

Customer profiling, segmentation, and sales prediction using machine learning

Department of Computer Science & Engineering

B. Mahitha(HU21CSEN0600088) Y. Jaswanth(HU21CSEN0600026)
R. Sahasra(HU21CSEN0600184) B. Rohan(HU21CSEN0600181)

Guide Name:
Dr. T. Sasi Vardhan
Assistant Professor
Dept of AI and DS



INTRODUCTION

Objective: The goal of this project is to leverage machine learning techniques to enhance our understanding of customers, segment the market effectively, and accurately predict future sales outcomes.

Customer Profiling: Customer profiling involves creating detailed representations of ideal customers by analyzing various data types, including demographic, behavioral, and psychographic data.

Customer segmentation: Customer segmentation is the process of dividing a broad consumer or business market into sub-groups of consumers based on shared characteristics.

Sales prediction: The main goal of sales prediction is to forecast future sales by analyzing historical sales data, market trends, and customer behavior.

- 1.To develop an efficient system for customer profiling and segmentation through clustering algorithms such as K-Means and DBSCAN.
- 2.To conduct sentiment analysis of customer reviews with VADER to analyze customer satisfaction and preferences.
- 3.To forecast future sales patterns using regression models (Linear Regression, Decision Tree Regression, Random Forest Regression) and time-series forecasting methodologies (ARIMA, LSTM).
- 4.To combine customer segmentation and sentiment analysis with sales forecasting models for better forecasting and targeted marketing campaigns.

PROBLEM STATEMENT

Inefficient Customer Segmentation Traditional segmentation methods such as demographic segmentation, RFM analysis, and rule-based clustering fail to capture dynamic and behavioral patterns in customer data. Inaccurate Sales Forecasting Traditional sales forecasting techniques, such as moving averages or simple regression models, fail to capture complex demand patterns, seasonality, and external market factors. Limited Utilization of Customer Sentiment and Unstructured Data Customer feedback, product reviews, and social media interactions contain rich in sights but remain underutilized in business decision-making. Businesses lack automated sentiment analysis to assess customer satisfaction and in corporate it into predictive models. Lack of Real-Time Decision-Making Businesses need real-time, data-driven insights for quick decision-making, yet most customer segmentation and sales prediction models do not function dynamically. Existing systems lack automated dashboards and real-time analytics, making it difficult for businesses to adjust marketing and inventory strategies based on live data.

METHODOLOGY

Customer Segmentation Approaches

K-Means Clustering

K-Means Clustering is applied in order to segment the customers into separate clusters on the basis of the features such as age, income and spending score. The algorithm divides the data into k clusters, and each subscriber is assigned to one of the clusters if the subscriber is nearest to one of the cluster means.

DBSCAN Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is employed to handle complex customer data that may contain noise and irregular cluster shapes.

Sentiment Analysis

VADER Sentiment Analysis (with intensity scores)

For the analysis of customer feedback, VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis is utilized. VADER proves to be highly effective in evaluating sentiments in social media interactions and customer reviews by identifying polarity (positive, negative, neutral) and measuring intensity.

Sales Forecasting Models

Linear Regression

Linear Regression serves as a foundational model for forecasting sales based on variables such as marketing expenditure, product pricing, and customer demographics.

Time-Series Forecasting

ARIMA

ARIMA (AutoRegressive Integrated Moving Average) is employed for short-term sales prediction. It analyzes time-series data by incorporating previous values and errors, capturing trends and seasonality in historical sales data.

LSTM

LSTM (Long Short-Term Memory) networks are utilized for extended sales forecasting. As a form of recurrent neural network (RNN), LSTM has the ability to grasp temporal dependencies and patterns in sequential data.

ALGORITHMS

1. Clustering for Customer Segmentation

K-Means Clustering

Objective: Segment customers into groups with similar characteristics by minimizing the variance within each cluster.

d(x,y)=sqrt(sum_{k=1}^n(x_k-y_k)^2)

2. Decision Trees and Random Forest Classifier

- **Objective:** Partition the feature space to predict outcomes.
- **Entropy (for splitting):**

Entropy = -sum_{i=1}^k p_i log(p_i)

3.Gradient Boosting (e.g., XGBoost, LightGBM)

Objective: Sequentially improve predictions by learning residual errors.

where:

- t = number of iterations
- f_k = decision tree in each iteration
- Ω = regularization term to control complexity

L = sum_{i=1}^N Loss(y_i, y_i^{(t)}) + sum_{k=1}^t Omega(f_k)

RESULTS

Fig 1. Customer Segmentation



Fig . Sales and Order Frequency

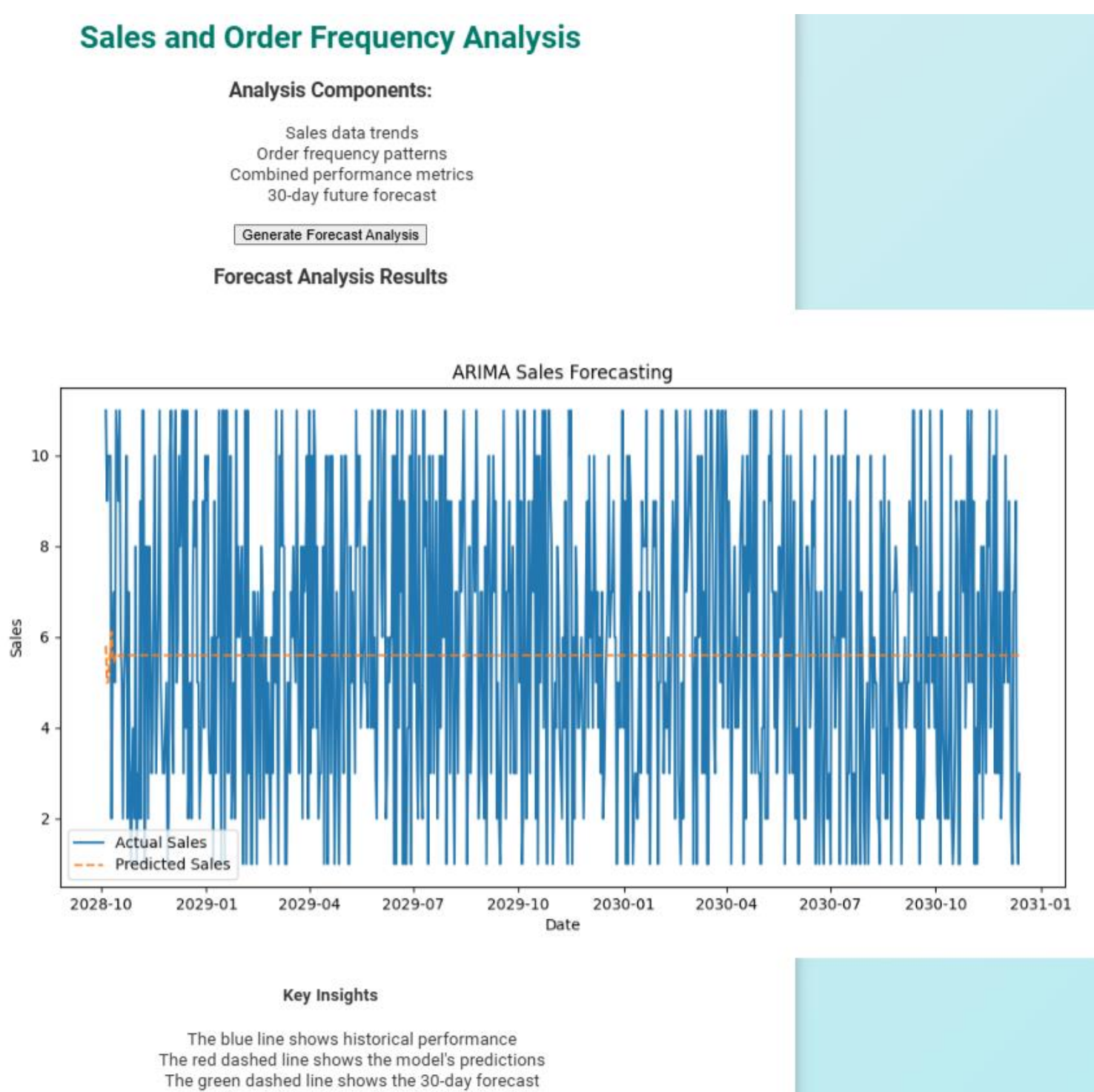


Fig 3. Sentiment Analysis of Customer

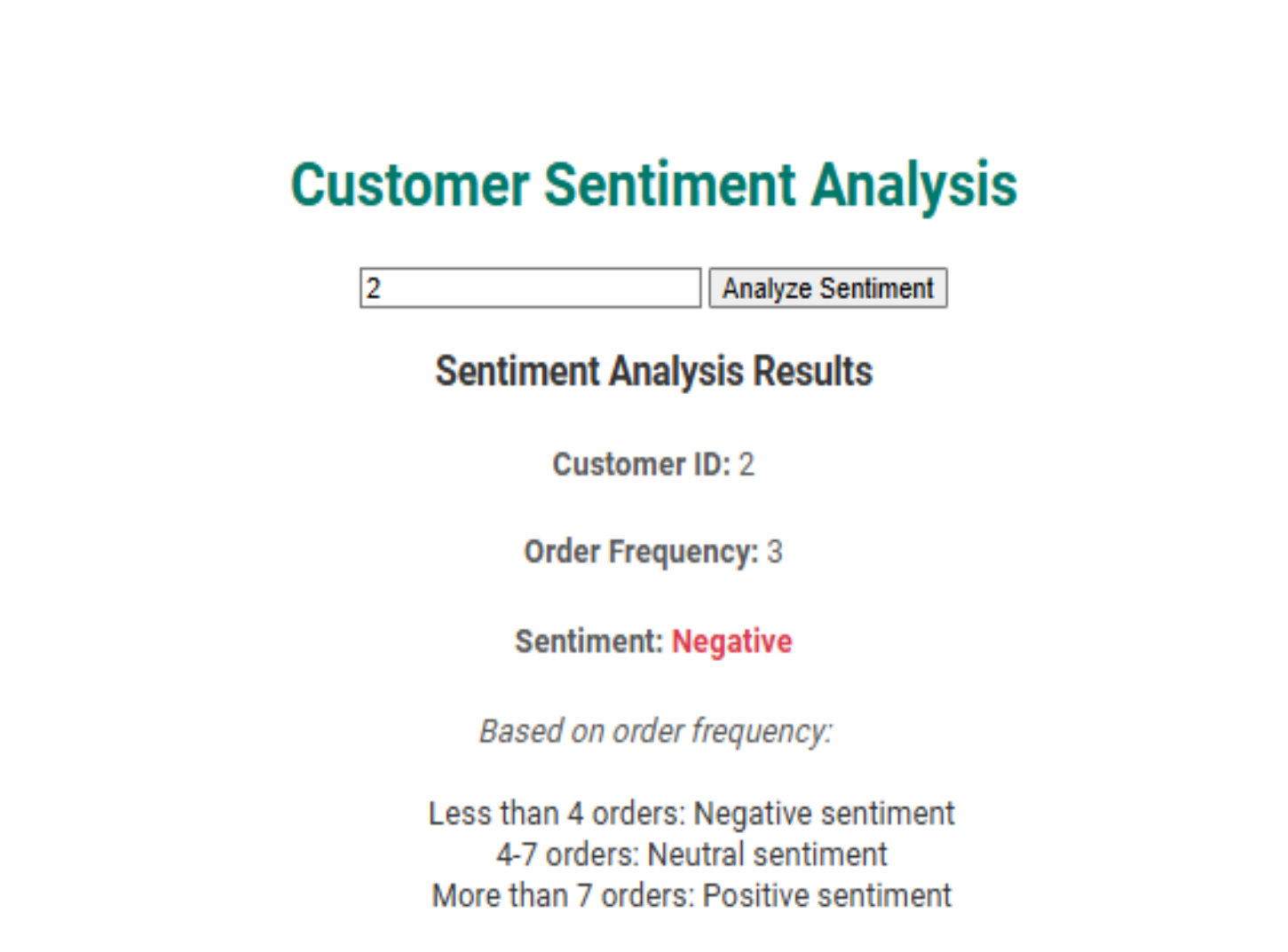
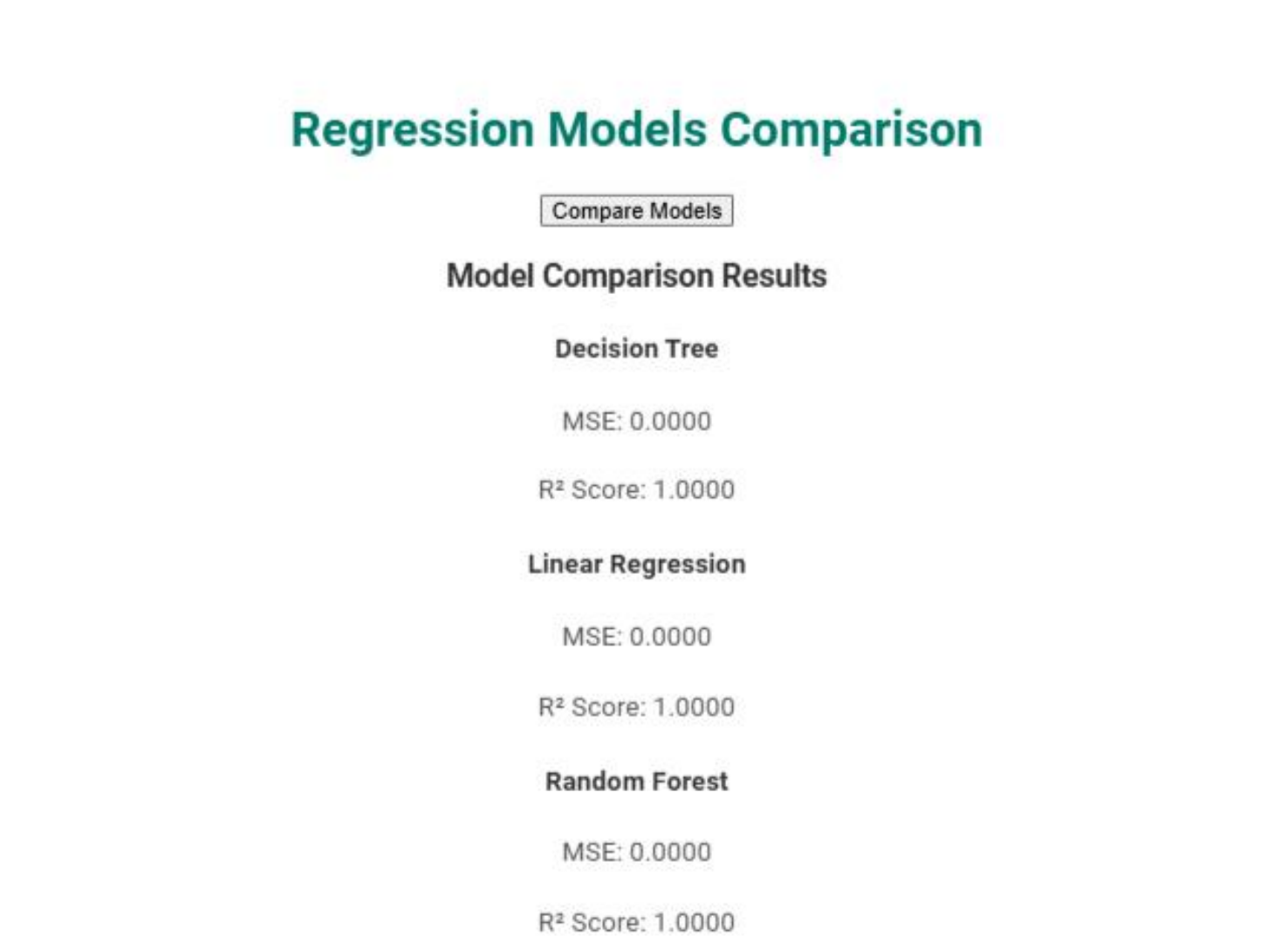


Fig 4. Regression Models Comparison



CONCLUSION

Customer profiling, segmentation, and sales prediction using machine learning collectively provide a powerful framework for understanding and targeting customers with precision. By leveraging advanced tools and technologies, businesses can create detailed customer profiles, segment their audience effectively, and forecast sales ML trends accurately. This integrated approach allows for highly personalized marketing strategies, optimized resource allocation, and proactive decision-making.

LITERATURE REFERENCES

[1] Kasem, M. S., Hamada, M., & Taj-Eddin, I. (2024). Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Computing and Applications, 36(9), 4995-5005.

[2] Alves Gomes, M., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. Information Systems and e-Business Management, 21(3), 527-570.

[3] Zhou, J., Wei, J., & Xu, B. (2021). Customer segmentation by web content mining. Journal of Retailing and Consumer Services, 61, 102588.

[4] Das, S., & Nayak, J. (2022). Customer segmentation via data mining techniques: state-of-the-art review. Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021, 489-507.