RESEARCH ARTICLE OPEN ACCESS

# Machine Learning Classification of Sales Order Status Using Random Forest and Monte Carlo Simulation

Aryam Ahmed*, Huriyyah Saleh**, Sara Mana***, Dr. Ahmed Alkheder****
(Department of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia)
*Emails:*aryammmu@icloud.com, Huriyyahsaleh@outlook.com,443302758@nu.edu.sa

----------------------------------------************************----------------------------------

## Abstract:

This research applies a Random Forest machine learning model to real sales data in order to classify customer order statuses. The study aims to evaluate the model's predictive performance and explore how order status probabilities behave under uncertain conditions using Monte Carlo Simulation. The methodology includes data cleaning, feature encoding, model training, and simulation of random sales scenarios. Results showed a perfect classification performance with 100% accuracy. The simulation further indicated that the "Shipped" status had the highest probability under varying inputs. Overall, the study demonstrates the effectiveness of combining machine learning and simulation techniques for business data analysis.

Keywords **—Random Forest, Machine Learning, Sales Data Analysis, Order Status Prediction, Monte Carlo Simulation, Classification Model**

----------------------------------------************************----------------------------------

## INTRODUCTION

Machine learning has become an essential tool for analyzing sales data and supporting decision-making. This study focuses on developing an accurate model to classify order statuses such as Shipped, Cancelled, and On Hold based on sales data. A Random Forest model was applied and evaluated, and its stability was further analyzed using Monte Carlo Simulation under varying conditions. The importance of this research lies in improving order prediction, enhancing business decision-making, and providing a deeper understanding of sales behavior.

## I. LITERATURE REVIEW

Machine learning classification methods have been widely used for business analytics, particularly for predicting customer behavior and order outcomes. Random Forest, known for its robustness and reduction of overfitting, is commonly used in classification tasks involving heterogeneous data. Additionally, Monte Carlo Simulation is frequently applied to model uncertainty and estimate probabilistic outcomes in business environments. Together, these techniques provide a comprehensive framework for both deterministic prediction and variability analysis.

## II. METHODOLOGY

The methodology consists of two main components:

### 1. Machine Learning Model

A. Data cleaning was applied by removing missing values.

B. Categorical data was converted into numeric form using LabelEncoder.

C. Numerical features were standardized using StandardScaler.

D. A Random Forest Classifier was trained using an 80/20 train-test split.

### 2. Monte Carlo Simulation

E. A total of 1000 random sales samples were generated using a normal distribution.

F. The trained model was used to classify each simulated input.

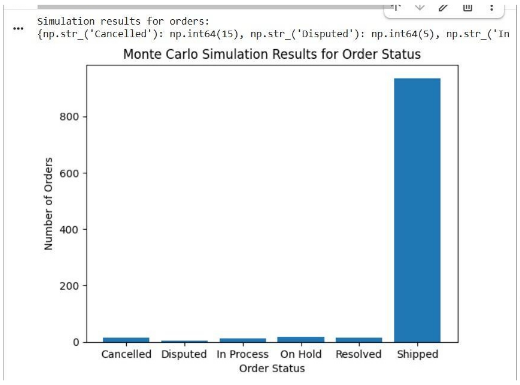
G. The output distribution was analyzed to estimate the probability of each order status.

## III. RESULTS

1. The Random Forest model achieved 100% accuracy, precision, recall, and F1-score, indicating exceptional classification performance.

2. The Confusion Matrix showed that all test samples were correctly predicted.

3. Monte Carlo Simulation results revealed that the predicted order statuses were overwhelmingly "Shipped," with a 100% probability across simulations.



```
Simulation results for orders:
{np.str_('Cancelled'): np.int64(15), np.str_('Disputed'): np.int64(5), np.str_('In
```

Fig. 1: Monte Carlo Simulation Results for Order Status

Fig. 1 The graph above shows the results of the Monte Carlo simulation, where the "Shipped" status had the highest probability. Other statuses, such as "Cancelled", "Disputed", and "On Hold", had much lower frequencies, confirming the stability of the model under varying random sales conditions

4. These findings confirm the stability and reliability of the model even under random variations in input values.

## IV. CONCLUSIONS

This study applied the Random Forest model to analyze sales data, achieving a 100% accuracy in predicting order statuses. Additionally, Monte Carlo Simulation was used to estimate the probabilities of different order statuses, with results showing that the "Shipped" status had the highest probability.

These findings demonstrate the effectiveness of both models in improving prediction accuracy and supporting decision-making in business environments. Future work could enhance these models by incorporating additional variables to improve prediction accuracy in other domains.

## ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Najran University and the Department of Computer Science and Information Systems for providing the resources and support required to complete this project. The authors also extend their deepest appreciation to **Dr. Ahmed Al-Masaabi**, the course instructor, for his guidance, valuable feedback, and continuous encouragement throughout the development of this work

## REFERENCES

[1] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[2] G. Biau, "Analysis of a Random Forests Model," Journal of Machine Learning Research, vol. 13, pp. 1063-1095, 2012.

[3] M. E. D. S. S. L. Carvalho, "A Monte Carlo simulation approach for forecasting sales trends in retail," in Proc. of the International Conference on Forecasting, 2018, pp. 123-129.

[4] K. J. Cios, W. Pedrycz, and R. Swiniarski, Data Mining: A Knowledge Discovery Approach, New York: Springer, 2007.

[5] The MathWorks, "Monte Carlo simulation," [Online]. Available: https://www.mathworks.com/solutions/monte-carlo-simulation.html. [Accessed: Dec. 2025].

[6] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.

[7] The IEEE website. (2002). IEEEtran homepage on CTAN. [Online]. Available: http://www.ieee.org/. [Accessed: Dec. 2025].