*Prediction of Tesla stock using machine learning.*

Nallani Lakshmi Sri

*School of computer science and Artificial intelligence..*

*SR University.*

Warangal, India.

email:2203a52230@sru.edu.in

***ABSTRACT*:**

**Technological advancements have revolutionized stock market forecasting, with machine learning methods proving more accurate than traditional statistical approaches. Comparing various models, it was found that integrated algorithms like Random Forest exhibit superior performance in predicting Tesla's stock closing prices. These methods mitigate risks associated with investing while maximizing dividends, crucial for effective resource allocation and macroeconomic expansion in a country's financial market.**

***Keywords— Stock Markets Forecasting, Machine Learning,***

***Traditional statistical Approaches, closing prices , Investing.***

1. INTRODUCTION

Stock markets, comprising stockbrokers and dealers, facilitate trading of stock shares, enhancing liquidity and investor appeal. However, stock market investment carries inherent risks due to the potential for rapid value fluctuations, making forecasting stock prices a challenging endeavor. Statistical methodologies, being linear, often fail to account for sudden spikes or drops in stock prices, proving insufficiently accurate for the volatile and unpredictable nature of stock data. Various approaches, including It had Linear Models which consist of Regression, SVM, Random Forest, LSTM, and ARIMA models. Mixed Ownership Stock exchanges, like London and New York may be examples of this cater to publicly listed and privately held companies' shares, while investments via mixed-ownership share trade involve ordinary shares exchangeable publicly on rare occasions.

In an experiment utilizing Tesla stock data from 29 June , 2010, to 12 July, 2022, data is collected and processed from Kaggle, removing null values and normalizing it for model compatibility. Matlab is utilized to implement the aforementioned models, with performance comparison being a key objective. The dataset comprises 6 columns and 3031 rows, with 2121 data points allocated for training and 909 for testing. Training data spanned from 29 June, 2010, to 7 September , 2018, while the test data extends from 10 July, 2010, to 12 July, 2022. Data preparation involved removed extraneous information and scaling variable ranges for fair comparison across methodologies. The experiment aims to assess model accuracy and provide updated recommendations for future research and project expansion in stock price prediction.

|  |
| --- |
| XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE |

Data from Kaggle dated from June 29, 2010, until July 12, 2022 and used Tesla stick information was applied in the experiment which included comprehensive pre-processing step to handle null values and normalize the data. It was implemented by various types of machine learning and statistic models namely Lineal Regression, SVM, Random Forrest, LSTM and ARIMA which were used for predicting shorts rates. The way the databases are split to train and test set is 2121 for the training data and 909 for the testing data. This is to make sure that the evaluation is balanced between different model performances. The use of evaluation matrices such as MAE or RMSE might be used in the quest to measure the accuracy of the models. The purpose of the experiment is to show the updated recommendations that the project can employ for further research and adding more data to the stock models, after discovering the know-how with the extreme data of Tesla stock.

1. LITERATURE SURVEY

Luren Dai. The article attempt to use machine learning like k-nearest Neighbors algorithm and time series to analyze Tesla's stock price. Kayako . The paper aims at undertaking the study of such machine learning techniques like Linear Regression, Polynomial Regression, XGBoost, ARIMA, Prophet, and LSTM to understand and make predictions about Tesla's stock value, where LSTM is least error prone.Evaluation ways such as Mean square error, Root Mean Square Error, Mean Absolute Error, and Accuracy of trend prediction are introduced to assess the prediction accuracy of each model. Linear regression, polynomial regression, and XGBoost are among the models used in the analysis. Tesla stock data from 2016 to 2021 sourced from Kaggle is utilized in the study.

Indrani Bhattacharjee. A comparison study was undertaken on classic stats methods and AI sorcery for foretelling stock costs, uncovering that neural web model are the peak precision. The exponential Smoothing method uses a smoothing constant to maximize prediction accuracy from the last prediction. The Weighted Moving Average (WMA) method uses weights with previous values to predict future values, with different WMA units showing varying prediction errors.

Dr. Chaitanya Kishore Reddy.The PDF focuses on analyzing Tesla stock prices using machine learning algorithms. Python was chosen as the language for the project, utilizing linear regression for stock price forecasting. The document includes figures illustrating data processing, linear regression, and data splitting.

Hao Li. The PDF compares the performance of four models in predicting Tesla stock prices Linear regression stands out with the highest R-squared value of 0.85 for Tesla stock from 2020 to 2022 Super vector regression (SVR) is a part of Support vector Machine (SVM) and is popular for its wide range of applications in regression estimation

Sasha S. Yamada. The PDF discusses the application of machine learning models, including Kalman filtering and LSTM architectures, for stock price prediction over a 10-year range. It explores the performance of different algorithms in predicting future stock prices, with a focus on both lowvolatility (e.g., Microsoft) and high-volatility (e.g., Tesla) stocks. The study uses historical stock prices of individual technology sector stocks (e.g., MSFT, AAPL, TSLA) and market indices (e.g., NASDAQ, S&P 500) to train and test the machine learning models

Subhash Chand Agrawal.Stock market prediction by Machine Learning algorithms is crucial for investors to estimate future stock values accurately. Factors like demand and supply, corporate results, and popularity influence stock values, making accurate predictions essential for decisionmaking. Linear regression is commonly used in stock market forecasting, while deep learning-based non-linear regression models outperform linear regression in capturing stock market data changes.

Jianyao Li.LSTM variants like Vanilla, Stacked, Bidirectional, and CNN LSTM are compared in predicting Tesla's stock price Stacked LSTM, with multiple hidden layers, shows promising results in stock prediction.LSTM, a type of RNN, retains prior-state information and is crucial for learning long-term dependencies in stock market forecasting

Hendri Mahmud Nawawi.The PDF focuses on predicting Tesla's stock prices using deep learning models like MLP and LSTM based on data from investing.com and Yahoo Finance. The models are evaluated using metrics like MAE, MSE, and RMSE to assess the accuracy of the predictions. The study aims to forecast future Tesla stock prices by analyzing historical data from 2015 to 2021, emphasizing the importance of technical indicators and neural networks for accurate predictions.

Michele Coiro. The PDF examines Tesla's stock investment potential based on historical data, social media sentiment, and investor behaviorThe article says that Tesla is seen positively on social media and in old-fashioned news, route to political backing and absence of challengers. It recommends using text analytics and AI for a deeper look at social media data about Tesla's stocks. According to the Twitter study, there aren't many negative tweets about Tesla, but a lot of people seem thrilled and proud to invest in the company.

Amey Bhadamkar. The PDF focus on predict of Tesla Inc. sock value by sentiment analysis and delves into the correlation between tweets by Elon Musk and Tesla's stock price. Methods of Machine Learning, especially sentiment analysis on social networks, are implanted to enrich market foretelling models. PDF debates the adoption of the Facebook Prophet technique for prophesying time series data, accentuate precision, velocity, and flexibility.

Gustav Edma. Twitter data analysis was used to predict intraday stock return for Tesla, Inc., utilizing sentiment extraction methods like VADER and SVM. The discussion dimension of Twitter data, including retweets and volume of posts, was found to positively impact trading activity and predict stock returns. Data cleaning procedures were implemented to adjust for time zone and calendar differences between Twitter and Tesla stock data sets.

III. PROPOSED APPROACH

1. *Overview:*

The prevalent stock value prophetic setup joins a multifarious approach, fishing on basic analysis, technical analysis, and market emotion analysis. Basics analysis delve into a corporation's financial health, examining factors such as revenue, earning, and growth potential. On another foot, technical analysis scrutinizes historical and pricing data for identifying patterns and trends. Market sentiment analysis plays a pivotal role in appraising public sentiment and marketplace trends, often leveraging social media and news sources. Furthermore, algorithmic trading and quantitative analysis employ automatic trading tactics and mathematical models to capitalize on market inefficiencies. Despite the widespread use of those approaches, the innate unpredictability of financial markets limits the preciseness of prophets. To tackle this dilemma, machine learning and artificial intelligence (AI) methods are progressively integrated into prophetic models, aiming to enhance accuracy by leveraging sizeable amounts of data and detecting complex patterns that may evade traditional approaches.

Creating a sock indication system involve several kestle stepos. First, historic stock prized datums is being collecting and pre-processed, includes datums cleansing' and feature engineering'. Mechanism learning model, like diagonal regression or more advancement uns, are then training on a parting of datums and testing for predictively accurate. External factoros like new sentiment and market trends may be incorporated to enhance predictions! It's a complex process required experimentation, model fine-tuning, and bustling testing. In these simplifying example, we using diagonal regression to predicte' Tesla's stock prized based on history datums, but real-world systems employ more sophisticated modelos and extensible datums pre-processing for more accurately predilections.

1. *Features Extraction:*  Features extracteds from informations regarding the stocks price predictin' system and 'em goals, as architectural frameworks, include!

*a. Objectives:*  - Accurate predictiveness is super crucial! It helps the investors and analysts big time in handling those risky investments in the ever-changing market vibes with all the ups and downs.

* Comparison of Performance: Compare regression techniques to find best method for predicting stock prices.
* Empowerin' daters wit' data-driven outsights can greatly empower them when makin' decisions!
* Adaptability: Foster continuous adaptability to changing market conditions.

*b. Architectural Framework Phases:*

* Data Process: Collecting like it's cleaning historical data of stock prices to be quality and reliability assured!
* Model Developing: Creating and delicate-tune models of regression (like Linear, Ridges, Lassos, KNN) while model considerations are they.
* Assessing: Evaluation of accuracy and performances of regression models for the purposes of comparative analysis conducting.

- Deployment: Provide a user-friendly interface for accessing predictions, emphasizing data security measures to protect financial information.

These features encompass the system's objectives, processes, and key considerations, outlining its functionality and framework for development and implementation.

IV. EXPERIMENTAL RESULTS

In this Section ,we explain the details of experimental and dataset results , then the results of our experiments.

*Experimental setup:*

For this project, I have obtained my dataset from Kaggle. This dataset contains 2814 rows of data and 7 columns (features) that we could focus on to build our prediction model i.e., I have used 7 attributes to predict the rise in stock of Tesla.

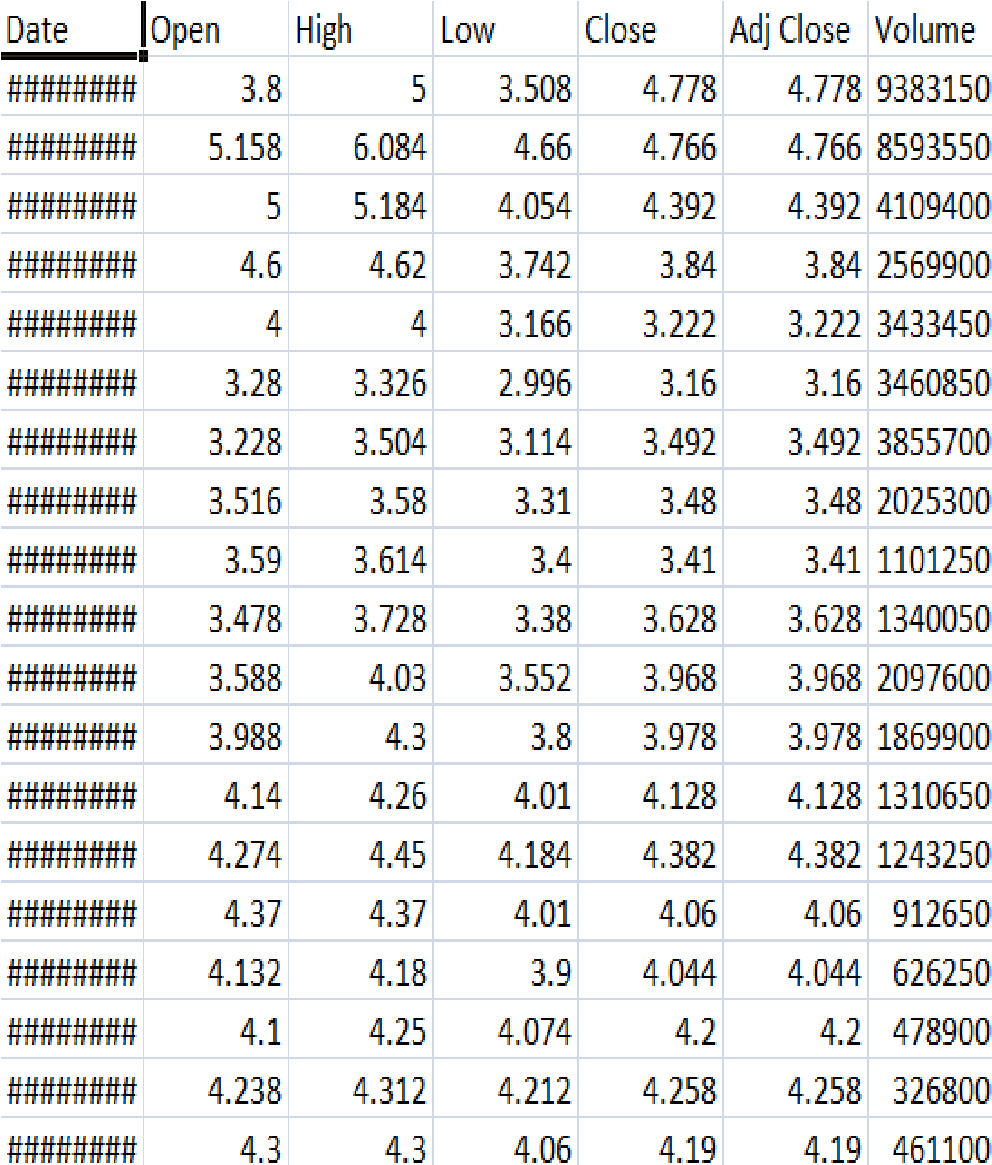


Table 1: Data set

1. Date: This column features the date of observation for each data point. They represented the day when the stock market started and as well as the time records were recorded.
2. Open: The "Open" price is the "TSLA" stock opening price on a trading day. That is the amount of money for which a security was offered on the market recently.
3. High: The "High", which stands for the highest price that TSLA shares managed to reach in the current trading day.
4. Low: The "low" denominated 142.59 is the lowest price where TSLA stock moved during trading today.
5. Close: The "Close" price is noted as TSLA's closing stock price at the end of the trading day. It is the latest price at which an order occurred, and it represents the last transaction that took place before the market closed.
6. Adj Close: The "Adjusted Close" price is the same as the close or closing price but will be adjusted when there is a change between the stock's value and any certain factors such as dividends and stock splits. It is one of the most effective historical performances analysis in today's times.
7. Volume: The "Volume" column indicates with which the given day’s activity on the TSLA stock was measured. It denotes the aggregate percentage of shares open during the day's trading session

These parameters are essential for conducting technical analysis, building predictive models, and making investment decisions. For stock prediction, analysts and data scientists often use historical data like this to develop models that attempt to forecast future stock prices or to identify patterns and trends in the market.

**Data Cleaning:**

Data cleaning consists of removing the incorrect, incomplete, inaccurate, irrelevant, or missing data from the data then modifying or replacing them according to the needs of the data! Data cleaning being one of the first elements of data science, is considered fundamental by the data science experts. Data is the most valuable thing for data analysis and machine learning. In computing and Business, Data is the key aspect of all areas. As we deal with today's world of data, we are likely to encounter missing, inaccurate, or erroneous values. If da information is corrupted, the process may be disrupted or the results become unreliable. My dataset has no string which means that I do not need to clean the data.

**DATA VISUALIZATION:**

Representing data with visualization means to depict 1data clearly in a pictorial or graphical representation. Data visualisation tools are a means of understanding various themes like showing trends, the patterns and detection of outliers in data. The data visualization tools and technologies will be playing key role when it comes to analyzing big amounts of data and making evidence-driven decisions. Visualization usage as a tool to cognize fought throughout centuries. More often than not, the type of data visualization used are charts, tables, graphs, maps, dashboards, and so on. However, the figures of our data set is shown in plots from each feature to price.

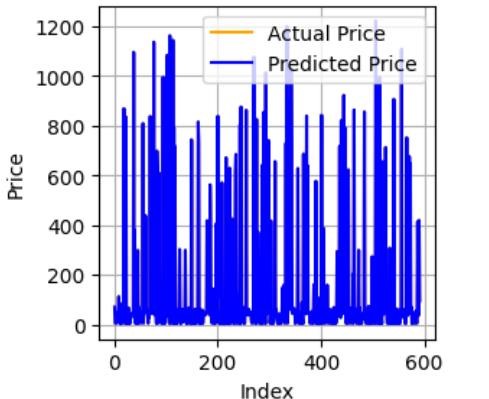
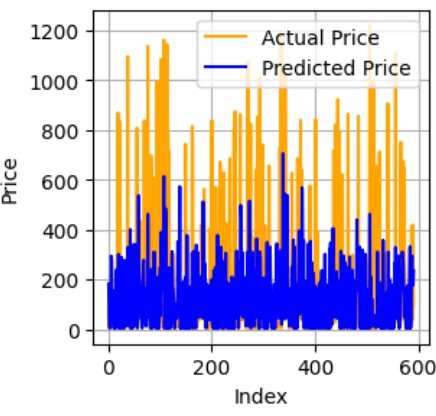


Figure1.Graph: Open price vs High price

(LINEAR REGRESSION) Figure 4. Graph: Open price vs Adj Close price

(KNN)

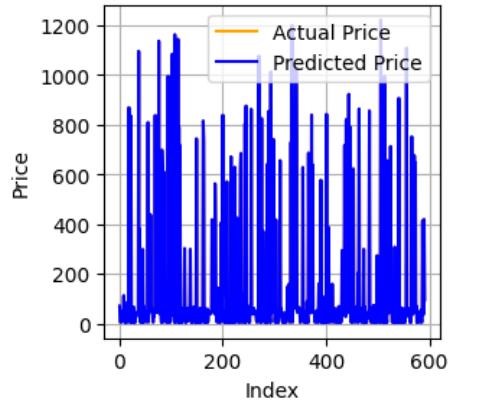
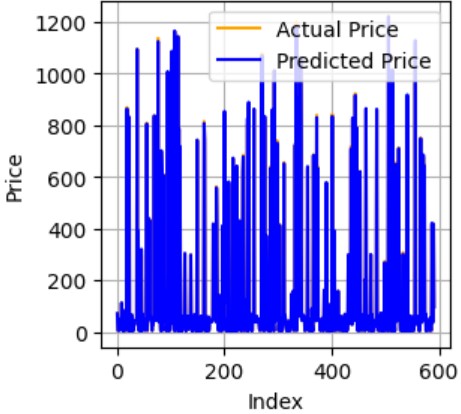
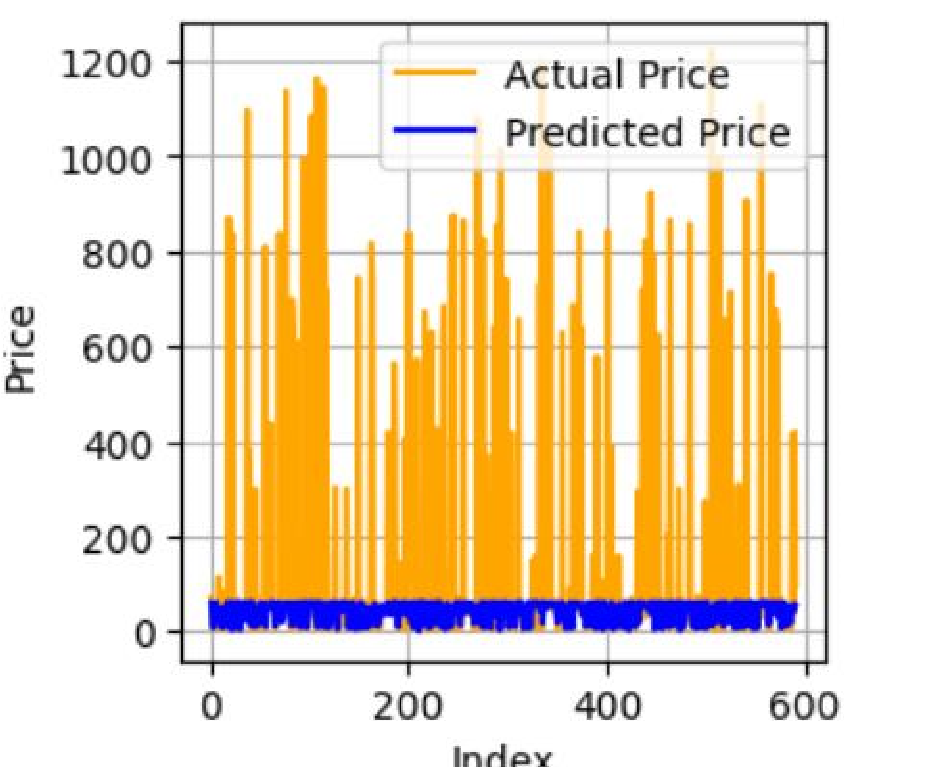
 

Figure 2.Graph: Open price vs low price Figure 5.Graph: Open price vs volume

(LASSO REGRESSION)

(RIDGE REGRESSION)

***METHODOLOGY:***

Enough methods are performed on the data to evaluate the data set and gather knowledge about the data. Let's perform some Machine Learning models and

Experimentation to create a model that helps us to achieve the goal I stated in the problem definition. In this, we talk about the various machine-learning algorithms used for the project. They are Linear regression, Logistic regression, Svm, Knn, Ridge regression, and Lasso regression.

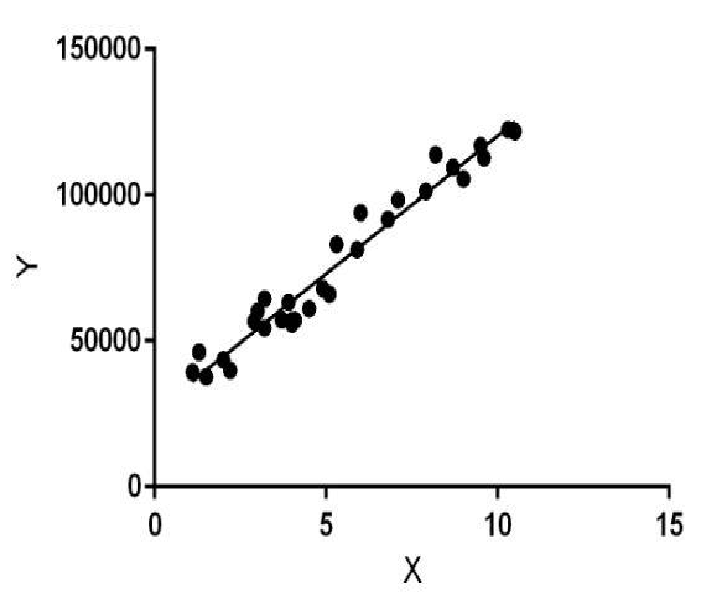
***Linear Regression:***

Linear regression serves as a supervised learning method, mapping out a linear connection between a dependent outcome and one or more independent factors. When only a single predictor is involved, it's termed Univariate Linear

Regression, while the involvement of multiple predictors defines Multivariate Linear Regression. The primary

Figure3.Graph: Open price vs Close price objective is to derive the optimal linear equation to forecast the dependent outcome based on the independent factors,  (SVM) essentially representing their correlation through a straight line. The slope of this line signifies the alteration in the

dependent outcome for a unit alteration in the independent factor. In regression analysis, the dependent outcome (Y) is forecasted from the independent factor(s) (X), employing various regression models or functions to achieve this task..



**Figure 6:** Linear Regression Formula and Examples of Multiple Linear Regression:

yi=β0 β1x1 β2x2 ... βxi ϵ in which for i=n observations: yi = established variable xi = function(impartial) variables β0 = y-intercept (regular time period) βp = slope coefficients for every explanatory variable ϵ=the version’s mistakes term (additionally referred to as

the residuals)

In the real property marketplace, linear regression can be carried out to expect housing expenses.

Independent variables Would possibely encompass Square pictures, wide variety of bedrooms, location, and distinct belongings features, even as the established variable is the housing charge.

In the enterprise global, linear regression may be used to are looking ahead to consumer churn based totally totally on various factors like purchaser tenure, utilization patterns, and customer service interactions. Churn (yes/no) is the dependent variable, and patron associated factors are the independent variables.

Retailers can use Linear Regression to Predict Future sales based on Factors such as Advertising spending, time of year, and past sales data. Sales are the dependent variable, and marketing budgets and time-related factors are independent variables.

***2.* RIDGE REGRESSION:**

Ridge Regression, regularization technique for Linear Regression,adds a regularized term to the cost function to prevent overfitting by constraining the weights. This term, controlled by the parameter 'alpha', utilizes the L2 norm to minimize the weights' magnitude. A higher 'alpha' strengthens regularization, reducing variance in estimates. Scaling inputs with Standard Scaler from sklearn is crucial due to the model's sensitivity to input scaling. Ridge regression enhances linear regression by mitigating overfitting, making it a valuable tool in machine-learning projects.

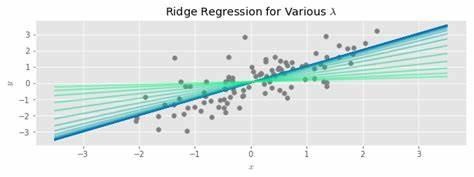


Figure 7: Ridge regression

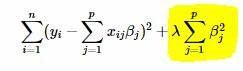


Figure 8:Formula for Ridge regression

EXAMPLE:

•When constructing a portfolio of investments, Ridge Regression can be used to find the optimal weights for different assets to maximize returns while minimizing risk. It helps to handle the issue of multicollinearity among different asset classes and ensures the model is more stable and accurate.

•In marketing, you can use Ridge Regression to

understand the impact of various marketing channels on sales. The goal is to identify which channels have the most influence while accounting for collinearity and preventing the model from becoming too complex.

•Ridge Regression can be applied to predict environmental variables such as pollution levels. Independent variables might include weather data, industrial activities, and traffic density, while the dependent variable is the pollution level. Ridge Regression helps in modeling and accounting for the interdependence of various factors.

***3. LASSO REGRESSION:***

In linear regression, the relation between Input Variables and Target Variable forms a line or hyperplane. Model coefficients are determined by minimizing the sum squared error between predictions and actual values. However, large coefficients in linear regression can lead to instability, especially in datasets with fewer observations than predictors. To address this, penalized linear regression introduces additional costs for large coefficients, known as the L1 penalty. This penalty minimizes coefficients' sizes and enables some to reach zero, effectively removing irrelevant features from the model.

***FORMULA FOR LASSO REGRESSION:***

Residual Sum of Squares λ \* (Sum of the absolute cost of the importance of coefficients) Where, λ denotes the amount of shrinkage. λ = zero implies all features are taken into consideration and it's far equal to the linear regression in which best the residual sum of squares is considered to build a predictive version λ = ∞ implies no function is taken into consideration

The Bias Increases with an Increase in λ

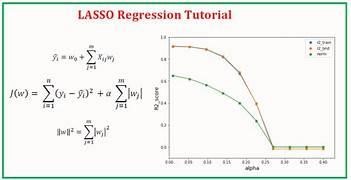


Fig 9: Lasso Regression **Example of Lasso Regression:**

In this section, depicted how ilk application the Lasso Regression algorithm. Firstly, allow presentin' a typical regression dataset. Takin' a housing dataset. A housing dataset be an regular machine learnin' dataset containin' 506 rows of records with 13 numerical input variables and a numerical intention variable. By applyin' a take a look at harness of repeatin' stratified 10-fold pass-validation with 3 repeaters, a naive model should fetch an average absolute mistakes (MAE) . A pinnacle-performin' shape can fetch an MEA in this take a look at . These yield boundaries for the performance expectation on this data put.

***SVM:***

Support Vector Machines (SVMs) is versatile supervised learning algorithm use for classifocation and regreshion tasks, preferring datas with many features or claration separation margins. They aims to finds a hyperplane with the large margin between classees in training data, facilitating accurately classifcation. In SVM, each data point is plotted in an N-dimensional space (were N is the number of features), and an optimial hyperplane is determine to separate classess.While inherent binary, SVMs can handles multiclass problems by creations a binary classifier for each classs in the dataset. This approache enables SVMs to effectively classifying datas acroses multiple classes!!!

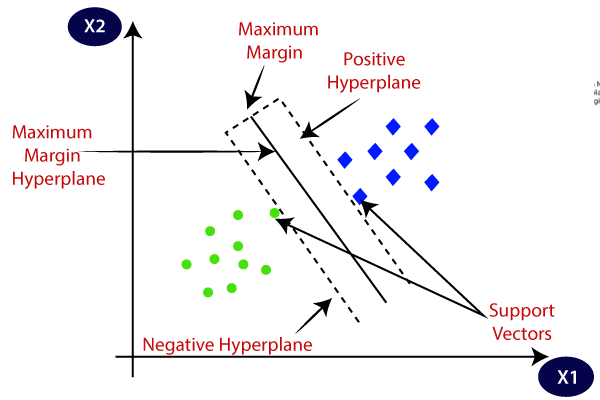


Fig 10. Support Vector Machine

***FORMULA:***

F(x) = signal(β0 β1\*X1 β2\*X2 ... βn\*Xn) f(x) is the choice function that predicts the maginficence label .X1, X2, ..., Xn are the feature values. β0, β1, β2, ..., βn are the coefficients to be found out at some point of the training technique. sign is the sign function, which returns +1 if the expression is positive and -1 if it's negative.

***EXAMPLES:***

Support Vector Regression (SVR) have ability to foretell numeric quantities, such as stock costs, by discover a hyperplane that finest fits the data while decrease deviations. Through SVR, instances consist of prophesying dwelling costs founded on features like square footages and total of sleeping rooms or foreseeing a corporation's earnings founded on previous data and financial signals!!!

***KNN REGRESSION:***

◆K-Narest Nawbor iz a simpl Machin Larning algoritm bassed on Supervised Leaning tecnics!

◆K-Narest Nawbor on the other hand is a simple Machine Learning algorithm that works on a basis of Supervised Learning technics.KNN algorithm store all the data.

◆The function of the KNN algorithm is not critical to application for Regression, but rather it is preferred for a Classification issue.KNN be an no-param fraudulent algorithm, that have not done any thinks on base data.

◆It be likewise known as an idle scholar algorithm coz it do not learn from the teaching group at once rather it saves the data package and at the point of categorization, it executes an action on the data package.

◆The KNN algorithm's training phase simply store the dataset, and when new data arrive, it classifying data into category that is much similar to new data!

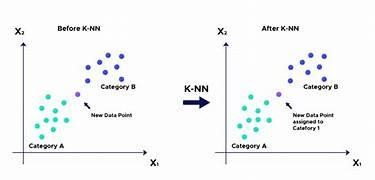


Fig 11. KNN REGRESSION

**Formula**

In the training stage, KNN algorithm just keeps a dataset. Which leads to classify new data or any information that is recently received into a group that resembles the nearest group.KNN be it Supervised Machine Learning Algorithm, can be for Classification and Regression. It is a forecastingoriented approach using majority vote at a distance in the feature space. Formalization of KNN might be described as.**For Classification:**

For a new data point x, KNN predicts class c by selecting the mode among the KNN classes in the training dataset.

**For Regression:**

For a new data point x, KNN predicts the target value by taking the mean (average) of the k-nearest neighbors' target values in the training dataset.

**Examples:**

1. KNN was trained on dataset measuring flowers, if there's a new flower like attributes K nearest neighbors, it predicts class based on common class among neighbors!
2. Housing' estimate price predicament KNN, a' it can estimating’ a house price by takin' the average of k nearby neighboring' houses with alike characteristics, such as square footage 'n bedrooms number.
3. KNN be used detect anomalies in credit card transactions through calculations distances between transactions and flagging those significantly different from their neighbors!

V. CONCLUSION

The utility of varied machinery learnings algorithm has significantly boost the domain of ancient sickness perception. Environment like cardiac problems, renal disorder, breasts cancer, and neurological sickness have tremendously profited from these algorithms. Considering the future, there's a persuasive necessity to establish more complicated machinery-learnings models to further intensify the accuracy and efficiency of sickness prediction. One crucial attribute of this development is the continuous calibration of those models after the primary training phase. Consistent finetuning can possibly result in superior model accomplishment and more credible predictions. This ceaseless adjustment and optimization of algorithms are crucial to guarantee that they stay effective as medical comprehension and data change .

Expanding da scope of datasets are imperative! To prevent overfittedting and enhance the precision of model deployments, it's essential to including a wider representation of demographics factors. Incorposating data from various populations can help ensures that da models are not biased or limited in applicability. In da search for greater model accuracies, it's essentially to utilizing more sophisticated feature selections techniques. By identifiering and inclusion the most relevant features, we can improvement da overall performances of machine learnings models. This not only leads to more preciseness predictions but also simplicities the model by focus on da most informating aspects of data !

The future disease machine prediction be a complex prediction model, regular calibrate model, broader and representative datasets, and implement advanced feature selection method. Essential step to improve accuracy, effectiveness, and applicability of machine learning algorithm in healthcare and diagnose disease.

REFERENCES

1. Luren Dai. The titile is Predict Stock Price of Tesla Based on Machine Learning. In Woodsworth College, University of Toronto.
2. Indronil Bhattacharjee Department of Computer Science and Engineering Khulna University of Engineering & Technology Khulna, Bangladesh.
3. Dr. Chaitanya Kishore Reddy.M1 1Professor & Dean-IT, NRI Institute of Technology.
4. Hao Li\* Department of Mathematical Sciences, University of Southampton, Southampton, the UK
5. Subhash Chand Agrawal Department of Computer Engineering and Application, GLA University
6. Jianyao Li\* Department of Management, Shandong University, Jinan,China.
7. Hendri Mahmud Nawawi Technical Information, Nusa Mandiri University,Jakarta,Indonesia.
8. Michele Coiro School of Professional Studies Centenary University,18.Eastern.Parkway .
9. Amey Bhadamkar. The title is Tesla Inc. Stock Prediction using Sentiment Analysis.
10. Desai Mitesh Madhusudan U.G. Student, Department of Information Technology, B. K. Birla college of Arts, Science and Commerce (Autonomous), Kalyan 421 306 , Maharashtra, India.