



# Sales Estimation of A Company Through ML Model

K. Sathyam <sup>1</sup>, M. Abdul Subhani<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

<sup>2</sup>Post Graduate, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

## Article Info

### Article History

Received : 25 March 2024

Published : 07 April 2024

### Publication Issue :

March-April-2024

Volume 7, Issue 2

Page Number : 318-325

## ABSTRACT

E-commerce platforms are developing more quickly as a result of the expanding impact of the Internet on people's lives. Both the number of users and the revenue generated by these platforms are trending upward. Strong government policy support in recent years has also created a favourable atmosphere for the growth of the e-commerce sector. The e-commerce sector has played a more significant part in the growth of the national economy as a result of this year's epidemic. In these situations, e-commerce platforms and businesses are growing in quantity and competitiveness. A platform needs to be able to better match user needs and perform well in all areas of coordination in order to keep its competitive advantage. Let's consider a problem statement, an e-commerce company wants to increase its sells to a certain amount. Now the challenge is to find the amount of investment on advertisement that will result the gain in sells. This project involves building the sales estimator for a company which utilizes the value of the money invested to predict the sales get by advertisements.

The Sales Estimation Project is an innovative application developed using Python programming language aimed at assisting businesses in accurately forecasting their sales figures. This abstract provides an overview of the project's objectives, functionalities, and significance.

The main objective of the Sales Estimation Project's goal is to leverage data analysis additionally machine learning techniques to predict based on past sales, projected sales data, industry patterns, and more pertinent elements. By harnessing the power of Python libraries like Scikit-learn, NumPy, and Pandas, the project offers advanced analytical capabilities for generating accurate sales forecasts.

Key functionalities of the Sales Estimation Project include feature engineering, model training, evaluation, and data preprocessing. The project

supports various machine learning algorithms

In conclusion, the Sales Estimation Project serves as a useful instrument for companies seeking to enhance their sales forecasting accuracy additionally make data-driven decisions. By harnessing the capabilities of Python programming language and machine learning algorithms, the project empowers businesses to optimize their operations and achieve better sales outcomes.

Keywords : Advertising, Sales, Python, Machine Learning, Marketing, Customer Segmentation, Predictive Analytics, Campaign Optimization, Revenue Generation, Data Analysis.

---

## I. INTRODUCTION

Big data-driven machine learning models have been incorporated into marketing in recent years with the goal of turning consumer data into actionable insights and improving forecasting accuracy for strategic decision-making. While a lot of research has been done on demand forecasting, less has been done to examine how marketing tactics like advertising and other promotional activities impact demand. Consequently, a precise demand-predicting Big data-driven machine learning models have been incorporated into marketing in recent years with the goal of turning consumer data into actionable insights and improving forecasting accuracy for strategic decision-making. While a lot of research has been done on demand forecasting, less has been done to examine how marketing tactics like advertising and other promotional activities impact demand. Consequently, a precise demand- predicting/sales estimation model has the potential to significantly advance company sustainability both theoretically and practically. This project's goal is to assess machine learning techniques that can accurately predict demand based on advertising costs. The study's prediction mechanism, which is based on a machine learning technique, has the potential to significantly advance

both academic research and commercial sustainability. This project's goal is to assess machine learning techniques that can accurately predict demand based on advertising costs. The study focuses on a machine learning-based prediction method. REGRESSION to address demand projections derived from advertising costs. Thus, a manufacturer's real market dataset made up of spending on advertising on Tv, social media, Newspaper, The accuracy of demand forecasting using selected machine learning techniques for sales and demand was analyzed and contrasted. Consequently, Multiple Linear Regression has been used here for delivering remarkably precise sales estimation projection findings based on advertising expenditures. In today's dynamic business landscape, accurate sales estimation has a crucial part in the accomplishment of and sustainability of any organization. Understanding and predicting sales figures not only facilitate informed decision-making but also enable businesses to allocate resources effectively, strategize marketing efforts, manage inventory, and plan for future growth.

The Sales Estimation Project aims to harness the power of data analytics and statistical modelling to develop robust algorithms that forecast sales trends with precision. By leveraging past sales information, industry trends, demographic information,

Considering additional pertinent elements, this project endeavours to provide actionable insights to businesses across various industries.

Through meticulous analysis and implementation of advanced predictive analytics techniques, the project seeks to address key challenges faced by businesses in estimating sales figures accurately. These challenges may include seasonality, fluctuating consumer behaviour, competitive dynamics, economic shifts, and unforeseen external factors.

Furthermore, the Sales Estimation Project is designed to be adaptable and scalable, catering to the unique needs and complexities of different businesses. Whether it's a small start-up looking to establish market presence or a multinational corporation seeking to optimize its global sales strategy, the project aims to deliver tailored solutions that drive profitability and expansion that is sustainable.

Ultimately, by utilizing data-driven insights to their full potential, the Sales Estimation Project endeavours to empower businesses to make informed decisions, mitigate risks, capitalize on opportunities, and thrive in today's fiercely competitive where Companies can have a virtual idea to spend on what advertisement media so that they can utilise and get maximum benefits. And with the help of this project we are bringing the use of technology in the field of Sales Estimation where Companies can get a clear idea where to Invest how much of Investment to get maximum profits.

## II. LITERATURE REVIEW

### **Examining the literature on Company sales estimation:**

Examining the literature on company sales estimation using Python and machine learning (ML) involves delving into various resources that discuss

methodologies, techniques, and tools for predicting and estimating sales figures. Explore different methods for sales forecasting, such as time series analysis, regression models, ensemble methods, and deep learning algorithms. Examine using Python libraries for data preprocessing such as Pandas and Scikit-Learn, feature engineering, model training, and evaluation. Review literature on gathering sales data from multiple sources (e.g., CSV files, ERP databases, market data), cleaning data, handling missing values, and preparing datasets for ML models. Learn about techniques for selecting relevant features, creating new features, and transforming data to improve the performance of sales estimation models.

### **An Overview of Machine Learning:**

A subfield of computer science and artificial intelligence called machine learning (ML) is concerned with using data and algorithms to help AI mimic human learning processes and progressively become more accurate. What is machine learning? Artificial intelligence, roughly defined as a machine's capacity to mimic intelligent human behaviour, includes machine learning as a subfield. Complex tasks are carried out by artificial intelligence systems in a manner akin to human problem-solving. Machine learning aims to teach machines to perform tasks more efficiently without the need for explicit programming. There are a few measures that must be taken to accomplish this goal. Data must first be gathered and prepared. Next, an algorithm, or training model, must be chosen.

### **Talks about feature selection technics and how well they work to Company Sales Estimation**

Feature selection techniques play an important part in the success of models for machine learning for company sales estimation. These techniques help

identify the most pertinent and impactful characteristics of the dataset, leading to improved model precision, reduced overfitting, and better generalization. Here are some commonly used feature selection techniques and their effectiveness in company sales estimation. Univariate feature selection works well when there are clear relationships between individual features and the target variable (sales). It helps in identifying key factors that directly influence sales performance. Feature importance from tree-based models is valuable for identifying influential features, especially when dealing with non-linear relationships and interactions. It helps prioritize features with the most predictive power for sales estimation. Forward and backward selection methods can be effective for smaller feature sets, allowing the model to consider different combinations of features and select the best subset for sales estimation. Analyzing pairwise correlations between features and the target variable can help identify features with strong correlations and eliminate redundant or highly correlated features.

### III METHODOLOGY

#### Approach

The methodology approach for company sales estimation using Python and machine learning involves several key steps and techniques. Clearly define the problem of company sales estimation, including the target variable (e.g., sales revenue), data sources. Gather past sales information, customer information, market patterns and more pertinent data sources.

Perform data cleaning, handle missing values, remove duplicates, and ensure data quality and consistency.

#### Implementation

Gather historical sales data, including features such as date, product/category, price, quantity sold, customer demographics, promotions, etc.

Preprocess the data using Pandas to handle missing values, remove duplicates, encode categorical variables, and perform feature engineering if needed (e.g., create new features like total sales, average sales per customer, etc.). Choose a suitable machine learning model from Scikit-Learn for sales estimation, like gradient boosting, decision trees, random forests, and linear regression. Utilizing the `train_test_split` tool from Scikit-Learn, split the pre-processed data into training and testing sets. Train the selected model on the training data using Scikit-Learn's fit function.

Optionally, To determine the ideal model parameters, do hyperparameter tuning using methods such as `GridSearchCV` or `RandomizedSearchCV`. Use Scikit-Learn's metrics module to calculate these metrics. Load the trained machine learning model and necessary data into the Streamlit app.

Design the user interface using Streamlit's interactive widgets (e.g., sliders, dropdowns, checkboxes) to allow users to input data or select options.

Write the logic to process user inputs, pass them through the model for prediction, and display the estimated sales results. Once you're satisfied with the Streamlit app's functionality and performance, you can deploy it to a web server or cloud platform for public access.

#### Characteristics

Characteristics of company sales estimation using Python and machine learning (ML) include several key aspects that define the approach and techniques used for predicting sales figures accurately. Utilizes

Pandas for data manipulation, cleaning, and preprocessing activities such as scaling numerical characteristics, encoding categorical variables, and handling missing values and removing outliers. Prepares the dataset by organizing it into features (independent variables) and the target variable (sales), ensuring data quality and consistency. Implements feature selection techniques from Scikit-Learn (e.g., SelectKBest, Recursive Feature Elimination) to identify relevant features that contribute most to sales estimation.

Performs feature engineering to create new features, derive meaningful insights, and capture important relationships between variables. Selects appropriate ML algorithms from Scikit-Learn's library, depending on the features of the problem and the data structure, such as support vector machines (SVMs), gradient boosting, decision trees, random forests, and linear regression.

#### **Data pre-processing**

Preparing the data is an essential step in company sales estimation using Python with machine learning (ML) libraries like Scikit-Learn and Pandas. It involves cleaning, transforming, and properly structuring the raw data in order to prepare it for analysis and model training.

Load the sales data into a Pandas DataFrame from a variety of sources, including CSV files, Excel sheets, databases, or APIs. Examine the dataset for any missing values (NaN or null values) and choose the best course of action for handling them, such as adding a specified value (such as the mean, median, or mode) or deleting rows or columns. Clean the data by removing duplicate records, irrelevant columns, or anomalies that could compromise the precision of sales estimation models. Develop fresh functionalities or transform existing features to capture valuable information and improve model

performance. Calculating total sales, average sales per customer, or sales growth rate. To train the machine learning model on training data and assess its performance on untested data, divide the preprocessed data into sets for testing and training.

#### **An Explanation of Machine learning Classifiers**

In the context of company sales estimation, machine learning classifiers play a crucial role in predicting and forecasting sales figures based on historical data and relevant features.

The first step is to gather historical sales data, including features such as time stamps, product categories, customer demographics, pricing information, promotional activities, and any other relevant data points.

This data is typically stored in a structured format such as a CSV file or a database, and it needs to be preprocessed and cleaned before training machine learning models. The process of feature engineering is turning unstructured data into useful features that machine learning algorithms may utilize to provide precise predictions.

For sales estimation, feature engineering may include creating new features such as total sales per customer, average sales per product category, seasonal trends, customer segmentation based on buying behavior, etc. Selecting an effective machine learning classifier for sales estimation is the next step after preprocessing the data and feature engineering. The type of data, the intricacy of the features-sales correlations, and the model's intended interpretability all play a role in the classifier selection.

## IV EXPERIMENTAL SETUP

### Pandas:

A potent data manipulation tool, Pandas and analysis library in Python that provides data structures like DataFrame and Series, along with functions for data cleaning, preprocessing, and transformation.

In the context of machine learning classifiers, Pandas is often used for loading and preprocessing datasets, addressing missing data, coding variables that are categorical, and creating feature-engineered datasets for model training.

### Streamlit:

Streamlit is a popular library for building interactive web applications and dashboards with Python. It allows developers to create user-friendly interfaces for machine learning models, data visualizations, and data-driven applications.

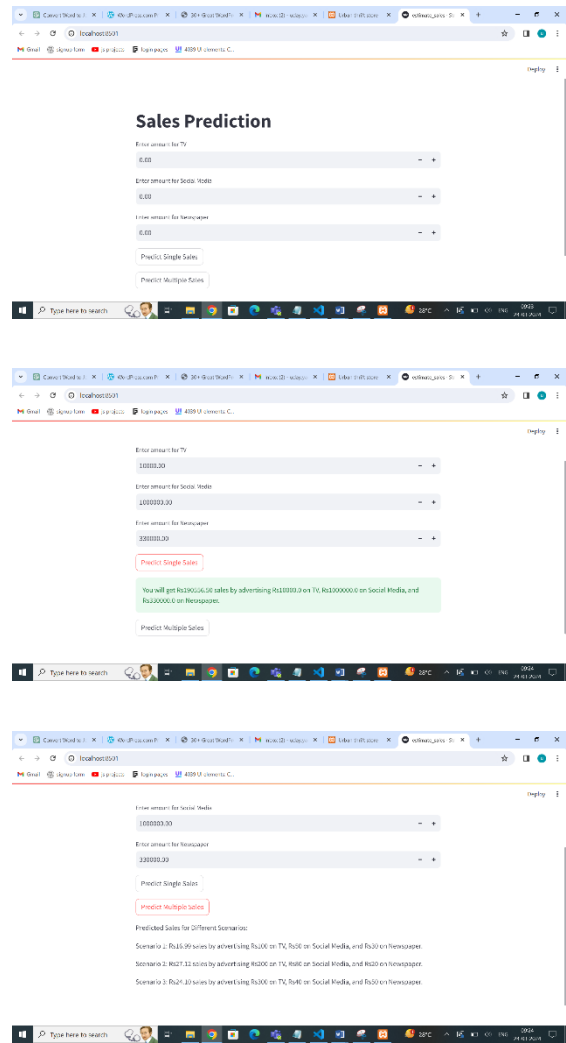
Regarding machine learning classifiers, Streamlit can be used to create web apps that accept user input (e.g., feature values for prediction), display model predictions, and visualize model performance metrics.

### Scikit-Learn (sklearn):

Scikit-Learn is a Python machine learning toolkit that offers several different algorithms for dimensionality reduction, clustering, regression, classification, and other uses. In the context of machine learning classifiers, Scikit-Learn offers implementations of popular classification algorithms such as k-nearest neighbors (KNN), logistic regression, decision trees, random forests, support vector machines (SVMs), and more.

Scikit-Learn is a flexible package for creating and assessing machine learning classifiers since it offers tools for feature selection, hyperparameter tuning, model evaluation, and pipeline development.

## V ANALYSIS



These Pictures Shows Single and Multiple Sales Predictions.

## VI DISCUSSIONS

### Interpretation of Results:

Interpreting the results of a company sales estimation analysis involves understanding the outcomes of the machine learning model and deriving actionable insights from them. Plot the predicted sales values generated by your ML model against the actual sales data on a time series plot or scatter plot to visually inspect how well the model

captures the sales trends and patterns. Translate the model's results into actionable business insights and recommendations for advertising strategies. Identify the key factors driving advertising spend and suggest strategies for optimizing advertising allocation, targeting specific audience segments, adjusting campaign budgets, or exploring new advertising channels. Prepare a comprehensive report summarizing the results of the advertising spend estimation analysis, including model performance metrics, visualizations, key findings, insights, and actionable recommendations.

### **Company Sales Estimation implications**

The implications of company sales estimation using machine learning (ML) are significant and can have a profound influence on different business areas operations. Massive amounts of historical sales data can be analyzed by ML algorithms, which can then be used to spot intricate patterns and trends that conventional statistical techniques would miss. This leads to more accurate and precise sales estimations, reducing forecasting errors and enhancing decision-making. Accurate sales estimations enable companies to forecast future sales trends more effectively. This helps in planning inventory levels, production schedules, resource allocation, and budgeting for marketing and sales initiatives, leading to improved operational efficiency and cost savings.

### **Benefits of company sales estimation**

Company sales estimation offers numerous benefits that can positively impact business operations and decision-making. Company sales estimation helps in predicting future sales trends with greater accuracy, allowing businesses to plan inventory levels, production schedules, and resource allocation more effectively. By accurately estimating sales,

companies can allocate resources such as marketing budgets, staffing, and production capacity in a more efficient and cost-effective manner, avoiding overstocking or underutilization of resources. Accurate sales estimation leads to better inventory management by ensuring optimal stock levels, reducing stockouts, minimizing excess inventory costs, and improving overall supply chain efficiency.

Sales estimation provides insights into customer preferences, buying behavior, and market demand, enabling companies to tailor their marketing strategies, campaigns, and product offerings to target the right audience effectively.

### **Drawbacks of company sales estimation**

While company sales estimation using machine learning and data-driven techniques offers numerous benefits, it's essential to acknowledge potential drawbacks and limitations associated with this approach. Sales estimation heavily relies on historical sales data. Poor data quality, incomplete data sets, or data inaccuracies can lead to biased predictions and unreliable forecasts. The training data may be overfitted by machine learning models, capturing noise or unimportant patterns that do not transfer well to new data. This can result in overly optimistic sales estimates that do not reflect actual market conditions. Complex machine learning models may be challenging to interpret and explain, making it difficult for stakeholders and decision-makers to trust the results and understand the underlying factors driving sales predictions. Sales estimation models may have limited scope and may not account for outside variables such as economic fluctuations, alterations in the behavior of consumers, competitor actions, regulatory changes, or unforeseen events (e.g., pandemics, natural disasters) that can impact sales.

## VII CONCLUSION

In conclusion, the Sales Estimation Project has significantly improved our ability to forecast sales trends accurately and make wise business choices. By leveraging sophisticated analytics techniques and insights derived from data, we have achieved greater precision in predicting sales figures, identifying market trends, and optimizing resource allocation.

Through thorough data validation processes, we have ensured the integrity and reliability of our sales data, providing a solid foundation for our estimation models. Additionally, the user-friendly interface and intuitive visualization tools have enhanced our ability to interpret and communicate insights effectively, facilitating collaboration and decision-making across departments.

Moving forward, continual refinement and enhancement of the Sales Estimation Project will be essential to staying competitive and adapting to evolving market dynamics.

## II. REFERENCES

- [1]. Armstrong, J. S. (2001). Principles of forecasting: A handbook for researchers and practitioners. Springer Science & Business Media.
- [2]. Chatfield, C., & Yar, M. (2019). Introduction to time series analysis and forecasting. John Wiley & Sons. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice. OTexts.
- [3]. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting methods and applications. John Wiley & Sons.

- [4]. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.