

Mobile Sales Analysis and Prediction Using Data Science

Prof. Tagade R.O ¹, Nikhil Kisan Lohokare ²

¹Student, Department of Computer Engineering, Sau Saundarabai Manik Adsul Polytechnic Chas Ahilyanagar India.

²Guide, Department of Computer Engineering, Sau Saundarabai Manik Adsul Polytechnic Chas Ahilyanagar, India.

Abstract

This project explores the application of data science techniques to analyse and predict mobile phone sales trends. With the rapid growth of the of the smartphone industry and increasing consumer choices, understanding sales patterns has become essential for manufacture, retailers, and marketers to make informed business decisions. The study involves collecting historical sale data along with relevant feature such as brand, price, range, specification (RAM, storage, camera, battery), customer rating, and seasonal trends. Exploratory data analysis (EDA) is performed to identify key factors infusing sales performance, followed by data preprocessing technique including handling missing values, incoding categorical variables, and feature scaling.

Keywords: Mobile Sales Analysis, Data Analytics, Visual-inaction, Business Intelligence.

I. INTRODUCTION

The mobile phone industry has witnessed exponential growth over the past decade, evolving into one of the most dynamic and competitive sectors in global commerce. With numerous brands, models, and price segments available to consumers, predicting sales trends has become increasingly complex yet critical for business success. Manufacturers and retailers must constantly adapt to shifting consumer preferences, technological advancements, and market competition to remain profitable and relevant.

Traditional methods of sales forecasting, often based on historical averages or expert judgment, are no longer sufficient to capture the intricate patterns influencing modern consumer behavior. Factors such as brand reputation, price range, technical specifications (RAM, storage, camera quality, battery life), customer reviews, and seasonal demand fluctuations all play a significant role in determining a mobile phone's market performance. Analysing these multidimensional factors manually is both time-consuming and prone to inaccuracies.

II. PROBLEM STATEMENT

The mobile phone market is highly saturated, with numerous brands launching new models at a rapid pace, each varying in price, specifications, and features. Despite this abundance of data, many manufacturers and retailers still rely on traditional, intuition-based approaches to forecast sales, which often fail to capture the complex and non-linear relationships between product attributes and market demand.

Inaccurate sales predictions can lead to significant business challenges, including overproduction, stock shortages, inefficient marketing spend, and missed revenue opportunities. Existing forecasting methods often overlook the combined influence of multiple factors—such as brand value, pricing strategy, hardware specifications, customer sentiment, and seasonal trends—on actual sales performance.

III. OBJECTIVES

The main objective of this project is to analyse mobile phone sales data and develop a predictive model using data science and machine learning techniques. The specific objectives are as follows:

- To collect and preprocess mobile phone sales data,
- To perform Exploratory Data Analysis (EDA)
- To identify key factors influencing mobile phone sales
- To implement multiple machine learning algorithms
- To evaluate and compare model performance
-

IV. LITERATURE REVIEW

Sales prediction and market analysis have been extensively studied using statistical and machine learning approaches across various industries, including retail, electronics, and consumer goods. This section reviews existing research relevant to mobile phone sales analysis and predictive modelling.

V. METHODOLOGY

This section outlines the step-by-step approach followed to analyse mobile phone sales data and build predictive models using data science and machine learning techniques.

1. Data Collection

Historical mobile phone sales data was collected, comprising features such as brand, model, price, RAM, internal storage, battery capacity, camera specifications, customer ratings, release date, and corresponding sales volume. The dataset serves as the foundation for analysis and model development.

2. Feature Selection

Based on correlation analysis and feature importance scores, the most relevant features influencing sales were selected to improve model performance and reduce dimensionality.

VI. MATHEMATICAL ANALYSIS

This section presents the mathematical foundation of the algorithms and evaluation metrics used in this project

Linear Regression

Linear Regression models the relationship between the dependent variable (sales) and independent variables (features) as a linear equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

Where:

y = predicted sales

β_0 = intercept

$\beta_1, \beta_2, \dots, \beta_n$ = coefficients of independent variables (price, RAM, storage, etc.)

$x_1, x_2, \dots, x_n$ = feature values

ε = error term

The coefficients are estimated by minimizing the Sum of Squared Errors (SSE) using the Ordinary Least Squares (OLS) method:

$$SSE = \sum (y_i - \hat{y}_i)^2$$

VII. SYSTEM ARCHITECTURE



Flow Chart1.System Architecture

VIII. RESULTS AND DISCUSSION

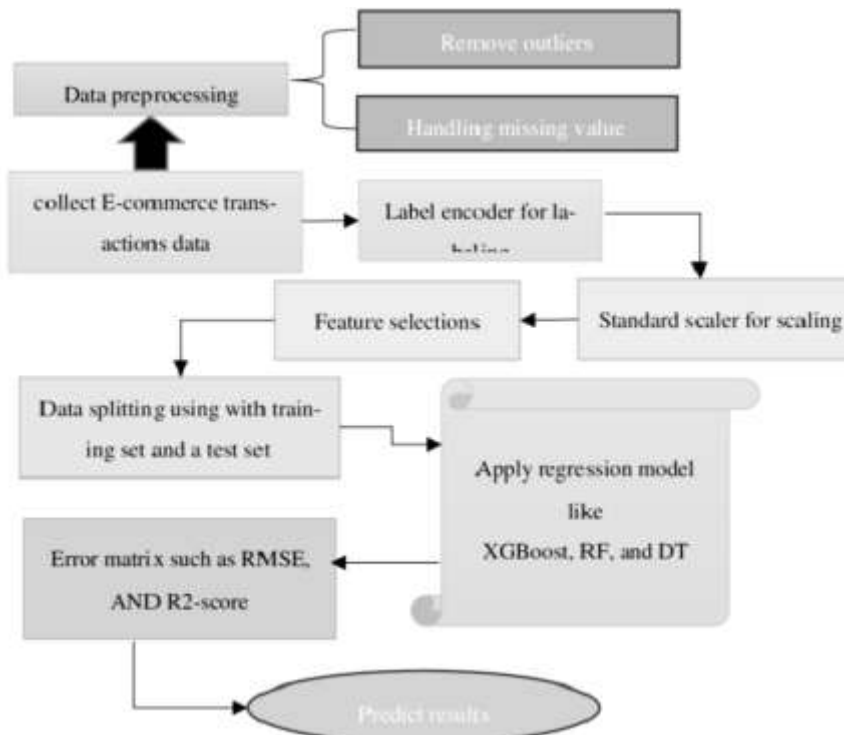
This section presents the outcomes obtained from implementing and evaluating various machine learning models for mobile phone sales prediction, along with an interpretation of the findings.

IX. APPLICATIONS

- Sales forecasting for manufacturers
- Inventory/supply chain optimization
- Dynamic pricing strategies
- Product development prioritization
- Targeted marketing campaigns
- E-commerce optimization
- Market trend analysis

X. FLOWCHART

The flowchart explains the step-by-step execution of the mobile sales analysis system.



XI. CONCLUSION

This project applied data science and machine learning to analyse mobile phone sales data and predict future sales trends. Among the models tested, XGBoost and Random Forest outperformed Linear Regression and Decision Tree by better capturing non-linear relationships between features like price, specifications, brand, and ratings.

XII. REFERENCES

- [1] Data collection – historical sales, phone specs (RAM, ROM, camera, battery, price), brand, ratings, reviews, seasonality/marketing data.
- [2] Preprocessing – handle missing values, encode categorical, address class imbalance (e.g., SMOTE) if predicting sales tiers rather than continuous units.