Topic 1
Cryptocurrency price prediction using sentiment analysis

This would predict cryptocurrency prices using sentiment analysis. As cryptocurrency popularity is increasing in the modern age and the money flow is increasing, this is making cryptocurrencies more volatile, and the patterns are changing. Some of the problems faced are as unlike the stock market cryptocurrencies are dependent on factors such as its technological progress and internal competition etc. I plan to get data from news agencies about specific tokens, also data of all the price changes in crypto from 2012 to predict the future prices and data from Twitter's API. As in this modern age, social media plays a big role in influencing the decision making of people, so my research would be mainly on google trends and raw tweets taken from Twitter. Some of the many places this data can be found is on Cryptocompare API: XEM and IOT historical prices in hour frequency, Pytrends API: Google News search frequency of the phrase “cryptocurrency”, Scraping redditmetrics.com: Subreddit “CryptoCurrency,” “Nem,” and “Iota” subscription growth. The tweets would be analyzed to create a sentiment score and compared to the price changes to that day to determine if a relationship between Twitter sentiment and cryptocurrency price changes could be determined.

Source 1
Abraham, Jethin; Higdon, Daniel; Nelson, John; and Ibarra, Juan (2018) "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis," SMU Data Science Review: Vol. 1 : No. 3 , Article 1.

https://scholar.smu.edu/cgi/viewcontent.cgi?article=1039&context=datasciencereview

This paper uses sentiment analysis to predict the prices of the two biggest cryptocurrencies to date(Bitcoin and Ethereum). This paper determines that rather than tweet sentiment, the volume of the tweets is the predictor of price direction. It uses a linear model and takes data from tweets and google data trends to accurately predict prices. The conclusion of this paper is that sentiment of tweets is not a reliable indicator for the price predictions, especially in cases of high variance; however, google trends and tweet volumes were highly correlated with the prices. Multiple linear regression was selected as the modeling algorithm of choice due to strong correlation metrics. In my initial pitch, I was planning on leaning towards only the sentiment of the tweets; however, the volume and the trends of the data also makes a huge difference in the prediction of prices.

Source 2

Valencia F, Gómez-Espinosa A, Valdés-Aguirre B. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy*. 2019; 21(6):589. https://doi.org/10.3390/e21060589

https://www.mdpi.com/1099-4300/21/6/589/htm

This paper uses both sentiment analysis and machine learning to predict the prices of 4 crypto currencies(Ripple, Bitcoin, Litecoin and Ethereum). The data was collected from the top 65 crypto exchanges in the world. The data for the sentiment analysis was collected from the raw tweets from the Twitter API. As per the paper of all the data collected through the tweets, 63% of it mentioned bitcoin. The conclusion for this paper is that the results were different for all cryptocurrencies and varied more in the smaller tokens such as litecoin. One hypothesis could be because litecoin has a smaller community which would mean it has fewer tweets and makes the results different. This paper proved that it is possible to predict the cryptocurrencies prices using these data sets. This is the reason I was only planning on using the data for the top 3 cryptocurrencies as because the price varies more and differently for the smaller coins which have less following. As seen in these papers, every model selects a specific amount of token/coins to be examined as a single model cannot determine price predictions for each coin because of factors like lower number of followings and variation in tweets.

Source 3

Stenqvist, E., & Lönnö, J. (2017). Predicting Bitcoin price fluctuation with Twitter sentiment analysis (Dissertation). Retrieved from
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-209191
https://www.diva-portal.org/smash/get/diva2:1110776/FULLTEXT01.pdf

This paper is based on the price prediction of bitcoin. It mentions that the prices of bitcoin are dependent on socially constructed opinions around it. The data collected for this was from coin desk publicly available API and Twitter API. It was determined that even 2 million tweets that were retrieved might be a low threshold to accurately determine the prices, however the model works most efficiently in determining the price of the 1 hour interval. This would vary between the intervals during the day(morning, evening, night). One aspect I noticed was that in almost every paper, including this one, the software used for sentiment analysis is VADAR which is used to determine if a tweet is positive, negative or neutral. Another aspect to mention is that this model had a 83% accuracy rate which could be increased if a better classifier is used which would identify financial and cryptocurrency terms and yield a more representative sentiment. Almost all papers use data from Twitter as its now widely known as a place to get advices/tip offs of buying and selling cryptocurrencies and stocks.

Source 4

Lamon, C., Nielsen, E., & Redondo, E. (2017). Cryptocurrency price prediction using news and social media sentiment. *SMU Data Sci. Rev*, *1*(3), 1-22.

http://cs229.stanford.edu/proj2017/final-reports/5237280.pdf

This paper analyzes the ability of daily news and social media data to predict the price of three cryptocurrencies(litecoin, ethereum and bitcoin). This prediction is done differently than a standard sentiment analysis.This approach was accomplished differently as it labeled data based on the price changes one day in the future rather than positive or negative sentiment. The data for this project was acquired from kaggle and using web scraping about 3600 news articles relating to cryptocurrencies was taken. As tokens have varied properties, different classifiers were used for different tokens. This model was tested with four classifiers(logistic regression,Linear SVM, Multinomial NB, and Bernoulli NB) to determine if there would be an improvement in the ability to detect large percent changes in prices. As a result the results were more accurate and as seen in most models made on this topic the results are efficient when there is an increasing trend rather than when the price falls. This is a different approach which I have not thought about before but keep these aspects in mind for my own model.

Source 5

Bharathi, S., & Geetha, A. (2017). Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems*, *10*(3), 146-154.

http://www.inass.sakura.ne.jp/inass/2017/2017063016.pdf

This model predicts price changes in the stock market. This model works on both to compare the results of sentiment analysis and technical analysis. The data for this model is taken from ARBK from Amman Stock Exchange(ASE). The Oracle database of Amman Stock Exchange(ASE) contains the prices of the 230 companies listed in the exchange from the year 2000. Moving average method is used as stock level indicators. Scaling system is used to determine sentiment for the words having positive/negative/neutral sentiment. The results for this model show significant improvement when sentiment analysis is used rather than only relying on indicators. More research will be done by me on the stock market analysis using sentiment analysis to see the differences of using sentiment analysis in the cryptocurrency and stock market. Papers which have used both indicators and sentiment analysis prove that the model is more accurate when both of these methods are used.

Source 6

Huang, X., Zhang, W., Huang, Y., Tang, X., Zhang, M., Surbiryala, J., ... & Zhang, J. (2021). LSTM Based Sentiment Analysis for Cryptocurrency Prediction. *arXiv preprint arXiv:2103.14804*

https://arxiv.org/pdf/2103.14804.pdf

This paper focuses on recurrent neural networks with long short-term memory(LSTM) by utilizing the sentiment analysis of social media to predict the real-time price movement of the digital currency. The data collected for this model is from a large-scale Weibo corpus from crawling Chinese microblogs on Sina-Weibo with the cryptocurrency keyword and is mainly for three cryptocurrencies(Bitcoin, Ethereum, and XPR). As the data used in the project is taken from Chinese entities and not from usual sites I have been researching on I thought the result would be different however the approach for the sentiment analysis is similar as the data is labeled into positive,negative or neutral sections and the results also show a similar efficiency rate which is close to 80%. I plan on using the same methods in my model, however, taking data from US-based social media sites.

Source 7

Wang, Y., & Chen, R. (2020, January). Cryptocurrency price prediction based on multiple market sentiment. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

https://scholarspace.manoa.hawaii.edu/bitstream/10125/63875/0109.pdf

This paper has a price prediction model of cryptocurrencies using a variety of machine learning and deep learning algorithms. In addition to this, using social media comment data can significantly improve the accuracy of the forecast. The data collected for this model was from the API of the cryptocurrency exchanges and from the relevant user comment sections using a web crawler. The data collected is for 6 coins(Bitcoin, Ethereum , Tether USD , EOS, Ripple, litecoin). A variety of machine learning algorithms are used to predict the price of cryptocurrencies, including BP neural network, RBF, support vector machine, and deep learning algorithm CNN, LSTM. The price prediction model is evaluated by comparing the accuracy of the five algorithms with the evaluation indicators. Although I am not planning on using all of these different machine learning algorithms, they still give a different perspective. I could evaluate and see which one would work better for my algorithm. The results show that the LSTM model has the best prediction effect, and the addition of the comments data sentiment can significantly improve the accuracy of prediction.

Pitch 2

This model would detect fires using the Keras Deep Learning library. As seen recently around the world in places such as the Amazon rainforest and a prominent part of Australia, wildfires are increasing in this era. These disasters are damaging to the ecosystem like damaging habitat and releasing carbon dioxide. This project can be built using k-means clustering. This model would be able to identify any forest fires hotspots along with the intensity of the fire at that particular spot which would result in either the model detecting if it's a wildfire or not. There is another way of making this using the neural network MobileNet-V2 or CNN which is more efficient, and I will be researching more on this. There is a data set compiled with over 1300 images that would be used to detect wildfires although datasets of higher magnitude are available if 1300 images would not be efficient. The data for this project can be found at https://drive.google.com/file/d/11KBgD_W2yOxhJnUMiyBkBzXDPXhVmvCt/view.

Source 1

Govil K, Welch ML, Ball JT, Pennypacker CR. Preliminary Results from a Wildfire Detection System Using Deep Learning on Remote Camera Images. *Remote Sensing*. 2020; 12(1):166. https://doi.org/10.3390/rs12010166

https://www.mdpi.com/2072-4292/12/1/166/htm

The paper uses a model in which the system relies on machine learning-based image recognition software and a cloud-based workflow capable of scanning hundreds of cameras every minute. Right now, the data is taken from ground cameras, but the researchers are working on developing a satellite-based system. The model is trained with the InceptionV3 architecture on our own dataset of images from the cameras. This system is still in its prototype phase and can produce false negatives or positives. For the model to be efficient, it is also important to get data of different types such as data of smoke and non-smoke related images. I planned on using data from an already built data set and training the model. The paper examined both pre-trained models vs training a model from scratch and found no difference between both.

Source 2

Zhang, Q., Xu, J., Xu, L., & Guo, H. (2016, January). Deep convolutional neural networks for forest fire detection. In *Proceedings of the 2016 international forum on management, education and information technology application. Atlantis Press*.

https://www.atlantis-press.com/article/25850411.pdf

This paper used a binary classification method to detect if there is a fire or not. The paper proposes to train both a full image and fine-grained patch fire classifier in a joined deep convolutional neural network(CNN). The dataset is acquired from various online resources. The proposed model in this paper consists of a full image CNN and local patch NN classifier, both classifiers share the similar deep neural networks. This paper demonstrates that the model NN-Pool5 consistently outperforms SVM-Raw and CNN-Raw. The neural network in most fire detection research papers is CNN which I am planning on using, and this paper tested, trained and compared the neural networks to see if one works better than the other.

Source 3

Park JH, Lee S, Yun S, Kim H, Kim W-T. Dependable Fire Detection System with Multifunctional Artificial Intelligence Framework. *Sensors*. 2019; 19(9):2025. https://doi.org/10.3390/s19092025
https://www.mdpi.com/1424-8220/19/9/2025/htm

This fire detection system gets data from multiple IoT gateways. The IoT gateways then send that data from sensors to the AI framework which then determines if it is a fire or not. The data includes most aspects such as humidity and temperature. The AI framework consists of CNN algorithm and the DNN algorithm. The image recognition algorithm CNN has proved its effectiveness in analyzing fires; however, the AI framework and multiple other ML algorithms on this model are used to improve its efficiency so fires are alerted faster. The adaptive fuzzy algorithms are also used in this model to reduce false negatives. This model is advanced and uses sensors to predict fires in real-time, which I am not aiming for. This paper is helpful as this too uses CNN for image analysis which I will be using on a data set which is already available.

Source 4

Green, M. E., Kaiser, K., & Shenton, N. (2020). Modeling Wildfire Perimeter Evolution using Deep Neural Networks. *arXiv preprint arXiv:2009.03977*.

https://arxiv.org/pdf/2009.03977.pdf

This paper proposes a wildfire spreading model that predicts the evolution of the wildfire perimeter. The fire spreading simulation uses a deep convolutional neural network(CNN) that is trained on remotely sensed atmospheric and environmental time series data. The data for this model was collected with the use of several United States government agencies and commercial APIs, followed by a multistep process of cropping, geo-synchronizing, and reprojecting data layers into a final tensor to sample from. For my model, I am not planning on using these advanced types of data sets such as which includes temperature and humidity etc, rather a data set which has over a thousand images which are needed by CNN to make an efficient model. This CNN based model was able to accurately detect changes in the atmosphere and detect if there was a fire or no.

Source 5

Mao, W., Wang, W., Dou, Z., & Li, Y. (2018). Fire recognition based on multi-channel convolutional neural network. *Fire Technology*, *54*(2), 531-554

https://link.springer.com/content/pdf/10.1007/s10694-017-0695-6.pdf

A novel fire recognition method based on a multichannel convolutional neural network is proposed in this paper to overcome the deficiencies such as detection using sensors and high definition cameras which can be inefficient because of factors such as car headlights/lamps which the cameras cannot differentiate between . From the public fire dataset and Internet, The data included 7000 images for training and 4494 images for test, and then ran experiments with the comparison of four baseline methods, including deep neural networks. The method MCCNN produces more efficient results. Although there is still progress to be made in producing accurate results in bad weather conditions such as rain and fog etc.

Source 6

Benzekri, W., El Moussati, A., Moussaoui, O., & Berrajaa, M. (2020). Early Forest Fire Detection System using Wireless Sensor Network and Deep Learning. *Int. J. Adv. Comput. Sci. Appl*, *11*(5), 496-503.

https://www.researchgate.net/profile/Ali-El-Moussati/publication/341776345_Early_Forest_Fire_Detection_System_using_Wireless_Sensor_Network_and_Deep_Learning/links/5ed38bab92851c9c5e6c324a/Early-Forest-Fire-Detection-System-using-Wireless-Sensor-Network-and-Deep-Learning.pdf

This model is based on collecting environmental wireless sensor network data from the forest and predicting the occurrence of a forest fire using Deep Learning. It uses a IoT forest fire system which provides 24/7 forest fire monitoring and detection. The proposed models in this paper Simple RNN, LSTM and GRU can improve the prediction accuracy and stability to make it more efficient. Although all models had an accurate result, the GRU model was the most efficient and had a shorter run time. This model, although efficient, uses live data from sensors and cameras etc, and if I were to use it, I would need different hardware to make this work. Which I am not planning on using as my model will be focused on getting the most efficient results using deep learning, and the data set would consist of only images rather than include other factors such as temperature, which I am not planning on using. Although this is not similar to my model, this demonstrates machine learning models I may plan to use.

Source 7

Harish, C., Prasad, D., & Bandari, M. DETECTION OF FIRE WITH DEEP LEARNING.

https://irjmets.com/uploadedfiles/paper/volume2/issue_9_september_2020/3329/1628083132.pdf

This paper develops a classification model using Deep Learning and Transfer Learning to recognise fires in images/video frames, to ensure fire is detected in a timely manner. A difference in this model is that it can detect fires in surveillance videos. The dataset consists of 2 classes - fire and nonfire. The dataset has 1678 fire images/video frames sourced from google. Various available pretrained models in Keras are used to extract the video frame features such as ResNet-50, InceptionV3 and InceptionResNetV2 models are used to extract the features. The model works in real-time and has the ability to send alert emails, also includes a user-friendly graphical interface. As this model is similar to what I am planning on making, except it also detects fires in videos. This paper makes the use of some of the same methods I am planning on working on with my model.