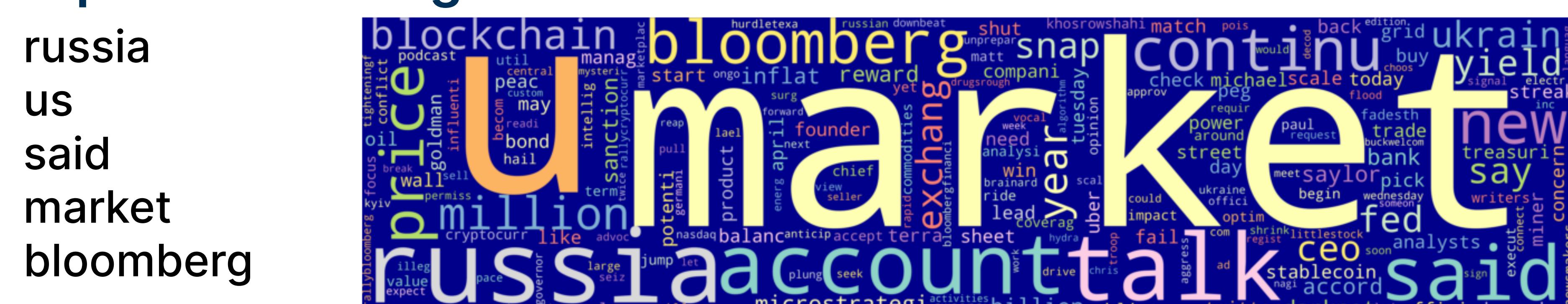
### Training Data Sentiment Label Distribution by Coin



### Top 5 Words Positive Sentiment Word Cloud



### Top 5 Words Negative Sentiment Word Cloud



## Evaluation Method

### Model Accuracy

We evaluate the model performance with the crypto price change prediction RMSE and Accuracy using k-fold validation. With the Twitter and News API, we query historical data and evaluate model prediction for that time frame.

### Optimization with Correlation

We investigated the effect of number of days lag in the model. N-day lag day indicate news and twitter are correlated with returns on n-day later. We studied the 1 day to 1 week lag and obtained the RMSE and accuracy for all 5 coins with the k-fold cross validation technique.

### UI and Model Usability Evaluation

User feedback was another method utilized to evaluate the application usage and value. We created a google survey to gauge user usage and opinions.

## Results

		RMSE	Accuracy
LinearDiscriminantAnalysis	tweets	0.309	0.691
	news	0.353	0.663
	concatenated	0.371	0.645
KNeighborsClassifier	tweets	0.415	0.641
	news	0.408	0.628
	concatenated	0.429	0.635
RandomForestClassifier	tweets	0.467	0.643
	news	0.375	0.644
	concatenated	0.361	0.687
GaussianNB	tweets	0.404	0.64
	news	0.671	0.489
	concatenated	0.753	0.443

Current stock research accuracy varies between 55 and 80 percent. The combination of sentiment signals from both Twitter and news outlets, for cryptocurrency prediction display similar performance to modern stock research.

One day lag performed best, indicating price change in the next day are more correlated to news and Twitter's data sentiment scores.

