# 7th International Conference on Communications and Cyber-Physical Engineering (ICCCE 2024)

**Springer**

**Title of Paper:-** Customer Satisfaction Prediction in Online Goods Delivery through Interpretable Predictive Models and Sentiment Analysis

**Paper ID:-** 574

**Track Name:-** *Track 2I : Transformative Applications of Machine Learning*

**PRESENTERS NAME & AFFILIATION:-** Akula Venkata Satya Sai Gopinadh, Sri Sathya Sai Institute of Higher Learning

**NAME OF ALL THE AUTHORS:-** Akula Venkata Satya Sai Gopinadh, SVSN Sarma, Gudipudi Radhesyam

# Introduction

❖ Ensuring customer satisfaction post-purchase is crucial for success in the competitive online shopping landscape, focusing on understanding customer preferences and delivery satisfaction.

❖ Post-purchase support is a key determinant of customer satisfaction, critical for customer retention and repeat purchases, beyond just advertising and promotions.

❖ Examining customer reviews provides insights into preferences and grievances, helping businesses improve and optimize the shopping experience for all patrons.

# Literature Survey

Machine Learning Approach to Predict E-commerce Customer Satisfaction Score (2023): Researchers from Thailand found that the Random Forest model best predicted customer satisfaction in online shopping, emphasizing the importance of product quality, delivery speed, and shipping duration.

Sentiment Analysis on Brazilian Portuguese user reviews: The study on Brazilian Portuguese user reviews compares document embedding techniques and classifiers, finding that TF-IDF with larger vocabulary sizes outperforms word embeddings in sentiment analysis and highlights vocabulary size importance for model generalization.

Optimising e-commerce customer satisfaction with machine learning: This paper uses four machine learning algorithms to predict customer satisfaction, finding that Random Forest achieved the highest accuracy. Key factors were meeting the estimated delivery date and delivery time, with feature normalization and imbalanced data treatment evaluated.

# Method



**Dataset:** Brazilian E-Commerce Public Dataset by Olist. **Source:** Kaggle

**Preprocessing:** Merging of 9 different files into a single file, and applying MINMAX Scaling(Numerical Columns).
**Imbalance Treatment**: SMOTE(Synthetic Minority Oversampling Technique).

**Feature Engineering:** Feature engineering enhanced models with new features: Estimated_number_of_days_df, arrival_punctuality, product_rating, and seller_rating. Sentiment analysis using DistilBERT generated review_positive_score, review_negative_score, and review_neutral_score for customer feedback analysis.

**Feature Selection:-** Feature selection using Random Forest identifies top features: product_rating, review_positive_score, review_negative_score, review_neutral_score, Estimated_number_of_days_df, seller_rating, and total_shipping_duration, enhancing model performance by removing redundant, irrelevant, or noisy features.
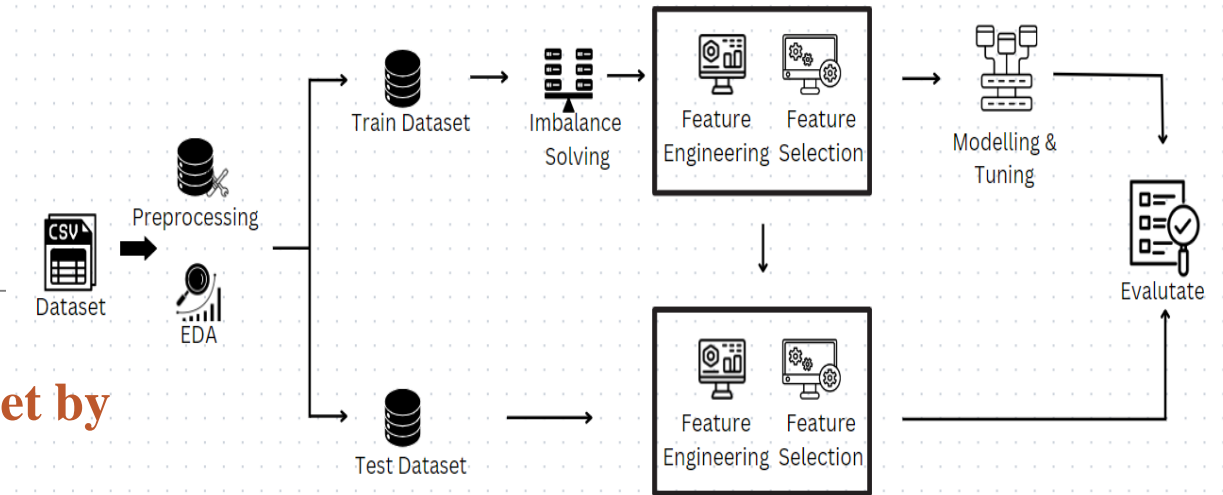
# Result

| | Models | Macro Precision | Macro Recall | Macro F1-Score |
|---|---|---|---|---|
| **Results** | **XGB** | 0.53 | 0.52 | 0.53 |
| | **RF** | 0.53 | 0.53 | 0.52 |
| | **DT** | 0.51 | 0.51 | 0.50 |
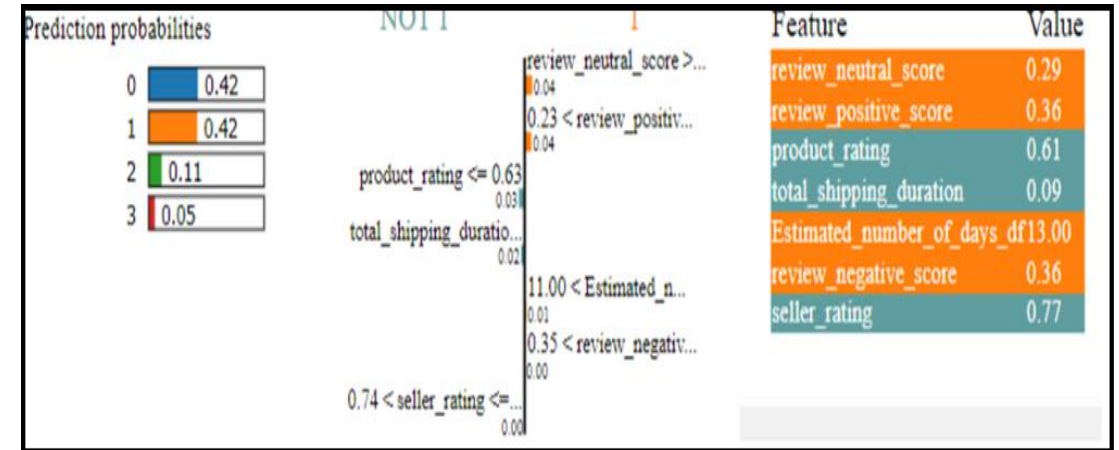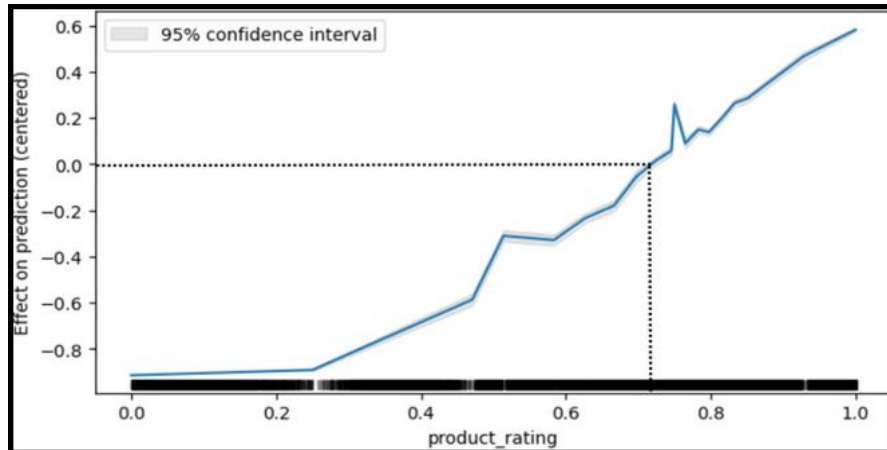
# Discussion

- Improving Prediction Accuracy: Enhance the prediction accuracy of customer satisfaction scores by identifying key influential features in e-commerce.

- Key Influential Features: Product and seller ratings, estimated delivery duration, and review sentiment scores.

- Impact on Business Strategies: Strong relationships between these factors and customer satisfaction aid in refining strategies, services, logistics, and support.

- Conclusion: Provides insights into customer satisfaction dynamics, optimizing operations, and enhancing customer experiences through advanced machine learning techniques.