Exploring Future Sales Prediction using Classification and Regression

Dr. Chaitanya Kishore Reddy.M[1]
Professor & Dean-IT, NRI Institute of Technology, A.P,
India-521212

G.Sravanth[2]
UG Scholar, Dept. Of IT, NRI Institute of Technology, A.P,
India-521212

P.Mounika[3]
UG Scholar, Dept. Of IT, NRI Institute of Technology, A.P, India-521212

Abstract:- Future forecast The greatest way to accomplish the targeted marketing objectives is through sales analysis. It would be better to advance your career once you have the capacity to make strategic decisions in sales forecasting. When predicting sales numbers, it is important to remember that future product prices will have an impact on the volume of sales in addition to past sales data. Multivariate time series first appear to be the best model for this problem. Since there is only ever one price for a product at any given time in sales history, unlike in real life where history is not always repeating. It makes creating a multivariate time series more challenging. However, the price is more dependent on the expiration date for some seasonal or perishable products. This additional data can aid in the creation of a causal time series model that is more precise. The proposed remedy makes use of a univariate time series model, but includes the product's price as a factor that systematically affects the prediction. Based on previous sales data, the price influence is calculated using data correlation analysis and customizable price ranges to find products with comparable histories. This unique strategy is simple to compute compared to other methods and enables the pricing parameter for simulations and forecasts to be chosen in advance.

Keywords:- *Sales Predicton,Forecasting, Linear Regression,Random Forest,Classification, Regression, Decision Trees, Training set.*

I. INTRODUCTION

Future sales prediction is the practise of estimating a company's short- or long-term sales success in the future utilising historical data and sales records from previous years. One benefit of careful financial preparation is this. A sales forecast is an estimate of future revenue for a given time frame, such as a week, month, quarter, or year.

The forecast is typically based on information like the amount of deals that have already closed in your sales pipeline and the stage of the sales cycle they are in.

It enables a sales force to forecast which deals will complete and how much money will flow into the company as a result with a high degree of accuracy.

If we are trending below target, it offers us time to adjust the sales plan, allowing us to manage our revenue projections for the predicted timeframe.

It gives individual sales reps visibility into their ongoing deals, enabling them to anticipate future roadblocks and make necessary course corrections.

It makes it possible for the entire organisation to plan forward, looking at potential growth prospects in new use cases or even emerging markets.

The act of assessing future sales by foreseeing the volume of goods or services that a single salesperson, a sales team, or an organisation is likely to sell over the course of a specific time period, such as the upcoming week, month, quarter, or year is known as sales prediction, sometimes known as sales forecasting. There are many different approaches to estimating sales, but what's crucial is that they are all supported by data of some form. You may replace gut instincts with evidence-based decision-making by using sales prediction. There are two ways to forecast sales:

- *Since quantitative forecasting makes use of both recent trends and historical company data, it is the most precise.*

- *Qualitative forecasting is individualised and draws on expert opinions, market research, and surveys of future clients.*

In order to analyse previous company data and patterns, such as completed and won deals and win/loss records, sales forecasting software uses mathematical methodologies. This analysis results in an accurate report of predicted sales income. Forecast reports contrast the sales goals with the actual results and the projected results.

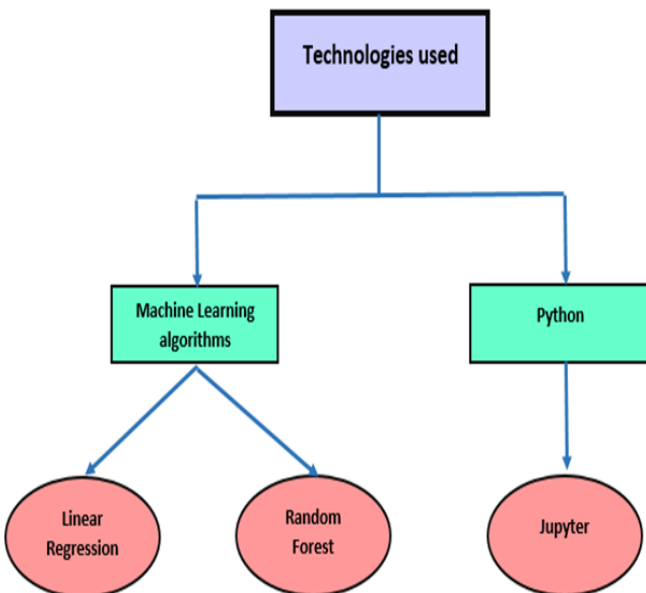Fig 1 Sales Forecasting

II. TECHNOLOGIES USED

Fig 2 Technologies used

➢ *Machine Learning*

Machine learning is a subfield of artificial intelligence (AI) and computational science that primarily concentrates on using data and algorithms to mimic how humans learn, progressively increasing its accuracy. It is one of the crucial components of the expanding discipline of data science.

To produce classifications or predictions or to find significant trends in data mining projects, algorithms are developed and refined using statistical approaches. Ideally, these insights have an impact on important growth metrics since they are used to drive decision-making within services and enterprises.

The market will increasingly want more data scientists as big data continues to expand. These are necessary to assist in determining the most appropriate business inquiries and the data required to respond to them.[Fig:3]
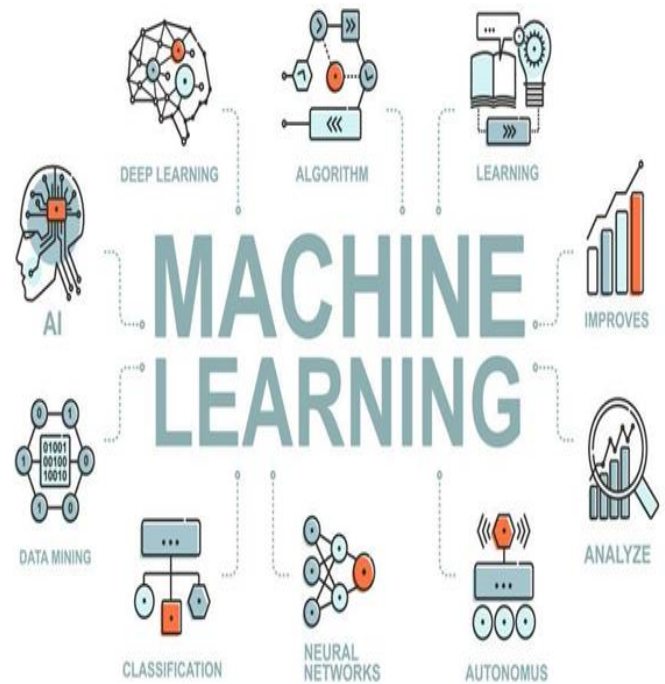

Fig 3 Machine Learning

• *Linear Regression*

A machine learning approach called linear regression is based on supervised learning. The regression task is carried out. Regression uses independent variables to model the desired predicted value.

Additionally, it is employed to determine the connection between factors and predicting.

Regression models vary depending on the amount of independent variables employed, the type of relationship among dependent and independent variables, and other factors.

The dependant variable in a regression goes by many different names. It may also be referred to as a regressand, endogenous variable, criteria variable, or outcome variable.

The independent variables may also be referred to as predictors, exogenous variables, or regressors.[Fig:4]

The task of predicting a dependent variable's value (y) with an inverse relationship involving x (input) and y (output) is carried out using linear regression.

Hence, linear regression is so named. In the diagram above, a person's pay is represented by Y (output) and their job experience by X (input). The line that fits a model the best is the regression line.
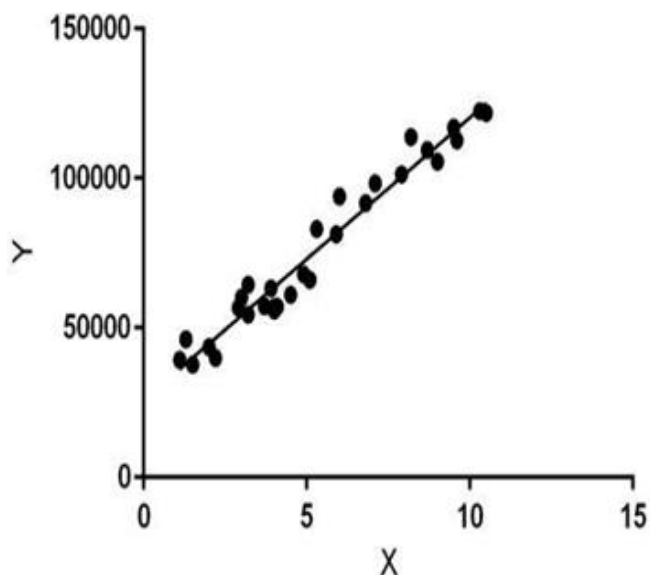
Fig 4 Linear Regression

- *Random Forest*

The grouping technique known as Random Forest uses several decision trees with a method known as Bootstrap as well as Aggregation, often known as bagging, to accomplish both classification and regression analyses.

The primary idea underlying this is to merge many decision trees in selecting the final output instead of depending on individual choices trees.

As the primary learning models, Random Forest uses a variety of decision trees.

For each model, sample datasets are created by randomly selecting rows and features from the dataset.

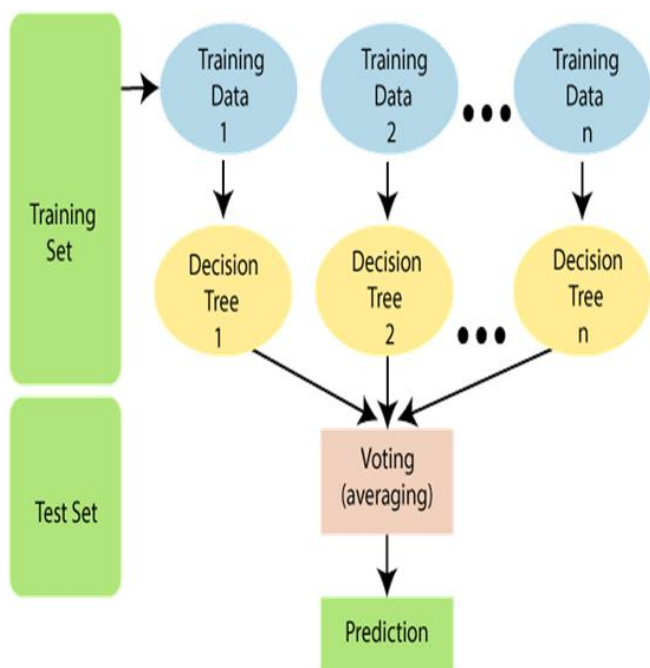The Bootstrap component is this.[Fig:5]

Fig 5 Random Forest

➢ *Python*

You may create and share documents with live code, equations, visuals, and narrative text using the free and open-source Jupyter Notebook web tool. Included in this are data transformation, classification, statistical modelling, data visualisation, machine learning, and data cleaning. Python is one of the nearly 40 programming languages supported by Jupyter. The Jupyter Notebook software itself can be installed using Python.

Using Anaconda, install Jupyter: Use the Anaconda Distribution to install Python and Jupyter, which also contains additional frequently used tools for scientific data and computational science. The most recent version of Python 3 for Anaconda is available here for download. Install Anaconda as soon as it has been downloaded.[Fig:6]

Using PIP, install Jupyter Notebook by running python3 -m python install --update pip. install Jupyter using python3 -m pip
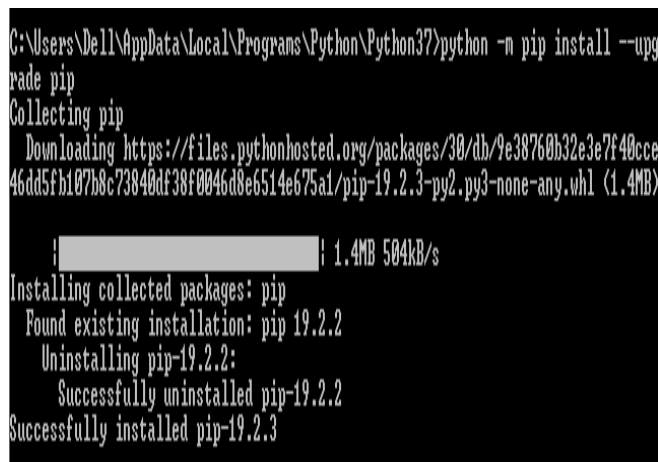
Jupyter Notebook is the command used to launch it.

Fig 6 Python

III. SOFTWARE REQUIREMENTS SPECIFICATION

SRS is a comprehensive description of the system's projected performance. At the conclusion of the requirements engineering phase, it is often approved[Fig:7].It outlines how software systems would interact with each other, internal components, hardware, other programmes for communication, and human interactions with them in a variety of situations that are similar to real-world ones.[Fig:7]
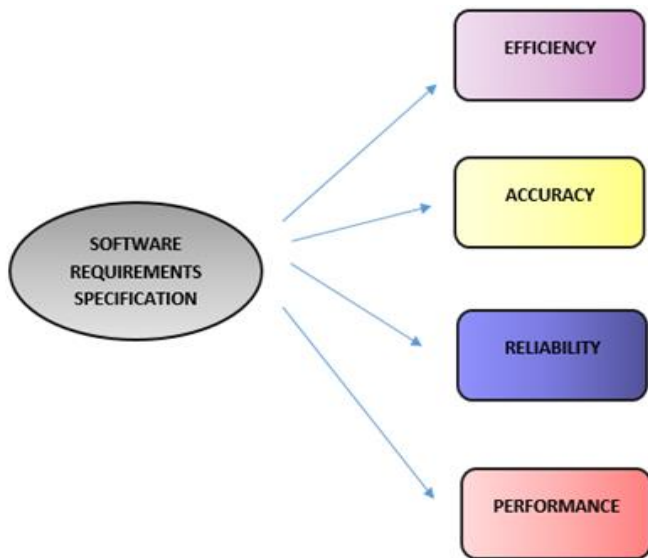
Fig 7 SRS

IV. EXISTING SYSTEM

To put it simply, "future sales prediction" is a method used to categorise or forecast future sales for a retail outlet or other business. This method's major objective is to precisely predict future sales in order to prevent product loss. Through the use of algorithms for machine learning and models, future sales can be predicted.

Sales forecasting is done utilising information about previous performance. Management of the sales pipeline and forecasting go hand in hand. Sales representatives and managers can give information for a trustworthy sales forecast by getting a precise picture of success, dedication and velocity for each contract. a projection of sales provides a precise forecast of future revenue performance by utilising a multitude of data sources.

Typically, historical performance information is used to construct sales projections. A detailed report that forecasts the amount of sales that an agent, team, or organisation will make each week, month, quarter, or year is known as a sales forecast.

Historical prediction: This method for predicting makes use of current or appropriate data.

Forecasting of opportunity stages stage forecasting: Your sales funnel is related to this forecasting technique.

➢ *Disadvantages of Existed System*

• *Uncertainty*
Nothing can be guaranteed to occur in the future. Global pandemics, economic shifts, and competitive behaviour are a few examples of variables that are simply impossible to forecast. Your forecasts may out to be meaningless due to outside influences.

• *Inaccuracy*
The majority of conventional forecasts assign a weight to each step of a negotiation, and these weights get better as the deal approaches a win or loss. They do not, however, take into account the length of time the arrangement has been in the works, the level of interaction between the representatives and the stakeholders, the recentness of the participation, if the close date remains the same or changed, and so on. As a result, because it does not take into account these decisive aspects, this kind of forecast is not accurate.

• *Lack of Sales History*
Newer companies or startup companies may find it difficult to estimate sales because sales forecasting models frequently use prior data to project future sales. Some methods need at least two years' worth of historical data to produce a reliable forecast.

V. PROPOSED SYSTEM

To enhance the accuracy of predictions, we suggest using both a random forest approach and a linear regression technique.

The relationship among a scalar answer and one or a few explanatory factors can be modelled **linearly** using the linear regression method. By using the input variable as a basis, this method forecasts the output variable. It's a dependent variable that you're after. Independent variables are those that you use to forecast the value of another variable.

When solving classification and regression issues, the random forest algorithm, a sort of supervised learning algorithm, employs ensemble techniques. During training, the algorithm builds a large number of decision trees and outputs the mean as well as the mode of each tree's forecast.

➢ *Advantages of Proposed System*

• *Advantages of Linear Regression*

✓ *Simple to Implement*
It's fairly simple to implement linear regression and get good results. In addition, compared to other complicated techniques, these models may be trained quickly and efficiently even on platforms with limited computational capability.Comparing linear regression with alternative machine learning techniques, linear prediction has a significantly reduced temporal complexity.The linear regression's mathematical formulae are also rather simple to comprehend and interpret. Given this, the linear regression method is a pretty simple algorithm to use.

✓ *Results on datasets with linear separation*
In order to determine the kind of a connection between variables, linear regression frequently used to arrange linearly separable datasets almost flawlessly.

- *Advantages of Random Forest*

✓ In general, random forests, or neural networks, calculate the relevance of variables. Additionally, they provide a better way of handling missing data. The variable that is present in a given node the most often is used to fill in any missing values.

✓ Out of all the categorization techniques that are currently available, random forests offer the highest accuracy.

✓ It can also manage enormous amounts of data with thousands of different variables. When a class in the data is more rare than the other classes, it can effectively balance the data sets.

✓ Additionally, this strategy manages variables the quickest, making it appropriate for challenging tasks.

VI. SYSTEM ARCHITECTURE

Fig 8 System Architecture

➢ *Gathering Data*

In order to answer particular inquiries, investigate hypotheses, and assess results, gathering data is the act of obtaining and analyzing information on relevant variables in a systematic manner.[Fig:8]

➢ *Analysis of Data*

Data is used in predictive analysis to foretell future patterns or events. Additionally, it uses previous data to project prospective outcomes that motivate us to make strategic decisions. [Fig:8]

➢ *Choosing Best Sales Forecasting Model*

The top sales forecasting model techniques, such as "Regression" and "Classification," are picked in this module. [Fig:8]

➢ *Sales Forecasting*

Sales forecasting is the procedure of assessing upcoming sales by estimating the volume of goods or services that one salesperson, a team of salespeople, or an organisation is expected to sell over the course of a specific time period, such as a week, every month, period, or annual.[Fig:8]

➢ *Sales Forecasting Results Evaluation*

When using forecasting techniques, it is important to estimate the circumstances. To test methods, evaluation procedures are required.[Fig:8]

VII. FUTURE SCOPE

A forecast of sales to come within a specific time frame, typically for one month or a year, is known as a sales prediction. The projection is based on information like the amount of commitments presently in the channel of sales and where in the sales process each of them are. Forecasting sales offers several wonderful advantages. The ability to effectively allocate resources for future growth and manage product loss is made possible by a company's ability to predict sales accurately. Sales projections empower executives to adjust their course early and help set criteria for future direction. Sales managers can stop merchandise from being destroyed by using sales prediction. Quotes for sales and revenue expectations can be aligned, allowing revenue leaders to maximise their chances of success.

VIII. CONCLUSION

Sales forecasting is an essential step in the strategy planning process because it enables businesses to foresee future performance. It gives them the ability to foresee not only fresh opportunities but also potentially detrimental tendencies. A mission statement is essential for any organisation since it defines its goals and serves as a guide for decisions. Businesses may effectively assess their costs and revenues with the use of sales forecasting, which then allows them to project their performance over the short- and long-term.

REFERENCES

[1]. https://www.geeksforgeeks.org/mllinearregression/#:~:text=Linear%20Regression%20is%20a%20machine,relationship%20between%20variables%20and%20forecasting.

[2]. https://spotio.com/blog/best-sales-forecasting-software/

[3]. https://www.tandfonline.com/doi/full/10.1080/21642583.2014.999389?scroll=top&needAccess=tr

[4]. https://corporatefinanceinstitute.com/resources/data-science/random-forest/

[5]. https://iq.opengenus.org/advantages-and-disadvantages-of-linear-regression/

BIOGRAPHIES

Dr. Chaitanya Kishore Reddy. M is currently working as a Professor and Dean in the Department of Information technology at NRI Institute Of technology, Pothavarappadu, Agiripalli, Krishna(dist), India. He received Ph.D. in Computer Science and Engineering and M. Tech in Computer Science and Engineering at Jawaharlal Nehru Technological university, Kakinada. He has Published 40 research papers in various National and International Journals and International Conferences. He is a member in ISTE, CSI, and IAENG. His research areas are Mobile Ad-hoc Networks, IoT, and Cloud Computing.

G.Sravanthi is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. She has done a project on future sales prediction. She has finished an internship at Blackbucks and she has done 2 NPTEL certificates.

P.Mounika is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. She has done a project on future sales prediction. She has finished an internship at Blackbucks and she has done 2 NPTEL certificates.