**Using artificial intelligence to predict cryptocurrency price with Project CryptOL™ software application**

|  |  |  |
| --- | --- | --- |
| **Alexander Gribtsov**  Department of Computer Science and Software Engineering  Monmouth University  400 Cedar Ave, West Long Branch, NJ 07764, United States of America | **Chris Woszczak**  Department of Computer Science and Software Engineering  Monmouth University  400 Cedar Ave, West Long Branch, NJ 07764, United States of America | **John Guseman**  Department of Computer Science and Software Engineering  Monmouth University  400 Cedar Ave, West Long Branch, NJ 07764, United States of America |

|  |  |
| --- | --- |
| ***Abstract***  Background: Implementation of different types of artificial intelligence algorithms can produce fit data that predicts the price of different cryptocurrency types such as BitCoin.  Related Work: Two related projects were observed for cross comparison of data performance and overall cryptocurrency price forecasting experience.  Architecture: Project CryptOL is a web based full stack application. The design, implementation and deployment of this project is covered in detail.  Testing: Embedded into the CryptOL application as part of the design of the project is a classical false negative and false positive testing technique of fitness of running data.  Discussion: During the development of the project all members of the team became familiar with most common data science terminology and functionality of standardized techniques.  Conclusions: Project CryptOL is a great example of how standardized data science techniques and algorithms can be applied together to present a strong indicator of future cryptocurrency price.  ***Keywords***  Algorithms, artificial intelligence, bitcoin, cryptocurrency, machine learning, online machine learning, LSTM, neural networks, data science, survival modeling, time series, linear regression, logistic regression.    I. Background  In the field of data science, artificial intelligence (AI) is best suited for collecting large datasets and performing evaluations that result in conclusions that can be measured. Implementation of different types of AI algorithms can produce fit data that predicts the price of different cryptocurrency types such as BitCoin. Countless data algorithms have been designed and implemented into many programs that perform analysis of the likelihood of an event or a prediction of the next probable sequence in time. Models  The second project competitor is Nirav from Pirimid Fintech and his open-source LSTM Neural Net model. After doing research on this project the superiority of this model became very clear. The team will be integrating this open-source model as part of a multi algorithm package. While CryptOL™ web-based application allows us to integrate multiple AI models for cross comparison, the quality of each model and the accuracy of their predictions is coveted. In the future, multiple algorithms will use parts of each currently functioning prediction algorithm in synchrony. Therefore doing research on our competitors and predecessors in this case allowed us to improve our product immediately up on the discovery from research. III. Architecture The goal of CryptOL is to use machine learning to predict the trend of Bitcoin prices. The project team selected three machine learning models for the experiment: linear regression, autoregressive integrated moving average (ARIMA), and long short-term memory (LSTM). To evaluate the success of the models, the project examines the directional accuracy of the predictions. That is, if a model predicts the price of Bitcoin will rise and it does rise, then a successful prediction has occurred. In addition, the focus of CryptOL is on short-term price predictions 15 minutes, 1 hour, and 1 day into the future. Cryptocurrency markets are highly volatile. Therefore, the project team concluded that it would be more difficult to arrive at long-term predictions. *A.* *Project Design* While CryptOL uses three separate machine learning models, the project exhibits a singular design to tie the components together. The figure below illustrates the general design of the project. | which are most famous are; linear regression [1], logistic regression [2], long short-term memory (LSTM) [3], time series [4], survival modeling [5], online machine learning [6], neural networks [7], to name a few.  These types of data manipulation algorithms when properly combined with feature engineering can produce very firm results. In combination and side by side cumulatively this data can give an observer great probability statistics when making a fit prediction of a future cryptocurrency price.II. Related Work The complex subject of crypto currency price prediction is constantly flooded with new information, opinions and techniques. After doing research and technical analysis of competitor software our group came up with two software products that rival and/or resemble our current project CryptOL™.  Two related projects were observed for cross comparison of data performance and overall crypto price forecasting experience. Nirav Prajapati’s (from Pirimid Fintech) open source work using machine learning and Crypto Forecast: AI Predictions using neural networks were observed.  First let’s look at Crypto Forecast, the AI predictions Android app. This product is a fully finished product that uses a neural net model for predictions. While this application is fully finished and utilizes one of the algorithms that CryptOL™ app uses in this case the neural network algorithm, the limitation is in that single algorithm and therefore this competitor’s best feature is the interface and quality of the GUI as well as how some statistical data is presented. Where CryptOL™ in its current state is an unrefined GUI with just 3 working algorithmic models and multiple additional algorithms in the works.  **Data Generation**  CryptOL obtains its data using the Yahoo Finance API. The number of historical Bitcoin prices the project collects depends on the price prediction interval selected by the user. 15-minute predictions collect prices from the previous 60 days. This results in a data set with 5,760 price points ((1,440 minutes in a day X 60 days) / 15-minute intervals). 1-hour predictions collect prices from the previous 60 days resulting in a data set of 1,440 price points. Finally, the 1-day predictions collect prices from the previous 365 days. Unfortunately, this results in a data set with only 365 price points.  **Splitting the Data**  Machine learning algorithms generally require that the data collected be split into at least two separate sets before a model is trained. One set is called the training set and the other is the test set. A machine learning model is trained using the training set. Predictions are made by applying the test set to the newly trained model.  CryptOL™ uses 80% of the data for training sets and 20% of the data for test sets.  **Model Training, Predictions, & Evaluation**  This will be covered in the **IMPLEMENTATIONS** section of the paper.  **User Interaction**  The user interacts with CryptOL through a graphical user interface. The GUI is built using a service called Anvil. Client requests are made on the front-end of the website. The python code used to make the predictions is run on the back-end of the Anvil service. The user selects an algorithm and a time interval if the option is offered. After some time, a prediction is returned to the user. At this point the user can decide whether to utilize the prediction result of the algorithm. *B. Project Implementation* **Linear Regression**  CryptOL™ implemented a multiple linear regression model. Linear regression makes it possible to predict the value of an independent variable based on the values of one or more independent values, or features. For this project, the dependent variable is the price of Bitcoin in 15 minutes, 1 hour, or 1 day. The dependent variables are the features that the project team selected.  This project employs the ordinary least squares method of linear regression. In the case of just two variables, one dependent and one independent, ordinary least squares regression seeks to fit a straight line through the plotted |
|  |  |

|  |  |
| --- | --- |
| points. This fitted line is meant to model a relationship between the dependent and independent variables. The best fitting line is one that minimizes the distance between the values predicted by the line and the actual values of the regression model is by looking at the coefficient of determination, or R-Squared. The R-Squared ranges from 0 to 1. The number measures the amount of variance in the dependent variable caused by changes in the independent variable.  While the CryptOL’s linear regression models often displayed R-Squared scores greater than 95%, this is likely not an ideal outcome. This is probably evidence that our models are overfitting the training dataset. In other words, the model exhibits desirable outcomes with the training data, however it reacts poorly when confronted with new data.  53 features were selected for the model. Three of these features are the prices of Bitcoin, Ether, and Dogecoin. The remaining 50 features include the 50 previous of Bitcoin at the selected time interval. The linear regression model trained using these features was able to predict the direction of the market with an accuracy as high as 72% sometimes.  **Time Series (ARIMA)**  CryptOL™ also implemented an Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA model is used for time series forecasting. In general, a time series forecasting model will try and predict the future value of something based upon known past values.  There are three parts to the ARIMA model. Autoregressive refers to the number of lags used to determine the future value of something. For instance, is there a high correlation between the current value and the value one, two, three, etc. lags behind. For CryptOL, we saw the best results using a lag value of one. This means that if the user wishes to predict the price of bitcoin 15 minutes into the future, then CryptOL’s ARIMA model will pay close attention to the price in the immediate 15 minutes before.  The next part of ARIMA is the ‘integrated’ aspect. Generally, time series forecasting requires the data to be stationary. This means that the mean and variance of the sample data remain constant. Unfortunately, this is often not the case with Bitcoin prices. Integrating the values, in our case the price of Bitcoin, just means subtracting one value from its adjacent value. This differencing results in sample data that is stationary. This differencing can take place as many times as it is necessary to arrive at stationary data. The price of Bitcoin, like most time series data, only had to be differentiated once to achieve stationary data.  The final aspect of the ARIMA model is the moving average. It is possible to make the ARIMA take into | account the moving average for a certain number of periods while making its prediction. CryptOL chose not to take into account the moving average for two reasons. First, experimenting with different moving averages did not drastically affect the results of the predictions. Second, including the moving averages greatly increased the time it took for the model to return a prediction.  **Long short-term memory (LSTM)**  Long short-term memory (LSTM) is a form of recurrent neural networks (RNN). Unlike other RNNs, LSTM utilizes a cell state in order to store information over long periods of time. It does this by using a three gate system input gates, output gates, and forget gates, these gates maintain the flow of information over time. When compared with other RNNs, this cell state makes it advantageous when considering long-term predictions. Like ARIMA, this model is a time series forecasting model that can predict future values based on known previous values.  CryptOL implemented a sequential stacked LSTM model, using a one layered approach with 256 neurons. This model takes into consideration three features in order to make its prediction. They include the open, high, and close values at 15-minute intervals.  When creating a LSTM model, it is important to minimize the training loss & validation loss, while keeping the values of each close to prevent over or under fitting. To achieve this combination numerous factors were considered, tested, and optimized. However, hyperparameters such as number of epochs, number of neurons, and the type of activation function being used have shown in our case to correspond to the model's accuracy and consistency the most.  After testing, in our case, the optimal number of epochs to produce the best accuracy and consistency possible was 80. Higher numbers increased prediction time and showed no significant improvement in accuracy. Lower numbers decrease prediction time, but provide more inconsistent predictions.  As mentioned earlier, 256 neurons are used when compiling the model. There is not a one size fits all approach to determining the number of neurons, after experimenting with various other values such as 32, 64, 128, and 512. Smaller values decrease prediction time, but also decrease the consistency of the model's accuracy. Values of 512 or higher |

|  |  |
| --- | --- |
| made no significant improvements, and dramatically increased the time of prediction.  The only convincing options when considering activation function were Tanh and ReLU, as they both produced less loss than others while experimenting. ReLU showed a significant decrease in time taken to produce predictions. However, with this in mind, CryptOL™ decided to use the Tanh activation function because in our case, average accuracy is improved around 5% compared to ReLU. After all of the optimizations and many smaller tweaks, this model has predicted with accuracy as high as 68%.  *C. Project Deployment*  Once comfortable with our accuracy and consistency in predicting trends in the crypto market, CryptOL™ created its own web service via Anvil as previously mentioned. Anvil is able to provide our models with a reliable domain so investors and people with cryptocurrency interests alike can use them. Anvil was chosen because of its many features. The two most important include a built in database and the unique ability to implement all aspects of the project with strictly Python, making deployment simple when compared to other options.  When CryptOL™ is made available to the public, updates to our web service will stem directly from user feedback. These updates will occur biweekly inorder to please users and improve our product. Updates to improve or add models will be made quarterly to ensure that there is time for proper testing and experimentation. IV. Experiment Testing The project evaluated models based on the directional accuracy of the predictions. In other words, the project was less concerned with predicting the precise value of Bitcoin in the future than with the direction of the market. A prediction was considered a true positive if the price of Bitcoin rose and our model also predicted a rise in price. A true negative prediction occurred when the price of Bitcoin fell and our model also predicted a fall in price. Both predictions are desirable. Therefore, the accuracy of a certain model can be determined by adding the true positives and true negatives and dividing this number by the total number of predictions returned by the model.  Embedded into the CryptOL™ application as part of the design project is a classical testing technique of fitness of | running data. False negative and false positive results using data set comparison is integrated into the application graphical user interface (GUI) as part of the end product of accurate data production. In addition to the integrated data testing of the models used in the project, independent data fit tests were performed on each algorithm. In one random week in April the linear regression algorithm was within 20% accuracy in predicting throughout the week.    The table above demonstrates the accuracy fit of the algorithm. All of the algorithms in the CryptOL™ project were within 20% of the correct price. V. Project CryptOL™ Discussion Going into the project the team had little to no experience in the field of data sciences. Well into the project all members of the team became familiar with most common data science terminology and functionality of standardized techniques, from online machine learning to feature engineering and many algorithmic techniques such as logistic regression, time series, survival modeling, long short-term memory (LSTM) were explored. Two O’Reilly books were used as reference to build the project framework: Hands-on Machine Learning with Scikit-Learn [8], Keras & TensorFlow and Introduction to Machine Learning with Python [9]. The design of this project took the team in the direction of a full stack web application designed in python relying heavily on TensorFlow library and ran on an Anvil full stack environment. VI. Conclusion Project CryptOL™ is a great example of how standardized data science techniques and algorithms can be applied together in an application with a GUI to present a user strong indicator of future cryptocurrency price. This |

|  |  |
| --- | --- |
| project concluded with a fully functional full stack web application that predicts cryptocurrency prices. The data testing conducted clearly shows the close fitness and accuracy of the data with a small error margin.  In conclusion the team agrees and presents that combining multiple machine learning algorithms into one graphical interface gives a user an extreme advantage when it comes to predicting the possible direction of the cryptocurrency market. Acknowledgment Project CryptOL™ was conducted as part of a capstone course in the Department of Computer Science and Software Engineering, Monmouth University. The course was administered and project CryptOL™ overseen by Samer Y. Khamaiseh, Ph.D.  Project CryptOL™ open source version of the code can be found on GitHub at: <https://github.com/tocold22/CryptOL> CryptOL™ is available for live for testing at: <https://www.itechnologyllc.com/cryptol> References [1] Cohen, Gil. “Forecasting Bitcoin Trends Using Algorithmic Learning Systems.” *Entropy* 22, no. 8 (2020): 838. doi:10.3390/e22080838.  [2] Lee, Brian K., Justin Lessler, and Elizabeth A. Stuart. “Improving Propensity Score Weighting Using Machine Learning.” *Statistics in Medicine* 29, no. 3 (2009): 337–46. doi:10.1002/sim.3782.  [3] Munim, Ziaul Haque, Mohammad Hassan Shakil, and Ilan Alon. “Next-Day Bitcoin Price Forecast.” *Journal of Risk and Financial Management* 12, no. 2 (2019): 103. doi:10.3390/jrfm12020103.  [4] “Introduction to Neural Networks for Time Series Forecasting.” *Machine Learning for Time Series Forecasting with Python®*, 2020, 137–65. doi:10.1002/9781119682394.ch5.  [5] Chan, Phyllis, Xiaofei Zhou, Nina Wang, Qi Liu, René Bruno, and Jin Y. Jin. “Application of Machine Learning for Tumor Growth Inhibition – Overall Survival Modeling Platform.” *CPT: Pharmacometrics & Systems Pharmacology* 10, no. 1 (2020): 59–66. doi:10.1002/psp4.12576.  [6] Tegen, Agnes. “Approaches to Interactive Online Machine Learning,” n.d. doi:10.24834/isbn.9789178770854.  [7] Felizardo, Leonardo, Roberth Oliveira, Emilio Del-Moral-Hernandez, and Fabio Cozman. “Comparative Study of Bitcoin Price Prediction Using WaveNets, Recurrent Neural Networks and Other Machine Learning Methods.” *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, 2019. doi:10.1109/besc48373.2019.8963009. | [8] Géron, Aurélien. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol,  CA: O'Reilly Media, Inc., 2019.  [9] Müller, Andreas Christian, and Sarah Guido. *Introduction to Machine Learning with Python: a Guide for Data Scientists*. Sebastopol: O'Reilly Media, 2018. |