

Airline Passenger Satisfaction Determinants

Abstract

Background: Understanding the factors influencing airline passenger satisfaction is essential for the aviation industry to enhance customer experience and loyalty. This research aims to identify key determinants of passenger satisfaction using exploratory data analysis (EDA) and machine learning algorithms.

Objectives: Investigate the impact of various factors including online boarding, in-flight services, flight distance, legroom service, age, ease of online booking, seat comfort, departure and arrival time convenience, baggage handling, gate location, and cleanliness on airline passenger satisfaction.

Methods: Employed EDA techniques and machine learning algorithms, including random forest, probit, and naive Bayes models. Applied undersampling and bootstrapping methods to address sample imbalance. Evaluated model performance using ROC/AUC curves and confusion matrices.

Results: Identified online boarding, in-flight services, flight distance, legroom service, age, ease of online booking, seat comfort, departure and arrival time convenience, baggage handling, gate location, and cleanliness as significant determinants of passenger satisfaction. Developed models achieved high accuracy ranging from 88% to 99%.

Conclusion: This research provides valuable insights into factors driving airline passenger satisfaction, offering actionable information for airlines to improve customer experience and loyalty. The robustness of the models and high accuracy rates underscore the effectiveness of machine learning approaches in analyzing passenger satisfaction data.

Data

The data utilized in this research paper originates from the "Airline Passenger Satisfaction" dataset, which is available on Kaggle. Access to the dataset can be obtained through the following link: <https://www.kaggle.com/datasets/mysarahmadbhat/airline-passenger-satisfaction> . This dataset encompasses a diverse range of features related to airline passenger experiences, including factors such as online boarding, in-flight services, flight distance, legroom service, age, ease of online booking, seat comfort, departure and arrival time convenience, baggage handling, gate location, and cleanliness. The dataset serves as a valuable resource for analyzing passenger satisfaction levels within the aviation industry.

Methodology & Results

Result 1: Initially, missing values were removed from the dataset to ensure data integrity. Subsequently, correlation analysis was conducted to identify and eliminate highly correlated features, reducing redundancy and potential multicollinearity in the dataset. The highest correlation is observed between departure and arrival delay and through a regression analysis it is observed that arrival delay is mostly caused due to departure delay. Therefore, only arrival delay is kept in the dataset.

See full results in the Appendix 1.

Result 2: The random forest algorithm was employed to analyze passenger satisfaction, utilizing three variations: the entire dataset, bootstrapping, and undersampling techniques. The results consistently highlighted online boarding, in-flight services, flight distance, legroom service, age, ease of online booking, seat comfort, departure and arrival time convenience, baggage handling, gate location, and cleanliness as crucial features influencing passenger satisfaction.

Result 3: A Probit model was developed to further explore the determinants of passenger satisfaction, with robustness checks performed using bootstrapping and undersampling methods (see graphs 2 – 7). The results demonstrated consistency across all methodologies, reaffirming the significance of the identified features in predicting passenger satisfaction levels.

See full results in the Appendix 2.

Result 4: Additionally, a naive Bayes model was utilized to assess the robustness of the findings, incorporating bootstrapping and undersampling techniques for validation (see graphs 8 – 13). The results remained robust across different methodologies, further validating the importance of the identified features in determining passenger satisfaction levels.

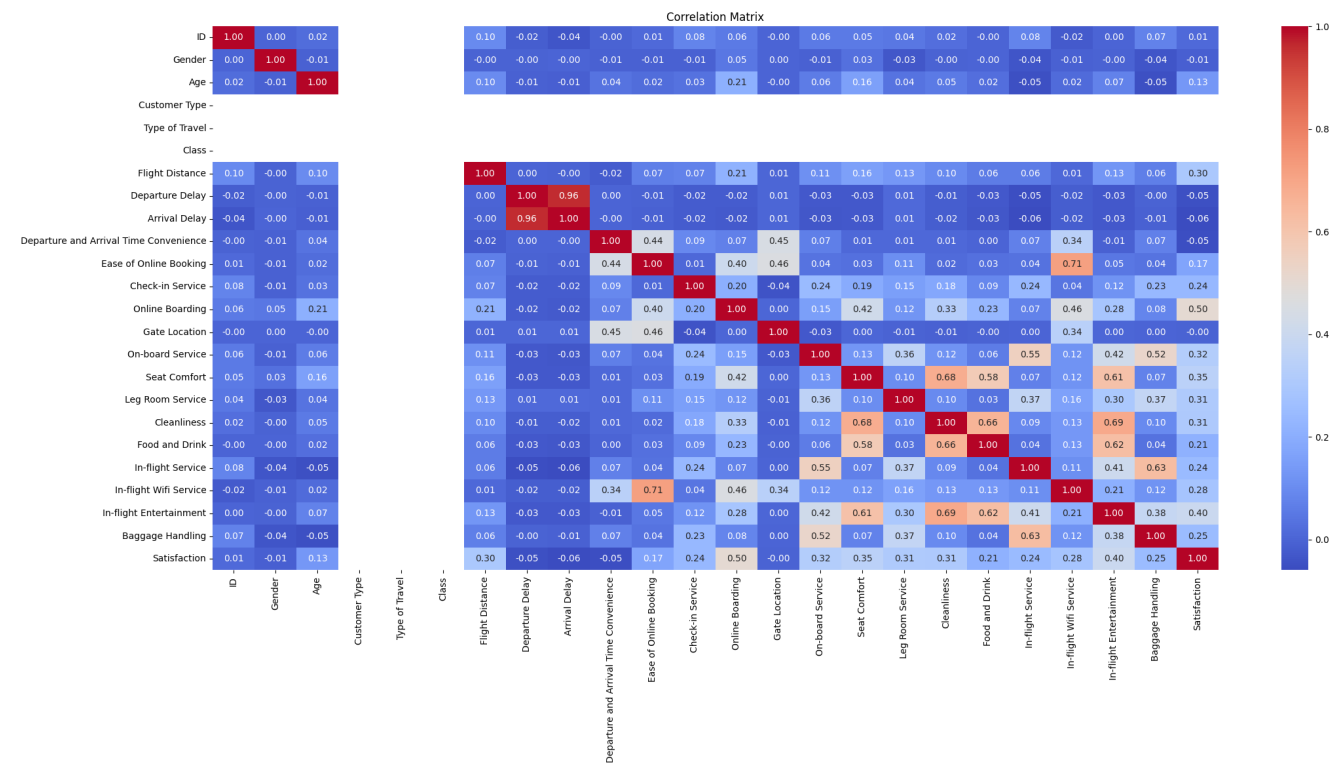
See full results in the Appendix 3.

Conclusion

In conclusion, this research sheds light on the multifaceted determinants of airline passenger satisfaction through a comprehensive analysis utilizing exploratory data techniques and machine learning algorithms. By systematically cleaning the data and employing advanced modeling techniques such as random forest, Probit, and naive Bayes, key features impacting passenger satisfaction were identified and validated. The robustness of the findings across different sampling methods underscores the reliability of the results. These insights offer valuable guidance for airline operators in prioritizing areas for improvement to enhance customer experience and foster loyalty. Moving forward, continued research in this area is crucial for staying attuned to evolving passenger preferences and maintaining competitiveness in the dynamic aviation industry landscape.

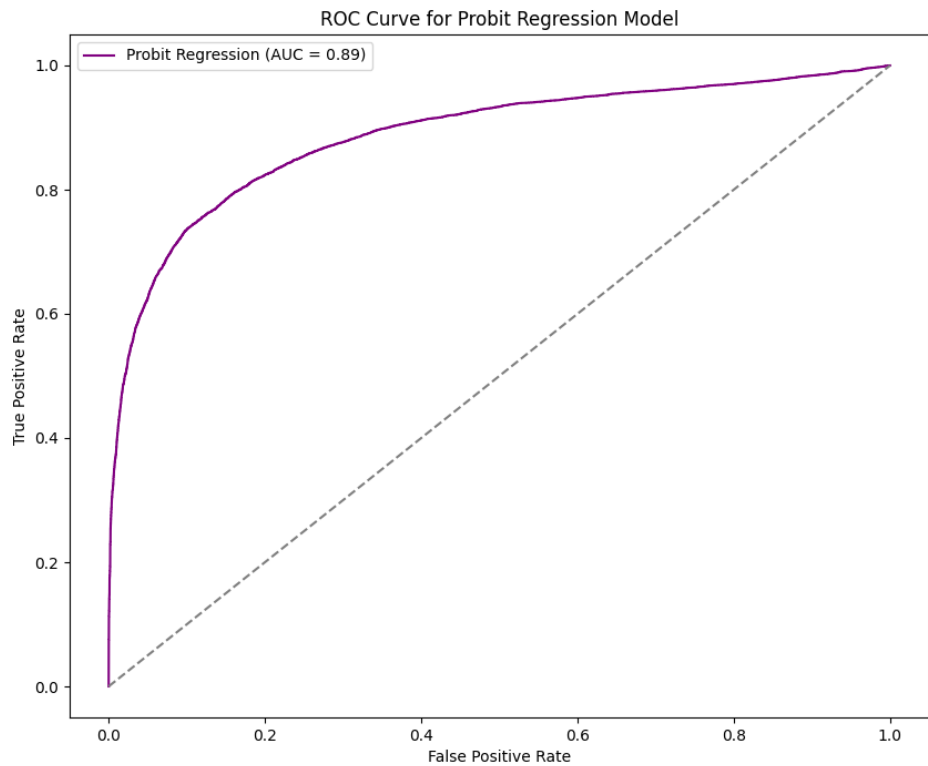
Appendix 1

Graph 1. Correlation Matrix

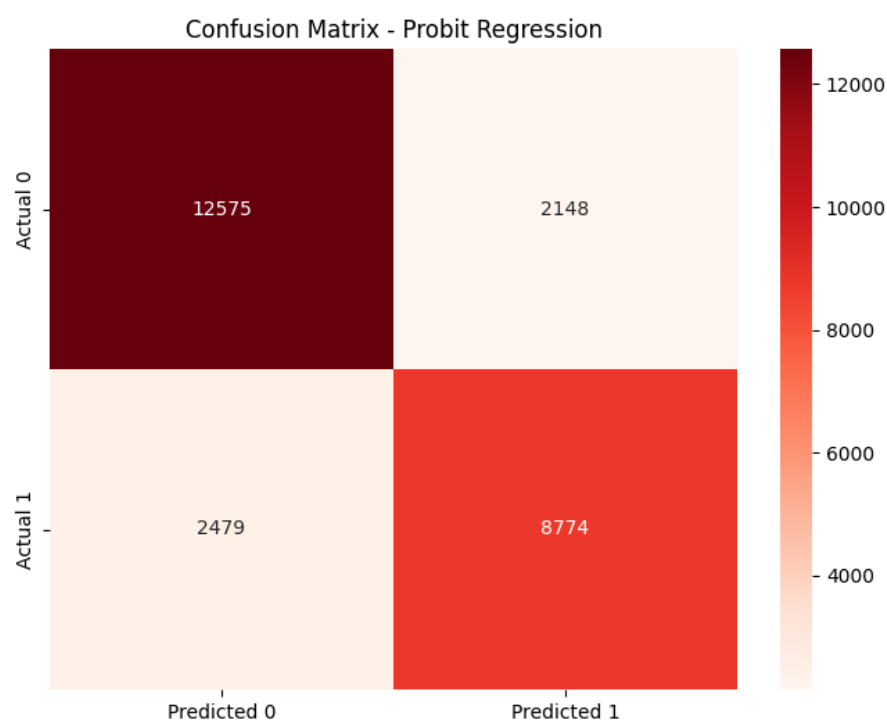


Appendix 2

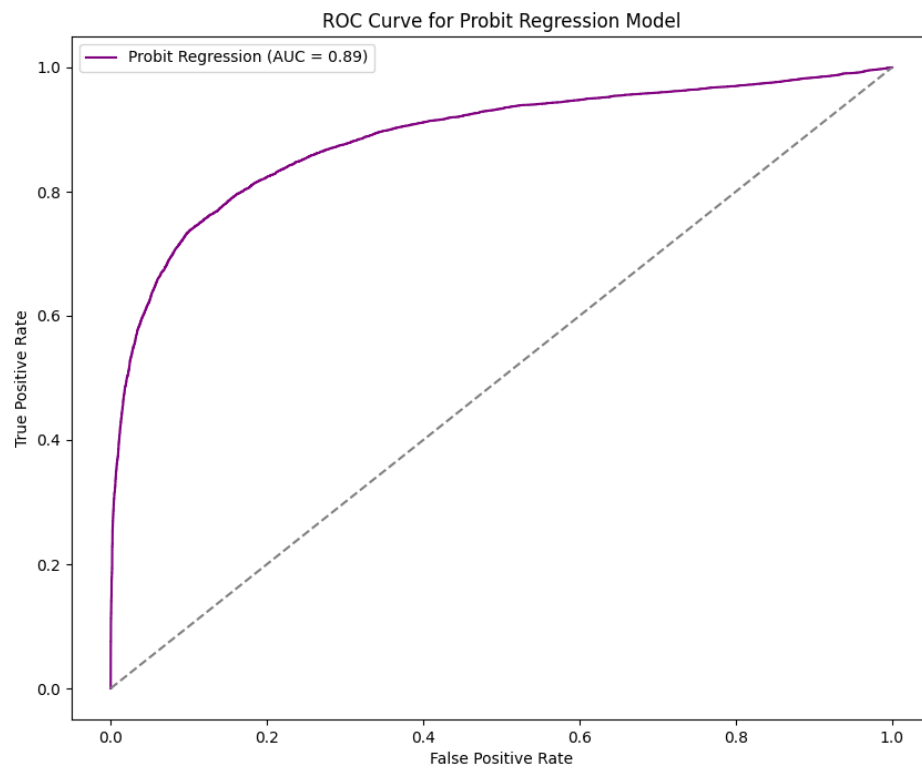
Graph 2. Simple Probit ROC/AUC



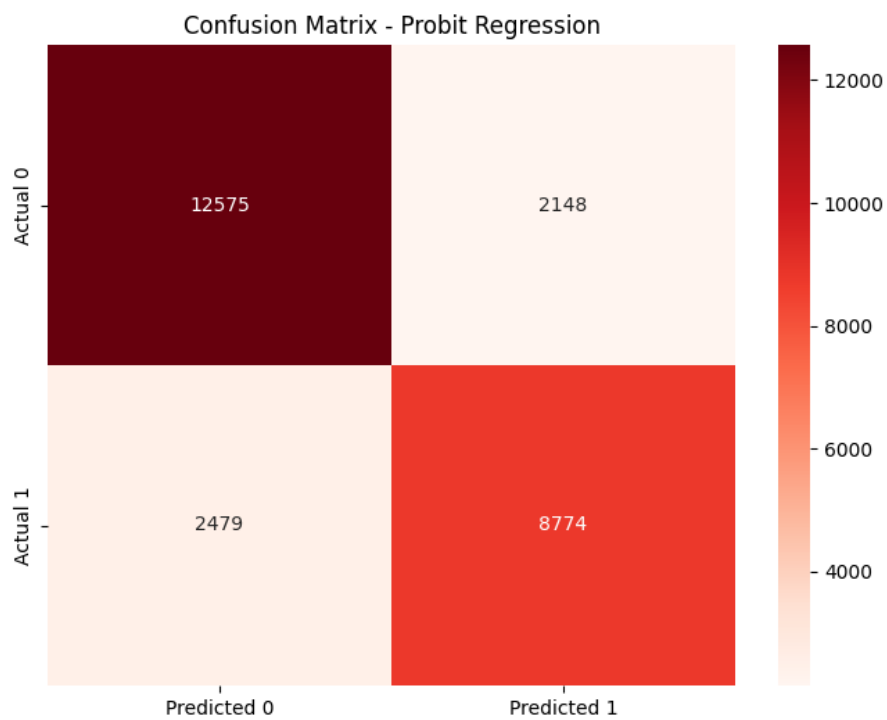
Graph 3. Simple Probit Confusion Matrix



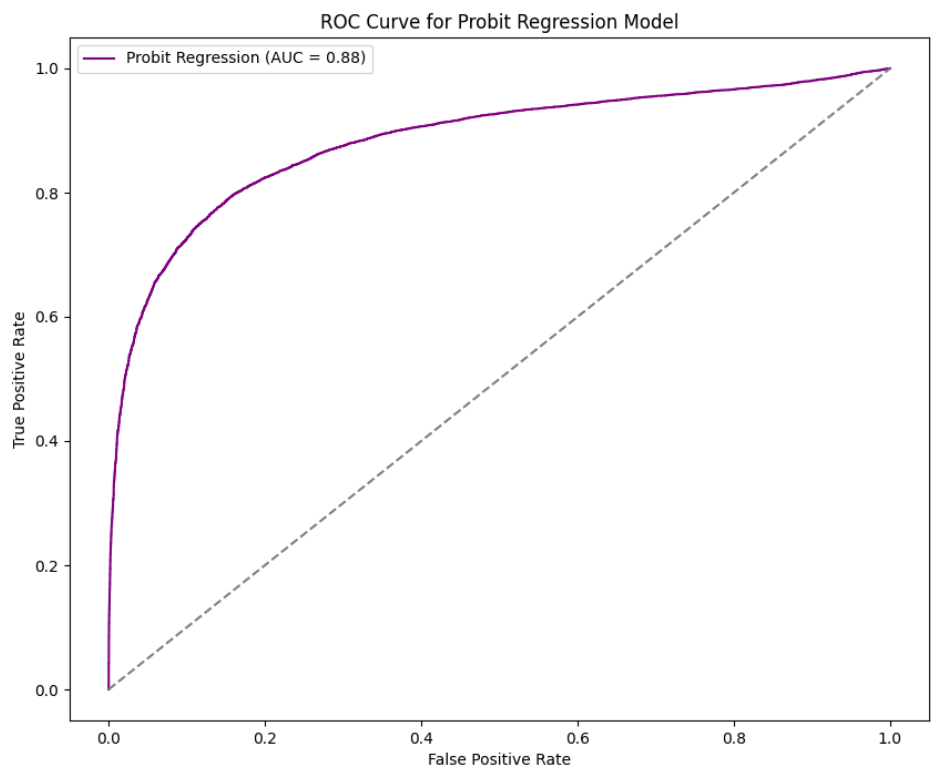
Graph 4. Bootstrap Method Probit ROC/AUC



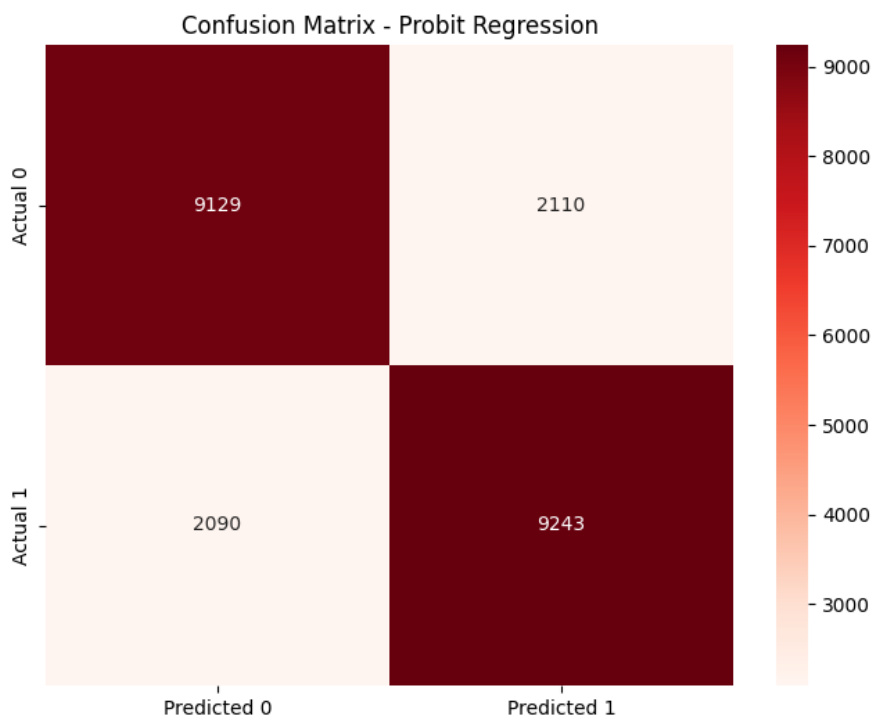
Graph 5. Bootstrap Method Probit Confusion Matrix



Graph 6. Undersampling Method Probit ROC/AUC

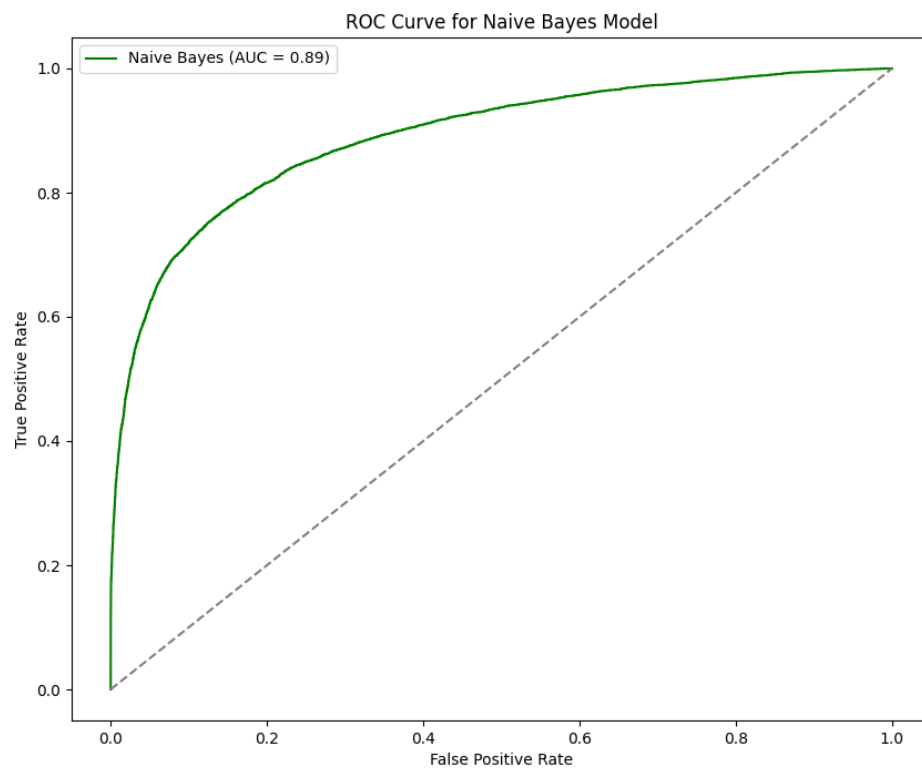


Graph 7. Undersampling Method Probit Confusion Matrix

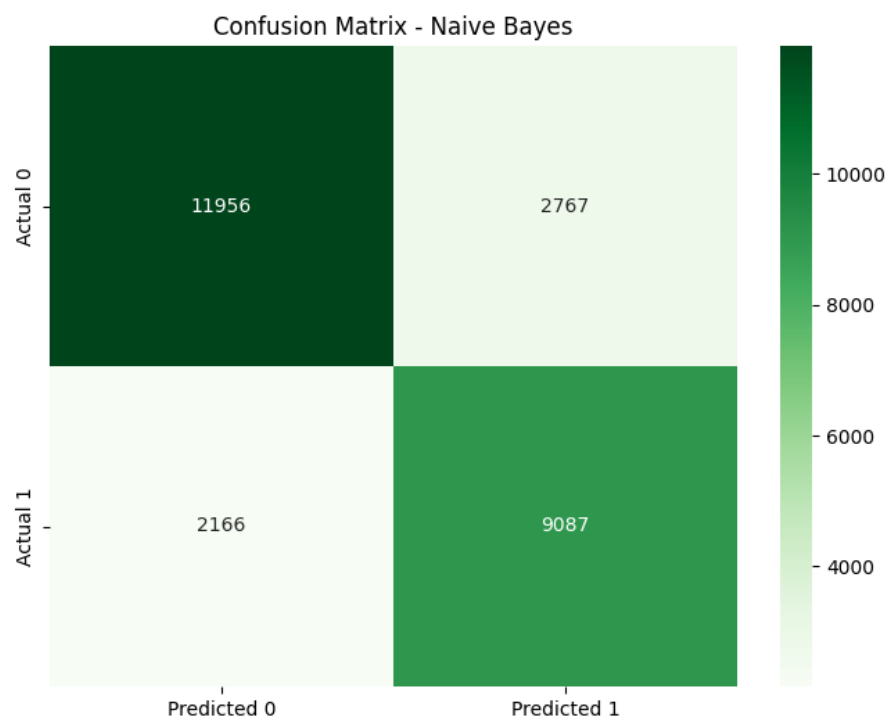


Appendix 3

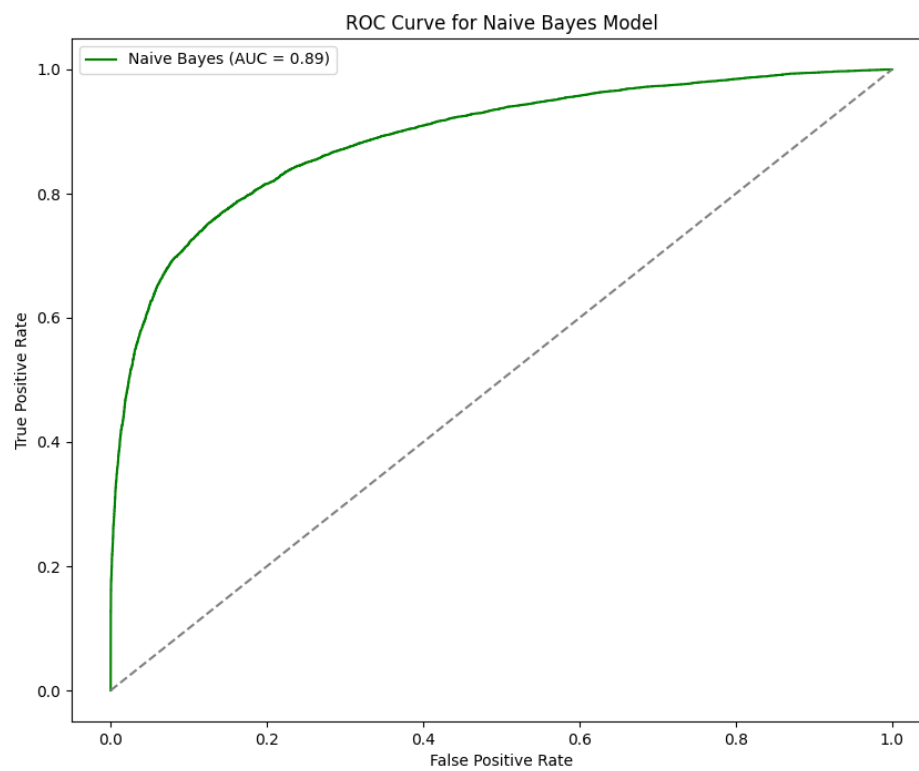
Graph 8. Simple Naive Bayes ROC/AUC



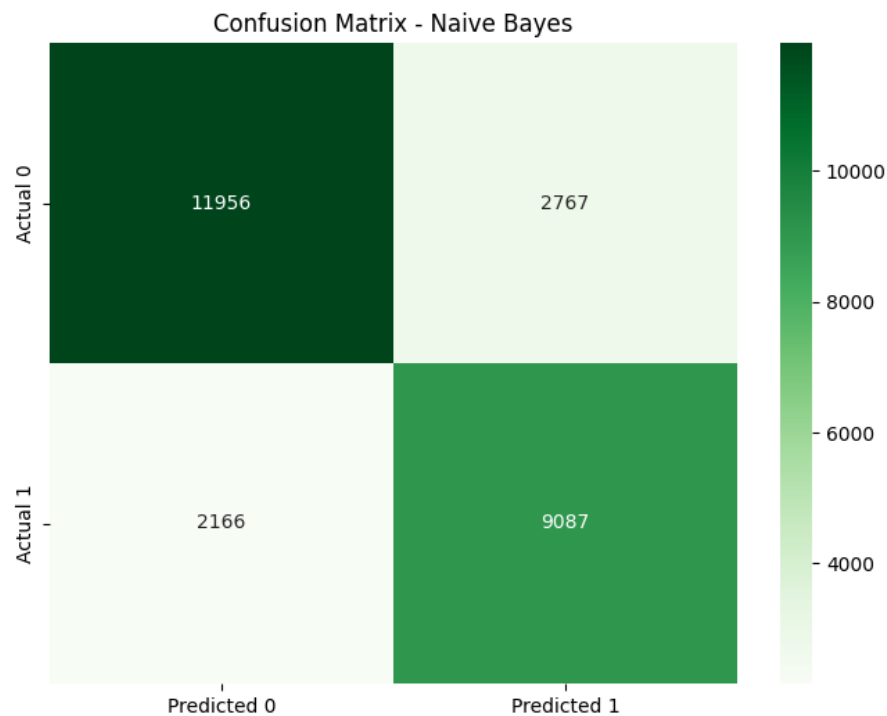
Graph 9. Simple Naïve Bayes Confusion Matrix



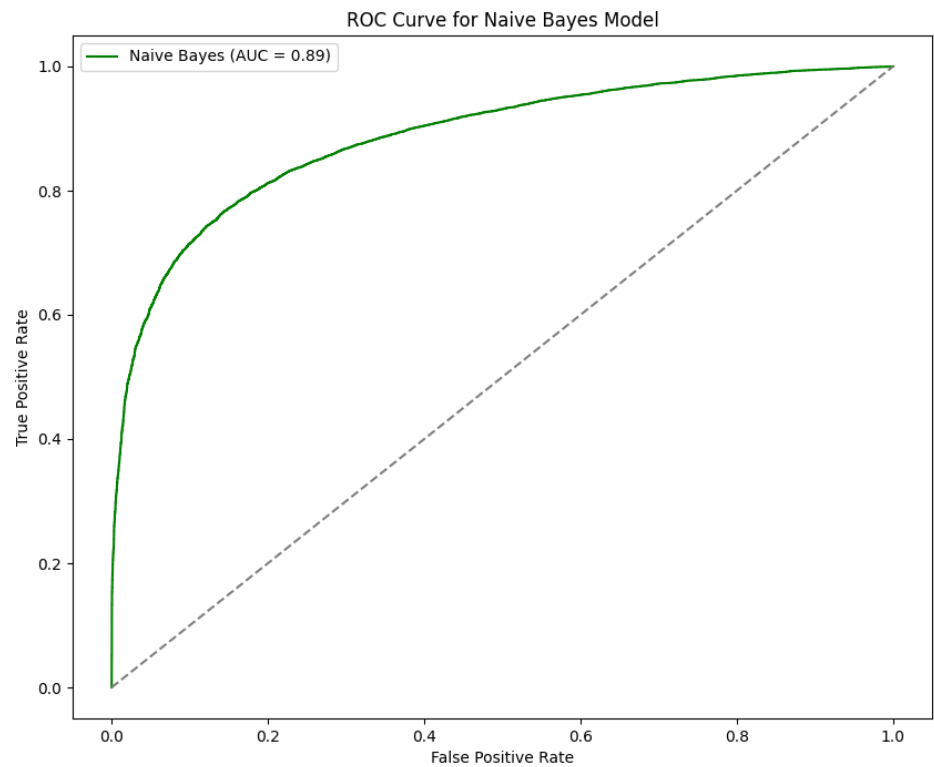
Graph 10. Bootstrap Method Naive Bayes ROC/AUC



Graph 11. Bootstrap Method Naïve Bayes Confusion Matrix



Graph 12. Undersampling Method Naive Bayes ROC/AUC



Graph 13. Undersampling Method Naïve Bayes Confusion Matrix

