International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

AI BASED CRYPTO CURRENCY RATE PREDICTION USING LSTM ALGORITHM

S.SATHISH KUMAR
Department Of Computer Science And Engineering,
PSN College Of Engineering And Technology,
Tirunelveli.

MRS.P.KANAGA LAKSHMI.,M.E
Associate Professor,
Department Of Computer Science And Engineering,
PSN College Of Engineering And Technology,
Tirunelveli.

Abstract— **The proposed system describes a cryptocurrency rate prediction system using Long Short- Term Memory (LSTM) neural networks. Cryptocurrency has gained Widespread acceptance in recent years, and predicting the rate of cryptocurrencies Is crucial for investors to make informed decisions. LSTM networks have been Widely used in time-series prediction tasks due to their ability to handle long-term Dependencies. The proposed system aims to predict the future rate of a Cryptocurrency by analyzing historical data using LSTM networks. The system Uses features such as opening price, closing price, highest price, lowest price, and Trading volume to train the LSTM network. The trained model is then used to Predict the future rate of the cryptocurrency. The system is evaluated using root Mean square error (RMSE) and accuracy score metrics. The results show that the Proposed system achieves high prediction accuracy, making it a valuable tool for Investors to make informed decisions about cryptocurrency investments.**

I. INTRODUCTION

Bitcoin (₿) is a decentralized digital currency, without a central bank or Single administrator, that can be sent from user to user on the peer-to-peer bitcoin Network without the need for intermediaries. Transactions are verified by Network nodes through cryptography and recorded in a public distributed Ledger called a blockchain. The cryptocurrency was invented in 2008 by an Unknown person or group of people using the name Satoshi Nakamoto. The Currency began use in 2009 when its implementation was released as open-source Software. Bitcoins are created as a reward for a process known as mining. They can Be exchanged for other currencies, products, and services. Bitcoin has been Criticized for its use in illegal transactions, the large amount of electricity (and Thus carbon footprint) used by mining, price volatility, and thefts from exchanges. Some investors and economists have characterized it as a speculative bubble at Various times. Others have used it as an investment, although several regulatory Agencies have issued investor alerts about bitcoin.A few local and national governments are officially using Bitcoin in some Capacity, with one country, El Salvador, adopting it as a legal tender. The Word bitcoin was defined in a white paper published on 31 October2008.

Transactions:

Transactions are defined using a Forth-like scripting language. Transactions consist of one or more inputs and one or more outputs. When a user Sends bitcoins, the user designates each address and the amount of bitcoin being Sent to that address in an output. To prevent double spending, each input must Refer to a previous unspent output in the blockchain. The use of multiple inputs Corresponds to the use of multiple coins in a cash transaction. Since transactions Can have multiple outputs, users can send bitcoins to multiple recipients in one Transaction. As in a cash transaction, the sum of inputs (coins used to pay) can Exceed the intended sum of payments. In such a case, an additional output is used, Returning the change back to the payer. Any input satoshis not accounted for in The transaction outputs become the transaction fee.The blocks in the blockchain were originally limited to 32 megabytes in Size. The block size limit of one megabyte was introduced by Satoshi Nakamotoin 2010. Eventually the block size limit of one megabyte created problems for Transaction processing, such as increasing transaction fees and delayed processing Of transactions.[30] Andreas Antonopoulos has stated Lightning Network is a Potential scaling solution and referred to lightning as a second-layer routing Network

Ownership:

Simplified chain of ownership as illustrated in the bitcoin whitepaper. In Practice, a transaction can have more than one input and more than one output.In the blockchain, bitcoins are registered to bitcoin addresses. Creating a Bitcoin address requires nothing more than picking a random valid private key And computing the corresponding bitcoin address. This computation can be done In a split second. But the reverse, computing the private key of a given bitcoin Address, is practically unfeasible. Users can tell others or make public a bitcoin Address without compromising its corresponding private key. Moreover, the Number of valid private keys is so vast that it is extremely unlikely someone will Compute a key-pair that is already in use and has funds. The vast number of valid Private keys makes it unfeasible that brute force could be used to compromise a Private key. To be able to spend their bitcoins, the owner must know the corresponding private key and digitally sign the transaction.

The network Verifies the signature using the public key the private key is never revealed.If the private key is lost, the bitcoin network will not recognize any other Evidence of ownership; the coins are then unusable, and effectively lost. For Example, in 2013 one user claimed to have lost 7,500 bitcoins, worth $7.5 million At

the time, when he accidentally discarded a hard drive containing his private Key. About 20% of all bitcoins are believed to be lost -they would have had a Market value of about $20 billion at July 2018 prices.To ensure the security of bitcoins, the private key must be kept secret. 10 If The private key is revealed to a third party, e.g. through a data breach, the third Party can use it to steal any associated bitcoins. As of December 2017, around 980,000 bitcoins have been stolen from cryptocurrency exchanges

Mining:

Mining is a record-keeping service done through the use ofComputer processing power. Miners keep the blockchain consistent, complete, And unalterable by repeatedly grouping newly broadcast transactions into a block, Which is then broadcast to the network and verified by recipient nodes. Each Block contains a SHA-256 cryptographic hash of the previous block, thus linking It to the previous block and giving the blockchain its name.

To be accepted by the rest of the network, a new block must contain A proof-of-work (PoW). The PoW requires miners to find a number called A nonce (number used once), such that when the block content is hashed along With the nonce, the result is numerically smaller than the network's difficulty

Target.[6]:ch. 8 This proof is easy for any node in the network to verify, but Extremely time-consuming to generate, as for a secure cryptographic hash, miners Must try many different nonce values before a result happens to be less than the Difficulty target. Because the difficulty target is extremely small compared to a Typical SHA-256 hash, block hashes have many leading zeros. as can be seen in This example block hash00000000000000000000590fc0f3eba193a278534220b2b37e 9849e1a770ca959

By adjusting this difficulty target, the amount of work needed to generate A block can be changed. Every 2,016 blocks (approximately 14 days given Roughly 10 minutes per block), nodes deterministically adjust the difficulty target Based on the recent rate of block generation, with the aim of keeping the average Time between new blocks at ten minutes. In this way the system automatically Adapts to the total amount of mining power on the network. As of September 2021, it takes on average 79 sextillion (79 thousand billion billion) attempts to Generate a block hash smaller than the difficulty target. Computations of this Magnitude are extremely expensive and utilize specialized hardware. The proof-of-work system, alongside the chaining of blocks, makes Modifications of the blockchain extremely hard, as an attacker must modify all Subsequent blocks in order for the modifications of one block to be accepted. As New blocks are mined all the time, the difficulty of modifying a block increase As time passes and the number of subsequent blocks increase.

Privacy and fungibility:

Bitcoin is pseudonymous, meaning that funds are not tied to real-world Entities but rather bitcoin addresses. Owners of bitcoin addresses are not Explicitly identified, but all transactions on the blockchain are public. In addition,

Transactions can be linked to individuals and companies through "idioms of use" And corroborating public transaction data with known information on owners of Certain addresses. Additionally, bitcoin exchanges, where bitcoins are traded for Traditional currencies, may be required by law to collect personal information. To Heighten financial privacy, a new bitcoin address can be generated for each Transaction.

Wallets and similar software technically handle all bitcoins as equivalent, Establishing the basic level of fungibility.

Data Mining for bitcoin prediction:

Data Mining or Knowledge Discovery in Databases (KDD), refers to the Process of investigation of hidden information patterns from different Perspectives for categorization into useful information. Currently, data mining and KDD are utilized interchangeably by statisticians, data analysts, and Information systems experts. This process includes different types of services Like web mining, pictorial data mining, audio and video mining, text mining And social media mining. The biggest challenge is to analyze the large data to Extract important information that can be used to generating predictive Knowledge. Data mining offers a set of techniques and tools for finding patterns And extracting knowledge in the bitcoin dataset that are difficult to detect with Traditional statistical methods. Hence, Data mining provides the methodology And technology to predict the rate of bitcoin with high accuracy and less costs.

Data Preprocessing:

The real-world databases are highly susceptible to missing, inconsistent and Noisy data due to their typically huge size and their likely origin from multiple, Heterogeneous sources . Data preprocessing is a data mining technique that is,Used to transform the CVDs patient dataset in a useful and efficient format . The

Commonly used phases in the process of data mining for extracting knowledge Data mining includes the following processes:

i. Data Cleaning: Inconsistent and noise data is eliminated.

ii. Data Integration: Multiple data from different sources are combined.

iii. Data Selection: Data having close connections are analysed and Retrieved from the database.

iv. Data Transformation: Transforming and strengthening the data With the performance of aggregation or summary operations.

v. Data Mining: It is a necessary process for the extraction of data Patterns by the application of intelligent methods.

vi. Pattern Evaluation: Patterns that are interesting and representing Knowledge, depending on intrusiveness is recognized.

1.2.2. Data Mining Techniques:

Data mining techniques give a confidence level about the predicted solutions In terms of consistency of prediction and accuracy. The antly to the success Of data mining and its extensive applications . The following are a few Techniques of datamining:

(i) Statistics: Statistics is the branch of mathematics, which deals with the Collection and analysis of numerical data by using various methods and Techniques. It is used in various stages of the data mining. For example,We can use statistics in the collection, sampling methods and analysis Stage. Data summarization, point estimation, and Testing a hypothesis Are the statistical techniques that find its extensive usage in data mining

(ii) Machine Learning: Machine learning refers to the process of generating A computer system that has the capability of acquiring data independently And integrating that data to generate useful knowledge. The use of machine learning in data mining is that machine learning enables us to Discover new and interesting structures and formats about a set of Previously unknown data.

(iii) Database Systems and Data Warehouses:

The database system mainly Emphasis on creating, maintaining and use of organizational and end-User databases. The principles of database systems are high in the data Models, query processing, query languages, storing data, methods of Optimization, indexing, and accessibility methods. Recently database Systems have made use of data mining and data warehousing facilities to Build systematized analyticalcapabilities on the database.

(iv) Information Retrieval: Information retrieval is the process of searching For information in the documents. Documents may be in the multimedia Or text form or may reside on the web. Effective search and analysis have Raised many challenging issues in data mining. Therefore, multimedia Mining and text data mining integrated with information retrieval Methods have become increasingly important.

(v) Artificial Intelligence: It is a branch of computer science that create Intelligent machines which can behave like a human, think like humans,And able to make decisions.

Data Mining Tasks:

Data mining uses several algorithms for performing a variety of tasks. These Algorithms examine the sample data of a problem and determine a model that Fits close to solving the problem.

a) Predictive Model: A predictive model enables the prediction of data Values by using the results which are known from the different sample Dataset. The data mining tasks that form the part of the predictive model Are:

(i) Classification: Classification refers to the process of determining a Model for describing the concepts or data classes. The model is Derived based on the analysis of a set of training data. The Classification task allows not only the review and analysis of current Data but also enable to predict the future behavior of sample data. The Derived model can be represented

in different forms, like Classification rules, that is, IF-THEN rules, mathematical formulae, Decision trees, naïve bayes and neural networks.

(ii) Regression: Regression Analysis is the most common statistical Modeling approach used in data mining which is used to measure the Average relationship between two or more variables in terms of the Original unit of data. It allows the future data values to be forecasted Depending upon the current as well as the past data values. Thistechnique is used by researchers in many areas, especially in health Sciences.

(iii) Time Series Analysis: Time series analysis is one of the data mining Predictive models enables to predict future values for the current set Of values and which is evenly distributed as hourly, daily, weekly,Monthly and yearly to draw a time series plot. This analysis uses the Present and the past data sample to estimate future value.

(b) Descriptive Model: A descriptive model allows us to determine the Relationships and patterns in a data sample. Descriptive model in data Mining tasks are:

(i) Clustering: Data items that have close resemblance with one another Are clubbed together in a single group is known as clusters which Are referred to as unsupervised machine learning techniques. It Enables to create new classes and groups depending on the study of Patterns and relationships between data values in a database.

(ii) Summarization: Summarization is one of the descriptive models in Data mining also be referred to as generalization or characterization. The use of summarization allows us to summarize a large chunk of Data containing in a web page or a document. This task searches for Specific data attributes and characteristics in the given large data Volumes which can be summarized.

(iii) Association Rules: Association rules enable the establishment of Association and relationships formed between large unclassified Data items depending on certain features and attributes. It defines Certain rules of associativity between data sets and then uses those

Rules to establish relationships.

(iv) Sequence Discovery: Sequence discovery is allowing us to Determine the sequential patterns that might exist in a large and unorganized database. It discovers the sequence in a data bank by Using the time factor i.e., associate the data items by the time at Which it was generated.

Data Mining Applications:

Data mining finds its applications in various fields in our day-to- day life. It is More useful for healthcare and pharmaceutical organizations, which produce Large volumes of data in their diagnostic and clinical activities to access the Doctors in making their clinical decision. Almost all the present-day varied Organizations use data mining in all the phases of their work and which can Apply the various concepts, methods, tasks, and techniques of data mining for Decision- making. It can be utilized in several applications like Insurance, Education, Business analysis, Fraud Detection, Computer security, News & Entertainment, Production Control, Health and Science Exploration

SURVEY

Karasu, S et al performed a Bitcoin prediction with Linear Regression (LR) and Support Vector Machine (SVM) from machine learning methods by using time series consisting of daily Bitcoin closing prices between 2012-2018. The prediction model with include the least error is obtained by testing with different parameter combinations such as SVM with including linear and polynomial kernel functions. Filters with different weight coefficients are used for different window lengths. For different

window lengths, Bitcoin price prediction is made using filters with different weight coefficients. 10-fold cross validation method in training phase is used in order to construct a model with high performance independent of the data set. The performance of the obtained model is measured by means of statistical indicators such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Pearson Correlation. It is seen that the price prediction performance of the proposed SVM model for Bitcoin data set is higher than that of the LR model. Velankar, S et al attempt to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value. For the first phase of our investigation, we aim to understand and identify daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. Our data set consists of various features relating to the Bitcoin price and payment network over the course of five years, recorded daily. For the second phase of our investigation, using the available information, we will predict the sign of the daily price change with highest possible accuracy Prediction: The Bitcoin's value varies just like a stock albeit differently. There are a number of algorithms used on stock market data for price prediction.

However, the parameters affecting Bitcoin are different. Therefore it is necessary to predict the value of Bitcoin so that correct investment decisions can be made.. The price of Bitcoin does not depend on the business events or intervening government unlike stock market. Thus, to predict the value we feel it is necessary to leverage machine learning technology to predict the price of Bitcoin. Demir, A et al proposed a new cryptology methods related to bitcoin price estimation and security. In our study, bitcoin price estimation was made by using machine learning methods using KAGGLE Bitcoin Dataset 2010-2019 data set. The methods used are long-short term memory networks, support vector machines, artificial neural networks, Naive Bayes, decision trees and the nearest neighbor algorithm. Obtained accuracy rates are 97.2%, 91.8%, 86.6%,
85%, 81.2% respectively.

Aggarwal, A et al performed a comparative analysis of various deep learning models on various parameters affecting bitcoin price prediction on the basis of RMSE values. The results show that the various deep learning models are evidently effective for the prediction of bitcoin price. LSTM results into the least RMSE value when bitcoin parameters are considered

for price prediction. On the other hand, when gold price is used as the parameter for bitcoin price prediction, LSTM does not show a positive correlation between the gold price and bitcoin price. As a result, the graph of actual versus predicted values shows a huge deviation. The results of bitcoin price prediction using twitter sentimental analysis show a positive correlation and is shown when a positive tweet is posted in context of bitcoin, the prices are expected to hike and when a negative tweet is passed related to the bitcoin, the prices do get affected and are expected to fall depending on the popularity of the twitterer which can be measured using the number of followers of twitterer on twitter. In future, further research can be done to quantify the effect of twitterer influence on the bitcoin price. To improve the accuracy of results further, live dataset input streams of various parameters affecting bitcoin price can be used for price predictions of bitcoin

Singh, D et al aims to utilize the properties of signed network and hence builds a Bitcoin alpha (whom-trust-whom) network. This trading network is 23 further scrutinized to estimate user's reputation i.e. predicting ranks of participating actors. Thus, we are able to provide score to individual user to recommend whether he is safe to make transactions with. In addition, novel method for link prediction by calculating strength of path between two unconnected nodes is proposed. This strength value considers all the features of the nodes, edges and neighbors giving results that have higher accuracy over previously defined methods.

Anupriya et al performed a prediction of Bitcoin close price by using the ARIMA model . The ARIMA model is found suitable for the prediction of bitcoin prices because this model is used for prediction of time series data. The forecast of future values is provided based on seasonality and trend present in the price data. In terms of visualizations, results are manifest by using R programming language. The obtained results are then compared with actual prices and percent mean error is calculated. The present mean error is found here less than 6% for most of the values

Serafini, G. et al investigate the predictive power of network sentiments and explore statistical and deep-learning methods to predict Bitcoin future price. In particular, we analyze financial and sentiment features extracted from economic and crowdsourced data respectively, and we show how the sentiment is the most significant factor in predicting Bitcoin market stocks. Next, we compare two models used for Bitcoin time-series predictions: the AutoRegressive Integrated Moving Average with eXogenous input (ARIMAX) and the Recurrent Neural Network (RNN). We demonstrate that both models achieve optimal results on new predictions, with a mean squared error lower than 0.14%, due to the inclusion of the studied sentiment feature. Besides, since the ARIMAX achieves better predictions than the RNN, we also prove that, with just a linear model, we may obtain outstanding market forecasts in the Bitcoin scenario.

Jang, H et al reveals the effect of Bayesian neural networks (BNNs) by analyzing the time series of Bitcoin process. We also select the most relevant 24 features from Blockchain information that is deeply involved in Bitcoin's supply and demand and use them to train models to improve the predictive performance of the latest Bitcoin pricing process. We conduct the empirical study that compares the Bayesian neural network with other linear and non-linear benchmark models on modeling and predicting the Bitcoin process. Our empirical studies show that BNN performs well in predicting Bitcoin price time series and explaining the highvolatility of the recent Bitcoin price.

PROBLEM STATEMENT

Bitcoin uses a peer-to-peer technology to operate with no central authority or banks.Bitcoin is open-source; its design is public, nobody owns or controls Bitcoin and everyone can take part.Digital currency bring into use as open source software in 2009 by pseudonymous creator Satoshi Nakamoto It is a cryptocurrency, so-called because it uses cryptography to control the creation and transfer of money.

DOMAIN DECRIPTION & EXISTING MODELS

There are two forms of data analysis that can be used for extracting models describing important classes or to predict future data trends.

These two forms are as follows –

- Classification
- Prediction

Classification models predict categorical class labels; and prediction models predict continuous valued functions.

Classification and Prediction Issues

The major issue is preparing the data for Classification and Prediction. Preparing the data involves the following activities .

• **Data Cleaning** − Data cleaning involves removing the noise and treatment of missing values. The noise is removed by applying smoothing techniques and the problem of missing values is solved by replacing a missing value with most commonly occurring value for that attribute.

• **Relevance Analysis** − Database may also have the irrelevant attributes. Correlation analysis is used to know whether any two given attributes are related.

• **Data Transformation and reduction** − The data can be transformed by any of the following methods. o Normalization − The data is transformed using normalization. Normalization involves scaling all values for given attribute in 26 order to make them fall within a small specified range. Normalization is used when in the learning step, the neural networks or the methods involving measurements are used. o Generalization − The data can also be transformed by generalizing it to the higher concept. For this purpose we can use the concept hierarchies. Note − Data can also be reduced by some other methods such as wavelet transformation, binning, histogram analysis, and clustering. Comparison of Classification and Prediction Methods Here is the criteria for comparing the methods of Classification and Prediction –

• **Accuracy** − Accuracy of classifier refers to the ability of classifier. It predict the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data. • Speed − This refers to the computational cost in generating and using the classifier orpredictor.

• **Robustness** − It refers to the ability of classifier or predictor to make correct predictions from given noisy data.

• **Scalability** − Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.

• **Interpretability** − It refers to what extent the classifier or predictor understands.

Types of Algorithms In Data Mining

a. Statistical Procedure Based Approach

There are two main phases present to work on classification. That can easily identify the statistical community. The second, “modern” phase concentrated on more flexible classes of models. In which many of which attempt has to take. That provides an estimate of the joint distribution of the feature within each class. That can, in turn, provide a classification rule.

Generally, statistical procedures have to characterize by having a precise fundamental probability model. That used to provides a probability of being in each class instead of just a classification. Also, we can assume that techniques will use by statisticians. Hence some human involvement has to assume with regard to variable selection. Also, transformation and overall structuring of the problem.

b. Machine Learning-Based Approach

Generally, it covers automatic computing procedures. That was based on logical or binary operations. That use to learn a task from a series of examples. Here, we have to focus on decision-tree approaches. As classification results come from a sequence of logical steps. These classification results are capable of representing the most complex problem given. Such as genetic algorithms and inductive logic procedures (I.LP.) are currently under active improvement.

Also, its principle would allow us to deal with more general types of data including cases. In which the number and type of attributes may vary.

This approach aims to generate classifying expressions. That is simple enough to understand by the human. And must mimic human reasoning to provide insight into the decision process.

Like statistical approaches, background knowledge may use in development. But the operation is assumed without human interference.

c. Neural Network

The field of Neural Networks has arisen from diverse sources. That is ranging from understanding and emulating the human brain to broader issues. That is of copying human abilities such as speech and use in various fields. Generally, neural networks consist of layers of interconnected nodes. That each node producing a non-linear function of its input. And input to a node may come from other nodes or directly from the input data. Also, some nodes are identified with the output of the network.

On the basis of this, there are different applications for neural networks present. That involve recognizing patterns and making simple decisions about them.

d. Classification Algorithms in Data Mining

It is one of the Data Mining. That is used to analyze a given data set and takes each instance of it. It assigns this instance to a particular class. Such that classification error will be least. It is used to extract models. That define important data classes within the given data set. Classification is a two-step process.

During the first step, the model is created by applying a classification algorithm. That is on training data set. Then in the second step, the extracted model is tested against a predefined test data set.

That is to measure the model trained performance and accuracy. So classification is the process to assign class label from a data set whose class label is unknown.

d. ID3 Algorithm

This Data Mining Algorithms starts with the original set as the root hub. On every cycle, it emphasizes through every unused attribute of the set and 29 figures. That the entropy of attribute. At that point chooses the attribute. That has the smallest entropy value.

The set is S then split by the selected attribute to produce subsets of the information.

This Data Mining algorithms proceed to recurse on each item in a subset. Also, considering only items never selected before. Recursion on a subset may bring to a halt in one of these cases:

• Every element in the subset belongs to the same class (+ or -), then the node is turned into a leaf and • labeled with the class of the examples

• If there are no more attributes to select but the examples still do not belong to the same class. Then the node is turned into a leaf and labeled with the most common class of the examples in that subset.

• If there are no examples in the subset, then this happens. Whenever parent set found to be matching a specific value of the selected attribute.

• For example, if there was no example matching with marks >=100. Then a leaf is created and is labeled with the most common class of the examples in the parent set. Working steps of Data Mining Algorithms is as follows,

• Calculate the entropy for each attribute using the data set S. • Split the set S into subsets using the attribute for which entropy is minimum. • Construct a decision tree node containing that attribute in a dataset.

• Recurse on each member of subsets using remaining attributes.

e. C4.5 Algorithm

C4.5 is one of the most important Data Mining algorithms, used to produce a decision tree which is an expansion of prior ID3 calculation. It enhances the ID3 algorithm. That is by managing both continuous and discrete properties, missing values. The decision trees created by C4.5. that use for grouping and often referred to as a statistical classifier.

C4.5 creates decision trees from a set of training data same way as an Id3 algorithm. As it is a supervised learning algorithm it requires a set of training examples. That can see as a pair: input object and the desired output value (class). The algorithm analyzes the training set and builds a classifier. That must have the capacity to accurately arrange both training and test cases. A test example is an input object and the algorithm must predict an output value. Consider the sample training data set

S=S1, S2,…Sn which is already classified.

Each sample Si consists of feature vector (x1,i, x2,i, …, xn,i). Where xj represent attributes or features of the sample. The class in which Si falls. At each node of the tree, C4.5 selects one attribute of the data. That most efficiently splits its set of samples into subsets such that it results in one class or the other.

The splitting condition is the normalized information gain. That is a nonsymmetric measure of the difference. The attribute with the highest information gain is chosen to make the decision. General working steps of algorithm is as follows,

Assume all the samples in the list belong to the same class. If it is true, it simply creates a leaf node for the decision tree so that particular class will select.

None of the features provide any information gain. If it is true, C4.5 creates a decision node higher up the tree using the expected value of the class.

f. K Nearest Neighbors Algorithm

The closest neighbor rule distinguishes the classification of an unknown data point. That is on the basis of its closest neighbor whose class is already known.

M. Cover and P. E. Hart purpose k nearest neighbor (KNN). In which nearest neighbor is computed on the basis of estimation of k. That indicates how many nearest neighbors are to consider to characterize.

It makes use of the more than one closest neighbor to determine the class. In which the given data point belongs to and so it is called as KNN. These data samples are needed to be in the memory at the run time.Hence they are referred to as memory based techniques

Bailey and A. K. Jain enhances KNN which is focused on weights. The training points are assigned weights. According to their distances from sample data point. But at the same, computational complexity and memory requirements remain the primary concern.

To overcome memory limitation size of data set is reduced. For this, the repeated patterns. That don't include additional data are also eliminated from training data set.

To further enhance the information focuses which don't influence the result. That are additionally eliminated from training data set.

The NN training data set can organize utilizing different systems. That is to enhance over memory limit of KNN. The KNN implementation can do using ball tree, k- d tree, and orthogonal search tree. The tree-structured training data is further divided into nodes and techniques. Such as NFL and tunable metric divide the training data set according to planes. Using these algorithms we can expand the speed of basic KNN algorithm. Consider that an object is sampled with a set of different attributes.

Assuming its group can determine from its attributes. Also, different algorithms can use to automate the classification process. In pseudo code, k-nearest neighbor algorithm can express,

K ← number of nearest neighbors

For each object Xin the test set do

calculate the distance D(X,Y) between X and every object Y in the training set

neighborhood ← the k neighbors in the training set closest to X

X.class ← SelectClass (neighborhood)

End for h.

Naïve Bayes Algorithm

The Naive Bayes Classifier technique is based on the Bayesian theorem. It is particularly used when the dimensionality of the inputs is high. The Bayesian Classifier is capable of calculating the possible output. That is based on the input. It is also possible to add new raw data at runtime and have a better probabilistic classifier. This classifier considers the presence of a particular feature of a class. That is unrelated to the presence of any other feature when the class variable is given.

For example, a fruit may consider to be an apple if it is red, round. Even if these features depend on each other features of a class. A naive Bayes classifier considers all these properties to contribute to the probability. That it shows this fruit is an apple. Algorithm works as follows,

Bayes theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c).

Naive Bayes classifier considers the effect of

the value of a predictor (x) on a given class (c). That is independent of the values of other predictors.

P(c|x) is the posterior probability of class (target) given predictor (attribute) of class.

P(c) is called the prior probability of class.

P(x|c) is the likelihood which is the probability of predictor of given class.

P(x) is the prior probability of predictor of class.

i. SVM Algorithm

SVM has attracted a great deal of attention in the last decade. It also applied to various domains of applications.

SVMs are used for learning classification, regression or ranking function. SVM is based on statistical learning theory and structural risk minimization principle. And have the aim of determining the location of decision boundaries. It is also known as a hyperplane.

That produces the optimal separation of classes. Thereby creating the largest possible distance between the separating hyperplane. Further, the instances on either side of it have been proven. That is to reduce an upper bound on the expected generalization error.

The efficiency of SVM based does not depend on the dimension of classified entities. Though, SVM is the most robust and accurate classification technique. Also, there are several problems.

The data analysis in SVM is based on convex quadratic programming. Also, expensive, as solving quadratic programming methods. That need large matrix operations as well as time-consuming numericalcomputations.

Training time for SVM scales in the number of examples. So researchers strive all the time for more efficient training algorithm. That resulting in several variant based algorithm.

SVM can also extend to learn non-linear decision functions. That is by first projecting the input data onto a high-dimensional feature space. As by using kernel functions and formulating a linear classification problem. The resulting feature space is much larger than the size of a dataset. That is not possible to store on popular computers.

Investigation of this issues leads to several decomposition based algorithms. The basic idea of decomposition method is to split the variables into two parts:

a set of free variables called as a working set. That can update in each iteration and set of fixed variables. That are fix during a particular. Now, this procedure have to repeat until the termination conditions are met

The SVM was developed for binary classification. And it is not simple to extend it for multi- class classification problem. The basic idea to apply multiclassification to SVM. That is to decompose the multi-class problems into several two-class problems. That can address using several SVMs.

J. ANN Algorithm

This is the types of computer architecture inspire by biological neural networks. They are used to approximate functions. That can depend on a large number of inputs and are generally unknown.

They are presented as systems of interconnected “neurons”. That can compute values from inputs. Also, they are capable of machine learning as well as pattern recognition. Due to their adaptive nature.

An artificial neural network operates by creating connections between many different processing elements. That each corresponding to a single neuron 35 in a biological brain. These neurons may actually construct or simulate by a digital computer system.

Each neuron takes many input signals. Then based on an internal weighting. That produces a single output signal that is sent as input to another neuron. The neurons are interconnected and organized into different layers. The input layer receives the input and the output layer produces the final output.

In general, one or more hidden layers are sandwiched between the two. This structure makes it impossible to forecast or know the exact flow of data. Artificial neural networks start out with randomized weights for all their neurons. This means that they need to train to solve the particular problem for which they are proposed. A back-propagation ANN is trained by humans to perform specific tasks.

During the training period, we can test whether the ANN’s output is correct by observing a pattern. If it’s correct the neural weightings produce that output is reinforced. if the output is incorrect, those weightings responsible diminish. Implemented on a single computer, a network is slower than more traditional solutions. The ANN’s parallel nature allows it to built using many

processors. That gives a great speed advantage at very little development cost. The parallel architecture allows ANNs to process amounts of data very in less time. It deals with large continuous streams of information. Such as speech recognition or machine sensor data. ANNs can operate faster as compared to other algorithms.

A decision tree is a predictive machine-learning model. That decides the target value of a new sample. That based on various attribute values of the available data. The internal nodes of a decision tree denote the different attributes

Also, the branches between the nodes tell us the possible values. That these attributes can have in the observed samples. While the terminal nodes tell us the final value of the dependent variable.

The attribute is to predict is known as the dependent variable. Since its value depends upon, the values of all the other attributes. The other attributes, which help in predicting the value of the dependent variable. That are the independent variables in the dataset. The J48 Decision tree classifier follows the following simple algorithm. To classify a new item, it first needs to create a decision tree. That based on the attribute values of the available training data.

l. Support Vector Machines

Support Vector Machines are supervised learning methods. That used for classification, as well as regression. The advantage of this is that they can make use of certain kernels to transform the problem. Such that we can apply linear classification techniques to non-linear data. Applying the kernel equations. That arranges the data instances in a way within the multi- dimensional space. That there is a hyperplane that separates data instances of one kind from those of another.

The kernel equations may be any function. That transforms the non-separable data in one domain into another domain. In which the instances become separable. Kernel equations may be linear, quadratic, Gaussian, or anything else. That achieves this particular purpose.

Once we manage to divide the data into two distinct categories, our aim is to get the best hyperplane. That is to separate the two types of instances. This hyperplane is important, it decides the target variable value for future predictions.

We should decide upon a hyperplane that maximizes the margin. That is between the support vectors on either side of the plane.

Support vectors are those instances that are either on the separating planes. The explanatory diagrams that follow will make these ideas a little more clear.

In Support Vector Machines the data need to be separate to be binary. Even if the data is not binary, these machines handle it as though it is. Further completes the analysis through a series of binary assessments on the data.

M. Sense Clusters (an adaptation of the K-means clustering algorithm) We have made use of Sense Clusters to classify the email messages. Sense Cluster available package of Perl programs. As it was developed at the University of Minnesota Duluth. That we use for automatic text and document classification. The advantage of Sense Clusters is that it does not need any training data; It makes use of unsupervised learning methods to classify the available data.

Now, particularly in this section will understand the K-means clustering algorithm. That has been used in Sense Clusters. Clustering is the process in which we divide the available data. That instances of a given number of subgroups. These sub-groups are clusters, and hence the name “Clustering”.

To put it, the K-means algorithm outlines a method. That is to cluster a particular set of instances into K different clusters. Where K is a positive integer. It should notice K-means clustering algorithm requires a number of clusters from the user. It cannot identify the number of clusters by itself. However, Sense Clusters has the facility of identifying the number of clusters.

PROPOSED SYSTEM

The goal of this work is to compare the accuracy of bitcoin price in USD prediction based on Long Short-term Memory (LSTM) network with self attention. Real-time price data is collected by Pycurl from Bitfine. LSTM model is implemented by Keras and TensorFlow. The proposed model used in this work is mainly to present a classical comparison of time series forecasting, as expected, it could make efficient prediction limited in short-time interval, and the outcome depends on the time period. The LSTM could reach a better performance, with extra, indispensable time for model training, especially via CPU.

- 1.Data Acquisition
- 2.Pre-Processing
- 3.Training
- 4.Prediction and Evaluation

Data set collection

The bitcoin data set form the year of 2013 to 2021 is collected from koggle website with attributes states. This data set contains complete information of staring price , closing price, open and end price .

Pre-Processing

Preprocessing is involves normalization of open and closeprice values .

Training

To split the data we will be using train_test_split from sklearn. train_test_split randomly distributes bitcoin data into training and testing set according to the ratio provided.

Prediction

Using these structured data and deep learning models to predict bitcoin which is an important issue in worldwide. In order to solve the problem of low accuracy of Long-Short Term Memory (LSTM) model in bitcoin prediction, this chapter presented a proposed model of LSTM model based on attention mechanism. The proposed model can learn the importance of each past value to the current value from the long sequence of cost data at the past moment, which makes it possible to extract more valuable features. Constructed a dataset using the bitcoin data in the core section of Wuhan for experiments, and the performance of the improved model is compared with the original LSTM model.

CONSTRUCTION OF ATTENTION-LSTM MODEL

LSTM Model

We will briefly introduce the principle of LSTM model. LSTM is a kind of recurrent neural network

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right)$$
$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right)$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_{t'}$$
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$
$$h_t = o_t \cdot \tan \mathrm{h}\left(C_t\right)$$

Among them, ft determines how much information we want to discard. it determines how much new information we should add. ot determines how much information we want to output. xt is the input at time t. ht−1 is the output of the previous gate, Wf , Wi, Wc and Wo is the weight, bf , bi , bc and bo is the bias, Ct−1 is the cell state at the previous moment, Ct is the cell state at the current

moment.

Attention-LSTM Model Since the model is difficult to learn information at a time far from the current time, and it may be important for the current value. To overcome the weakness, we tried to add an attention layer to the LSTM network. Referring to the attention implementation steps of [9], we can apply it to the LSTM model.

Among them, Xi, i ∈ (1, n) is the input, hi is the intermediate output result of each cell, hi are input into each attention model as H, and the elements of the next layer h i are used as H to calculate the similarity and weight coefficient, and finally get the attention coefficient.

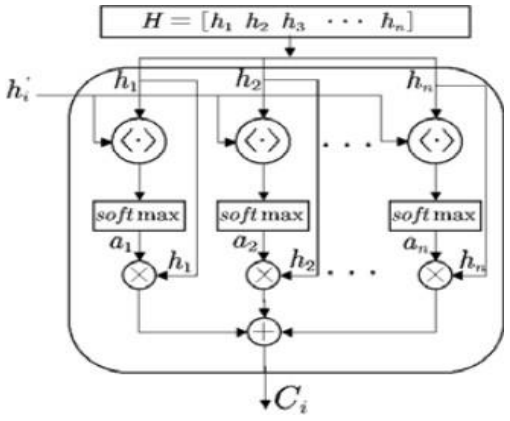

Figure The internal structure of attention model

Where, represents the dot product operation, which is used to calculate the similarity between the current element and the intermediate output result in the previous layer, and then normalized by the softmax function to obtain the corresponding weight coefficient ai. Finally, a weighted summation operation is performed to obtain the Attention value Ci. The formula used in the attention layer is as follows

$$H = [h_1\ h_2\ \cdots\ h_n]$$

$$H_i' = \left[h_i'\ h_i'\ \cdots\ h_i'\right]$$

$$sim_i = H_i' \cdot H^T$$

$$a_i = \frac{e^{sim_i}}{\sum_{j=1}^{L_h} e^{sim_j}}$$

$$C_i = \sum_{j=1}^{L_h} a_i \cdot h_j$$

In the above equations, uses vector H and H to calculate similarityto obtain weights, uses the softmax function to normalize the weight, uses the normalized weight ai and hi weighted sum. The result of weighted summation is the attention weight value Ci.The implementation of the Attention layer is to retain the intermediate output results of the input sequence by the LSTM encoder, and then calculate the similarity between the intermediate output results of the previous layer and the current output to obtain the weight factor, and finally obtain the attention coefficient.

Proposed algorithm 1

Input: covid-19 data

Output: A trained Attention-LSTM model.

1: Construct a dataset with a sliding time window,including Xt andZt.

2: Normalization Xt and Zt.

3: Input features matrix Xt and current disease vector Zt to A-LSTM network.

4: while training epoch does not reach the set value do

5: Put (Xt, Zt) into the Attention-LSTM network for forward propagation.

6: Calculate the attention weight corresponding to each element 7: Generate Yt

8: Calculate mean square error.

9: Use RMSProp update weights for A-LSTM network.10:

end while

11: return A trained Attention-LSTM model.

The performance of the LSTM model based on the attention mechanism is verified for long time series and large prediction lag time. All prediction models use the same data set and are built in the same way. In the LSTM model, we set 2 hidden layers, the number of hidden layer neurons is 64 and 64, and the learning rate is 0.05. The network optimizer is also RMSprop. The process of AttentionLSTM model training is shown in Algorithm 1.

Evaluation

Confusion Matrix

Confusion Matrix as the name suggests gives us a matrix as output and describes the complete performance of the model. Lets assume we have a binary classification problem. We have some samples belonging to two classes : YES or NO. Also, we have our own classifier which predicts a class for a given input sample. On testing our model on 165 samples ,we get the following result.

n=165	**Predicted: NO**	**Predicted: YES**
Actual: NO	50	10
Actual: YES	5	100

Confusion Matrix

There are 4 important terms :

- **True Positives :** The cases in which we predicted YES and the actual output was also YES.
- **True Negatives :** The cases in which we predicted NO and

the actual output was NO.

• **False Positives :** The cases in which we predicted YES and the actual output was NO.

• **False Negatives :** The cases in which we predicted NO andthe actual output was YES.

Accuracy for the matrix can be calculated by taking average ofthe values lying across the "main diagonal" i.e

$$Accuracy = \frac{TruePositive + TrueNegative}{TotalSample}$$

$$\therefore Accuracy = \frac{100 + 50}{165} = 0.91$$

Confusion Matrix forms the basis for the other types of metrics.

Area Under Curve

Area Under Curve(AUC) is one of the most widely used metrics for evaluation. It is used for binary classification problem. AUC of a classifier is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example. Before defining AUC, let us understand two basic terms :

• **True Positive Rate (Sensitivity) :** True Positive Rate is

defined as TP/ (FN+TP). True Positive Rate corresponds to the proportion of positive data points that are correctly considered as positive, with respect to all positive data points.

$$TruePositiveRate = \frac{TruePositive}{FalseNegative + TruePositive}$$

• **True Negative Rate (Specificity) :** True Negative Rate is defined as TN / (FP+TN). False Positive Rate corresponds to the proportion of negative data points that are correctly considered as negative, with respect to all negative data points.

$$TrueNegativeRate = \frac{TrueNegative}{TrueNegative + FalsePositive}$$

• **False Positive Rate :** False Positive Rate is defined as FP / (FP+TN). False Positive Rate corresponds to the proportion of negative data points that are mistakenly considered as positive, with respect to all negative data points.

$$FalsePositiveRate = \frac{FalsePositive}{TrueNegative + FalsePositive}$$

False Positive Rate and True Positive Rate both have values in the range [0, 1]. FPR and TPR both are computed at varying threshold values such as (0.00, 0.02, 0.04, …., 1.00) and a graphis drawn.

SOFTWARE SPECIFICATIONSOFTWARE

REQUIRED:

• IDLE 1.7

• PYTHON 1.7.6

HARDWARE REQUIRED:

• System : Windows Xp Professional Service Pack 2

• Processor : Up to 1.5 GHz

• Memory : Up to 512 MB RAM

5.1 PYTHON

The Python language had a humble beginning in the late 1980s when a Dutchman Guido Von Rossum started working on a fun project, which would be a successor to ABC language with better exception handling and capability to interface with OS Amoeba at Centrum Wiskunde and Informatica. It first appeared in 1991. Python 2.0 was released in the year 2000 and Python 3.0 was released in the year 2008. The language was named Python after the famous British television comedy show Monty Python's Flying Circus, which was one of Guido's favorite television programmes. Here we will see why Python has suddenly influenced our lives and the various applications that use Python and its implementations.

5.1.1 Why Python?

Now you might be suddenly bogged with the question, why Python? According to Institute of Electrical and Electronics Engineers (IEEE) 2016 ranking Python ranked third after C and Java. As per Indeed.com's data of 2016, the Python job market search ranked fifth. Clearly, all the data points to the ever rising demand in the job market for Python. It's a cool language if you

want to learn just for fun or if you want to build your career

around Python, you will adore the language. At school level, many schools have started including Python programming for kids. With new technologies taking the market by surprise Python has been playing a dominant role. Whether it is cloud platform, mobile app development, Big Data, IoT with Raspberry Pi, or the new Block chain technology.

Some key features of the language are:

• Python programs can run on any platform, you can carry code created in Windows machine and run it on Mac or Linux

• Python has inbuilt large library with prebuilt and portable functionality, also known as the standard library

• Python is an expressive language

• Python is free and open source

• Python code is about one third of the size of equivalent C++ and Java code

• Python can be both dynamically and strongly typed--dynamically typed, there is no need to define the type (int or float) of the variable

Python applications

One of the most famous platforms where Python is extensively used is YouTube. The other places where you will find Python being extensively used are the special effects in Hollywood movies, drug evolution and discovery, traffic 48 control systems, ERP

systems, cloud hosting, e-commerce platform, CRM systems, and whatever field you can think of.

Versions

At the time of writing this book, two main versions of the Python programming language were available in the market, which are Python 2.x and Python 3.x. The stable release as of writing the book were Python 2.7.13 and Python 3.6.0.

Implementations of Python

Major implementations include CPython, Jython, IronPython, MicroPython, and PyPy.

5.1.2 Installation

Here we will look forward to the installation of Python on three

different OS platforms, namely, Windows, Linux, and Mac OS. Let's begin with the Windows platform

CONCLUSION

Digital economy growing enormously with high market capital due to the emergence of information technology. Bitcoin, one the major crypto currency, with highest market capital among all other crypto-currencies. Prediction of prices accurately and price projects helps investor and traders, aiming this the proposed work for Bitcoin price prediction with Long Short Term Memory (LSTM) models is done. Price project for month through Long Short Term Memory (LSTM) models is provided in the work. Experimental results suggests that LSTM outperforms the other on price prediction

REFERENCES

Li, L., Arab, A., Liu, J., Liu, J., & Han, Z. (2019). Bitcoin

Options Pricing Using LSTM-Based Prediction Model and Blockchain Statistics. 2019 IEEE International Conference on

Blockchain (Blockchain).

Ali, M., & Shatabda, S. (2020). A Data Selection Methodology to Train Linear Regression Model to Predict Bitcoin Price. 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT).

Wu, C.-H., Lu, C.-C., Ma, Y.-F., & Lu, R.-S. (2018). A New Forecasting Framework for Bitcoin Price with LSTM. 2018 IEEE International Conference on Data Mining Workshops (ICDMW).

Jang, H., & Lee, J. (2018). An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. IEEE Access, 6, 5427–5437.

Serafini, G., Yi, P., Zhang, Q., Brambilla, M., Wang, J., Hu, Y., & Li, B. (2020). Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches. 2020 International Joint Conference on Neural Networks (IJCNN).

Anupriya, & Garg, S. (2018). Autoregressive Integrated Moving Average Model based Prediction of Bitcoin Close Price. 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT).

Singh, D., Kumar, L., & Jain, S. (2019). User ReputationAnalysis for effective Trading on Bitcoin Network. 2019 Twelfth International Conference on Contemporary Computing (IC3).

Aggarwal, A., Gupta, I., Garg, N., & Goel, A. (2019). Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction. 2019 Twelfth International Conference on Contemporary Computing (IC3).

Demir, A., Akilotu, B. N., Kadiroglu, Z., & Sengur, A. (2019). Bitcoin Price Prediction Using Machine Learning Methods. 2019 1st International Informatics and Software Engineering Conference (UBMYK).

Velankar, S., Valecha, S., & Maji, S. (2018). Bitcoin price prediction using machine learning. 2018 20th International Conference on Advanced Communication Technology (ICACT).

Karasu, S., Altan, A., Sarac, Z., & Hacioglu, R. (2018). Prediction of Bitcoin prices with machine learning methods using time series data. 2018 26th Signal Processing and Communications Applications Conference (SIU)