Investment Sentiment Analysis Capstone Project Report

Ramón.Torres

Northwest Missouri State University, Maryville MO 64468, USA
S556497@nwmissouri.edu

Abstract. The abstract should briefly summarize the contents of the paper in 150–250 words. The abstract should briefly summarize the contents of the paper in 150–250 words. The abstract should briefly summarize the contents of the paper in 150–250 words. The abstract should briefly summarize the contents of the paper in 150–250 words.

Keywords: natural language processing · financial analysis· sentiment analysis · capital preservation

1 Introduction

The planned domain is Natural language processing and visualization to demonstrate trends on investment options. The goal is to analyze sentiment analysis of experts, investors and institutional investment firms. Based on current world events and the generalization of the cryptocurrencies, particularly the bitcoin, I would like to analyze investment trends to preserve capital. I will use Kaggle data sources to obtained data sets to analyze financial trends and compare to financial reports and determine the sentiment or opinions of experts on this field . The process will include the following steps:

1. Explore for possible investment analysis scenarios 2. Search for data sources and save it to computer in CSV and MS Excel formats 3. Use python to organize and clean 4. Select appropriate data sources and columns 5. Design and run analysis model 6. Use Tableau to present visualization of relevant results 7. Present project report

Subsequent paragraphs, however, are indented.

2 Goals of this Research

Based on current world events and the proliferation of the digital currencies, it is beneficial to analyze investment trends to preserve capital and enhance return on investments for such instruments. The project is intended to:

1. review forecasting and predictive modelling for the and investment instruments

2. review of historical trends of the bitcoin for the period from 2015 to 2023.

3. generate a predictive tool for future applications to other types of investments and asset classes

This tool may identified parameters that influence the market price of the digital currency. This exercise may be useful for the individual, institutional, or highly speculative investor.

3 Literature Review and methodology approach

Based on the historic data, an estimate and trade model will be developed to identify market value trends. In addition, a time series analysis, the forecasting model may provide sequencing analysis and determine any comprised intervals. In general, an ARIMA model or a method for forecasting and prediction of future values based on historical time series. The predictive model follows the approach of the WalletInvestor. This a prediction site tool that uses machine learning algorithms to analyze and forecast cryptocurrency prices. The site provides daily price predictions for over 800 digital instruments and offers a long-term forecast for up to five years. Studies shows ARIMA model may result in predictions with the smallest for a day to seven days. Therefore, is feasible to use as a predictive method of Bitcoin for that period. On the contrary, a long, short term memory deep learning algorithm with good predictive performance for time series data. This method refers to the data of multiple periods before predicting the price of Bitcoin. This approach is that the prediction accuracy of the model that uses previous period is the best.

4 Data Collection and processing

Data source: Cryptory library for Python and https://www.kaggle.com/

datasets/adilbhatti/bitcoin-and-fear-and-greed. Data sources in CSV and MS Excel formats. Data scraping techniques: Use python to organize and clean. Select appropriate data sources and columns. Data attributes: Cryptograms bitcoin: seven rows, transaction date, open market value, low value, high value, total volume, closing values, market capitalization . Other specific data extraction details related to your project: Based on current world events, clean up items for dates before year 2019 and analyze effects after the COVID 19 Pandemic.

5 Data Cleaning

Using Python to upload the appropriate data sets. Review and check for null values, values before 2019 and possible duplicates. Made backups of original file and working file. Tools and techniques are used in the process: Pandas, Numpy Checking for missing values or deleting roes with missing results to avoid false or misrepresentation of results by improving the quality of the data. Seven columns and 2205 records. The definitions of important data attributes. date = date of transaction

Open = value of bitcoin when market open
Cloe = value of bitcoin when market close
High = highest value of bitcoin on a trading day
Low = lowest value of bitcoin on a trading day
Volume = total shares traded during a trading day
MarketCap = total value in dollars per trading day.

For analysis and forecasting: the independent variables volume and marke-CAp by month. The dependent variable will be next day open.

6 Exploratory Data Analysis

Exploratory data analysis is a stage where data is organized and process to determine the main attributes and characteristics of the data set. Generally, data is organized using a visualization methods, tools, or techniques, like graphs, charts, and tables. It is important in any given data analytics projects to identify trends, outliers, tendencies and prepare the scenario for in depth analysis and discussion of findings. It helps us avoid making any assumptions that might overstated, understated or plain wrong. This will avoid adding additional errors to the forecast or predictive models, using Python or R for example. In research methodology it set the stage for an appropriate statistical summary to perform hypothesis testing. Some of the exploratory data analysis techniques are univariable non graphical and graphical analysis, multivariable nongraphical analysis and graphical analysis. My project will include historical analysis of market values of the cryptocurrencies to consider the trends, frequencies, moving average and seasonality analysis. It may include time series analysis to determine variability rates and univariable correlation analysis among market values and prices. These tools are still under consideration. The techniques included analysis of market values of the cryptocurrencies to consider the trends, frequencies, moving average and seasonality analysis. The process is as follows:

Phase 1: Historical cryptocurrency market value analysis for the period of 2015 to 2019. This technique will provide market value trends.

Phase 2: Historical cryptocurrency trading volume analysis for the period of 2015 to 2019. This technique will provide volume trends.

Phase 3: Moving averages or mean values of cryptocurrencies main attributes, price and volume for the period of 2015 to 2019. This technique will facilitate the visualization of upward and downward trends.

Phase 4: Seasonality analysis on cryptocurrency market value for the period of 2015 to 2019. This technique will decompose the data by trends, seasonality, and noise. Therefore, providing the levels of variability on market value and volume.

Initial findings includes that the prices of cryptocurrencies showed a tendency of having market value under pressure for months of September and October of every year. This showed a downward trend on the values of the digital asset. Then after these periods of overselling, upward trends appear to present until

the end of each year. By December of every year, a buying rally reached a peak, as well as the value of the asset.

7 Predictive Modelling and Analysis Initialization and Setup

For capstone project research the Crypto library for Python was used. It provides historical data from coinmarketcap.com and bitinfocharts.com. These will be the basis for working with making forecasting of the crypto currency over time. Based on the historic data, an estimate and trade model will be developed to identify market value trends. In addition, a time series analysis, the forecasting model may provide sequencing analysis and determine any comprised intervals. In general, an ARIMA model or a method for forecasting and prediction of future values based on historical time series. The predictive model follows the approach of the WalletInvestor. This a prediction site tool that uses machine learning algorithms to analyze and forecast cryptocurrency prices. The site provides daily price predictions for over 800 digital instruments and offers a long-term forecast for up to five years. Studies shows ARIMA model may result in predictions with the smallest for a day to seven days. Therefore, is feasible to use as a predictive method of Bitcoin for that period. On the contrary, a long, short term memory deep learning algorithm with good predictive performance for time series data. This method refers to the data of multiple periods before predicting the price of Bitcoin. This approach is that the prediction accuracy of the model that uses previous period is the best. To determine which previous data points, have more influence on the serial correlation of the crypto currency. Statistical tools will include the application of regression analysis, moving averages, time series and forecasting. To determine which previous data points, have more influence on the serial correlation of the crypto currency. The training process will be fed form the re-sampling of the data will be completed to consider different frequencies and visualization techniques. Using data method for time-series decomposition pf trends, seasonality, and noise. Analysis of movement of the toe series will be performed. The testing process will be looking for to alternatives stationary series or no stationary series based on a statistical p-value. For the predictive model, the series parameters will be plotted to visualize correlation or partial correlation. For this capstone project will be developed using Python programming language, statistical tools, packages, and libraries. Steps will include:

Step 0: Design, organization, and initial setup

Step 1: Installing and Setting up required analysis tools

Step 2 Accessing libraries, analytical, statistical tools

Step 3: Access Crypto Library

Step 4: Accessing Data sets - pull data from start of 2015 to present day

Step 5: Main object definition and accessing historical bitcoin prices from bitinfocharts

Step 6: Additional historical bitcoin data from coinmarketcap

Step 7: Graphs configurations for Cryto Historical trends 2015 to 2023

Step 8: Resampling frequencies on monthly price, annual and quarterly

Step 9: Graphs configurations for Crypto frequencies

Step 10: Graphs configurations for seasonal data decomposition

Step 11 Seasonality analysis parameters definition

Step 12 Seasonality analysis for bitcoin market price per month, decomposition and visualization of results

Step 13 Approximation of parameters using correlation plots.

Step 14 Parameters for approximation definition using correlation.

Step 15 Model Selection and fitting

Step 16 STL-correlational decomposition and visualization of correlational results and residuals

Step 17 Model prediction

The initial analysis included market price trend analysis from 2015 to 2023, moving averages of market prices by days, months, quarters, and years. Also, seasonality analysis decomposition was illustrated for the market price. Price decomposition and statistical analysis shows seasonality trends and residuals. Initial findings includes that the prices of cryptocurrencies showed a tendency of having market value under pressure for months of September and October of every year. This showed a downward trend on the values of the digital asset. Then after these periods of overselling, upward trends appear to present until the end of each year. By December of every year, a buying rally reached a peak, as well as the value of the asset. The following graphs depict the previous initial results analysis. Additional findings and conclusions will be presented in the following tasks, assignment and Oleaf submittal.

Fig. 1. Historical bitcoin market prices from 2015 to 2023.

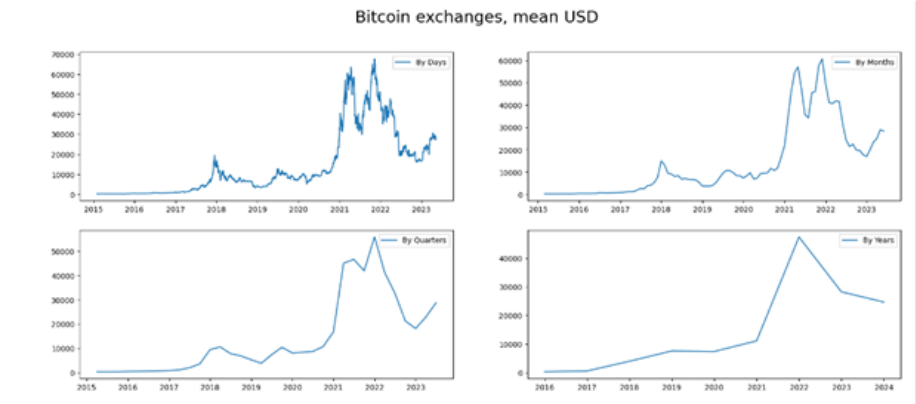

Fig. 2. Historical bitcoin market prices mean from 2015 to 2023 presented in days, months, quarters and years.

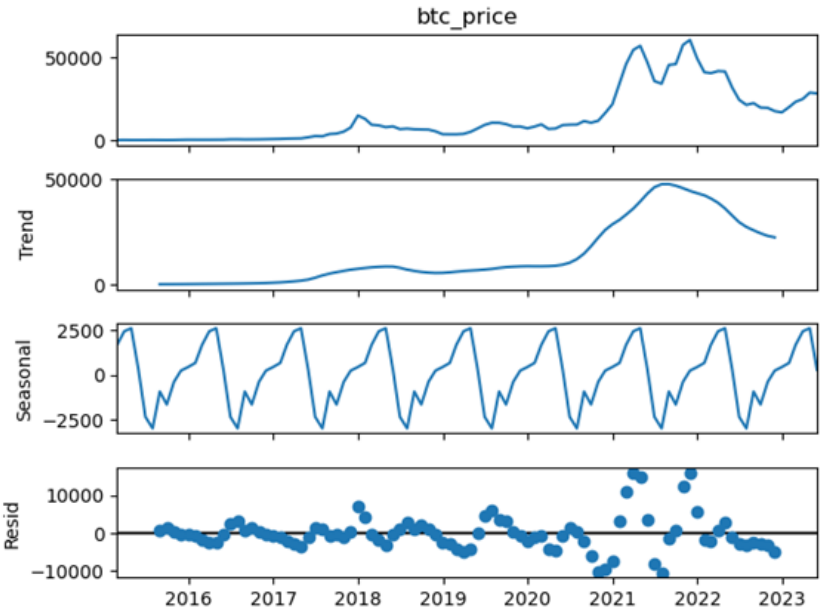

Fig. 3. Historical bitcoin market prices statistical seasonality analysis from 2015 to 2023 presented main parameters are trends, seasonality and residuals.

8 Discussion and findings

9 Conclusions

10 Ethical considerations and limitations

11 Recommendations

[]

References

1. Chein, J.: Analysis of bitcoin price prediction using machine learning. AlphaSene (2023)
2. Kashettar, S.: 5 best crypto prediction site tools, https://www.analyticsinsight.net/5-best-crypto-prediction-site-tools/
3. Sheynin, N.: 3 ways to apply natural language processing (nlp) in financial research. AlphaSene (2022)
4. Torres, R.: 44688-80-capstone-project, https://github.com/rtmtorresmorales/44688-80-Capstone-Project
5. Vimalachandran, A., Jayachandran, T.: The financial crash of 2020 and the retail trader's boon: a correlation between sentiment and technical analysis. SN Business Economics