CRYPTOCURRENCY TRADING PROJECTION

ANNAPAREDDY GANESHREDDY - 50442295

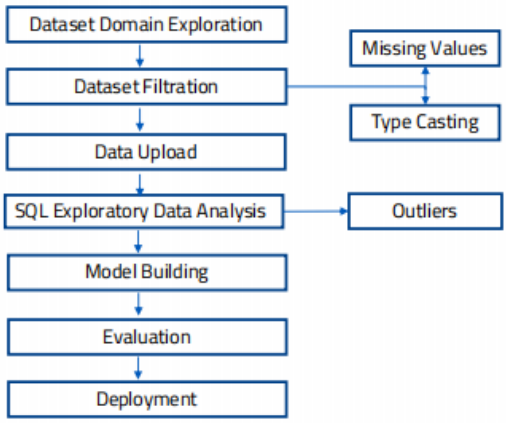
ABSTRACT :

Dynamic trading data concerning financial sector populates an uncertain behaviour over time. Understanding the patterns from the time-series historic data will benefit the business decisions in order to build a productive marketing strategy. Thus, an aim of the project is to make a cryptocurrency trading prediction by using machine learning algorithms. The data is collected from Kaggle open source platform which has about eight columns and more than 4800000 rows providing the historic information about the cryptocurrency stock prices. The main focus will be on the closing price and the corresponding time frame's moving average. These statistics will be a foundation to regard the strategy of either a buy or sell of the stock. The data-set will be split as a training content (to build a probabilistic model) and the testing content (to evaluate the developed model). Most of the feature engineering such as dealing with missing values, outliers and window frame buy sell computation will be handled through SQL. Then the updated dataset will be utilised to perform an exploratory data analysis before Implementing the machine learning algorithms. With this intention, few of the python library packages will be utilized for model building and data visualization.

INTRODUCTION :

Cryptocurrencies have been around for a long time and are quite volatile. Prices have varied dramatically over the last few years. As a result, putting up a trading strategy is incredibly tough. The goal of this specific project is to develop a model that will aid in the estimation of buy and sell trading. The current dataset CSV file includes bitcoin exchanges from January 2012 through September 2020. A target variable is constructed using the moving average idea.

FLOWCHART :



MOVING AVERAGE :

It's a time series approach for studying and predicting data patterns. The short-term moving average (SMA1) and long-term moving average (LMA1) are calculated using closing feature and it's mean (SMA2).

SMA 1 and SMA2 are used to create the a target variable. For the trading strategy, we will purchase if SMA1 > SMA2 and the signal is 1, and we will sell if SMA1 > SMA2 and the signal is 0.

DATA SUMMARY :

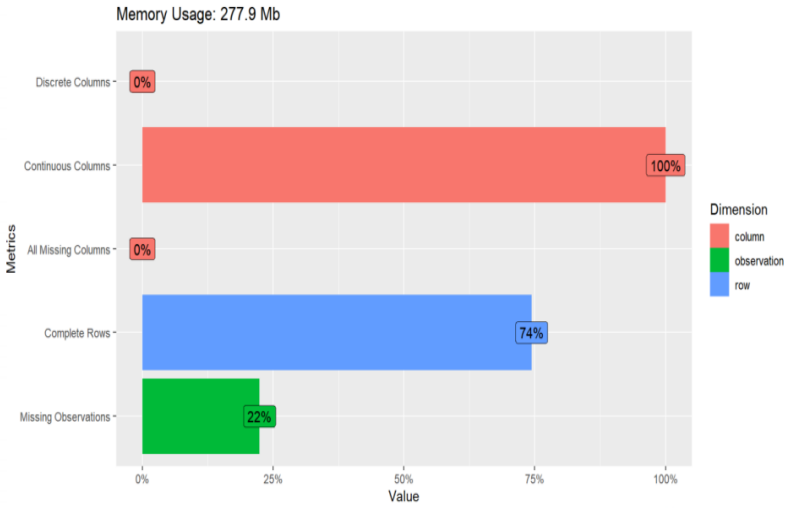
ROWS : 4857377

COLUMNS : 8

MISSING VALUES : 1243608*8

TOTAL OBSERVATIONS : 4857377*8

DISCRETE COLUMN : TARGET VARIABLE



DATA MODEL :

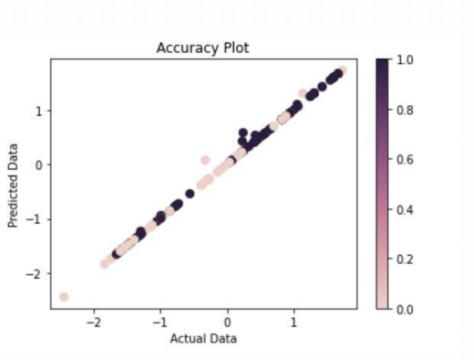
The data in the CSV file is a collection of strings that we separated into columns, filled in missing values, and converted to float. We constructed a historical data table and added the data using execute many, which includes 8 columns and 3.6 million rows, after creating a list of tuples. The target table was created using the historical data table, which has roughly 5 columns. Using their near feature and moving average, we calculated SMA1 and SMA2 in the Target table. SMA1 and SMA2 are used to create the signal.

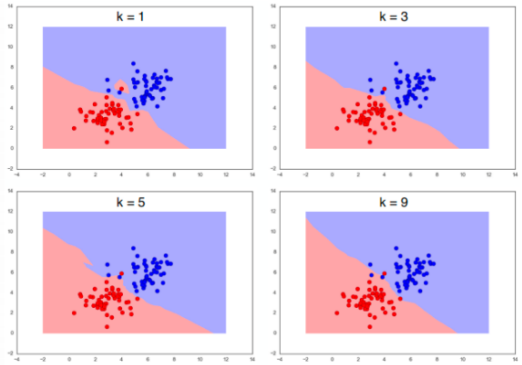
- We have used 3 models, they are

K-nearest neighbour :

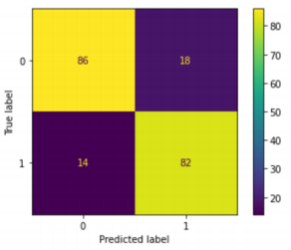
- With a k - value of 9, the k-nearest neighbour model obtains an excellent accuracy of 92.36 percent. When this k value is adjusted, the model's accuracy improves.

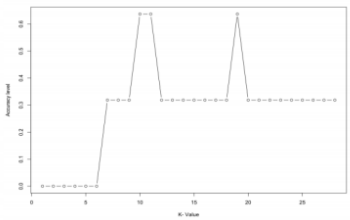
- For k = 9, 11, and 19, the accuracy plot shows that the model achieves good accuracy.





ACCURACY METRICS

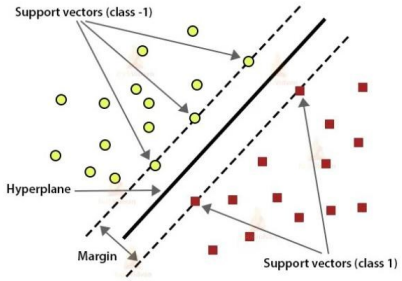




ACCURACY PLOT

SUPPORT VECTOR MACHINE :

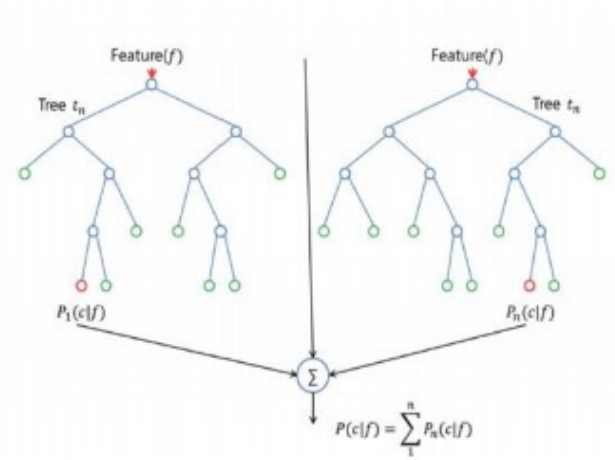
- Support vectors are the locations that are closest to the decision border and so define it.
- The maximum margin classifier's purpose is to find the linear boundary that minimizes the total distance between the line and the nearest point in each class.
- Linear Kernel, Polynomial Kernel, Gaussian Radial Basis Function (RBF), and Sigmoid Kernel are used to tune SVMs. The linear with a cost of ten has been chosen.
- Linear Kernel, Polynomial Kernel, Gaussian Radial Basis Function (RBF), and Sigmoid Kernel are used to tune SVMs. The linear with a cost of ten has been chosen.
- A perfect hyperplane with several support vectors running across the edge.



	precision	recall	f1-score	support
0	1.00	0.93	0.97	104
1	0.93	1.00	0.96	96
accuracy			0.96	200
macro avg	0.97	0.97	0.96	200
weighted avg	0.97	0.96	0.97	200

RANDOM FOREST :

- Bagging is enhanced by random forest, which creates ensembles of separate decision trees.
- As with bagging, the initial step is to train each tree on distinct samples from the data.
- As with bagging, the initial step is to train each tree on distinct samples from the data.
- The number of predictors to randomly choose at each split, the number of trees, and the minimum leaf node size are all parameters to tweak.



QUALITATIVE ANALYSIS(RESULTS) :

METHOD	*PRECISION*	*RECALL*	*F1-SCORE*
RANDOM FOREST	89%	0.89	0.89
SUPPORT VECTOR MACHINE	100%	0.93	0.97
K-NEAREST NEIGHBOUR	86%	0.83	0.84

CONCLUSION :

Such a Cryptocurrency Trading Projection can help in making trading strategy a little bit easy for people to look and take a make move on to buy the cryptocurrency or not.