James-tech4 /
JAMES-PHASE-3-DSPT-8-PROJECT

Code Issues Pull requests Actions Projects Wiki Security Insights Setting

Phase 3 Project

0 stars 0 forks 1 watching Branches Activity
Tags

Public repository

1 Branch 0 Tags Go to file t Go to file Add file Code

James-tech4 updated a69241e · 13 minutes ago

Data	updated	yesterday
Visualization Images	updated	2 hours ago
.gitignore	Added a git ignore file	4 hours ago
Hotel_Reservations.csv	updated	6 hours ago
Presentation.pdf	updated	30 minutes ago
Presentation.pptx	updated	42 minutes ago
README.md	updated	13 minutes ago
index.ipynb	updated	30 minutes ago
index.pdf	updated	30 minutes ago

Project Title: Model to Predict Hotel Booking Cancellations for Optimized Booking Management

Student Details

- Student Name: MUTHEE JAMES WACHIRA
- Student Pace: Part Time
- Instructor Name: Daniel Ekale

Table of Contents

1. Project Overview

This project aims to develop a predictive model to forecast hotel reservation cancellations. By analyzing historical booking data, the model will identify key factors influencing cancellations, enabling hotels to take proactive actions like overbooking, adjusting policies, or offering incentives. The goal is to reduce revenue loss, optimize resource allocation, and improve customer satisfaction by minimizing cancellations and no-shows.

2. Business Problem

Hotels face challenges in managing bookings effectively, as they cannot predict which reservations will be honored. This uncertainty impacts resource allocation, planning, and overall profitability. Predicting cancellations would help hotels optimize booking strategies, reduce losses, and improve customer satisfaction

3. Business Objectives

The objective of this project is to develop a predictive model to accurately forecast hotel reservation cancellations so as to enable hotels to:

- Optimize booking strategies and resource allocation based on cancellation predictions.
- Minimize revenue loss by proactively managing overbooking and cancellation policies.
- Improve customer satisfaction by offering targeted incentives to reduce cancellations

4. Target Audience

The Target audience for this project are:

- Hotel Managers
- Revenue Managers
- Hotel Owners/Executives

5. Data

The Hotel Reservations dataset used in this project was obtained from https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset.

6. Data Understanding

The data used for this project:

- Has 36275 rows and 19 columns
- Has No missing Values
- Has 5 columns with Categorical data types, 13 columns with Integer data type and one column with float data type
- Has no Multicollinearity
- Has columns with Outliers

7. Data Pre_processing

The following processes were carried out to the training and test features before building the models:

- Identified the dependent and independent variables to use for Modelling
- Split the resulting dataframe into train and test splits
- Converted categorical variables to Numerical Variables using One Hot Encoding
- Scaled some numerical columns using Standard Scaler
- Log Transformed the columns that had the greatest skewness

8. Modelling

In this project, 4 Models were created:

1. Logistic Regression Model-Baseline
2. Decision Tree Model-Tuned

3. Random Forest Model-Untuned
4. Random Forest Model- Tuned

9. Model Evaluation

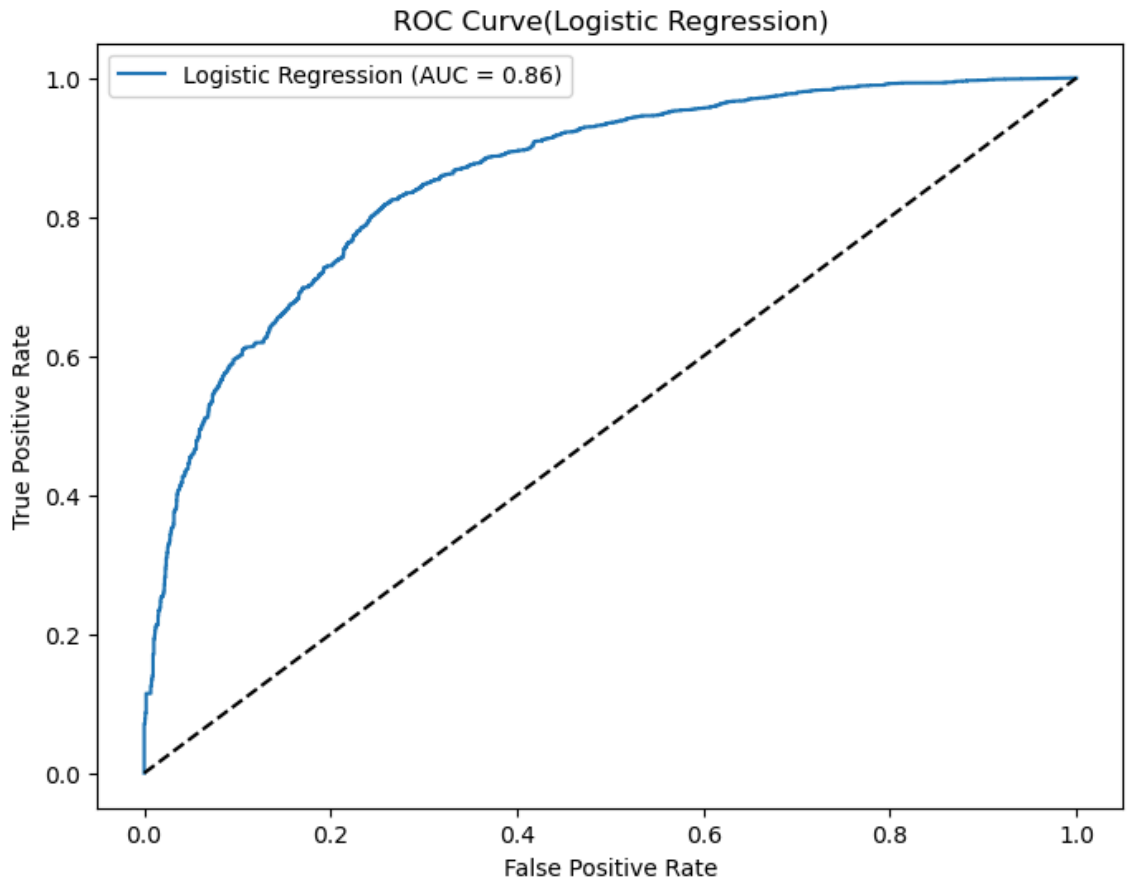
The Models were evaluated on two parameters, accuracy score and ROC-AUC score. Below are the outcomes of the evaluation for each Model

- Logistic Regression Model(Baseline) - Accuracy Score of 79% and AUC of 0.86
- Decision Tree Model(Tuned) - Accuracy score of 84% and AUC of 0.89
- Random Forest Model (Untuned) - Accuracy Score of 85% and AUC of 0.91
- Random Forest Model(Tuned) - Accuracy Score of 86% and AUC of 0.92

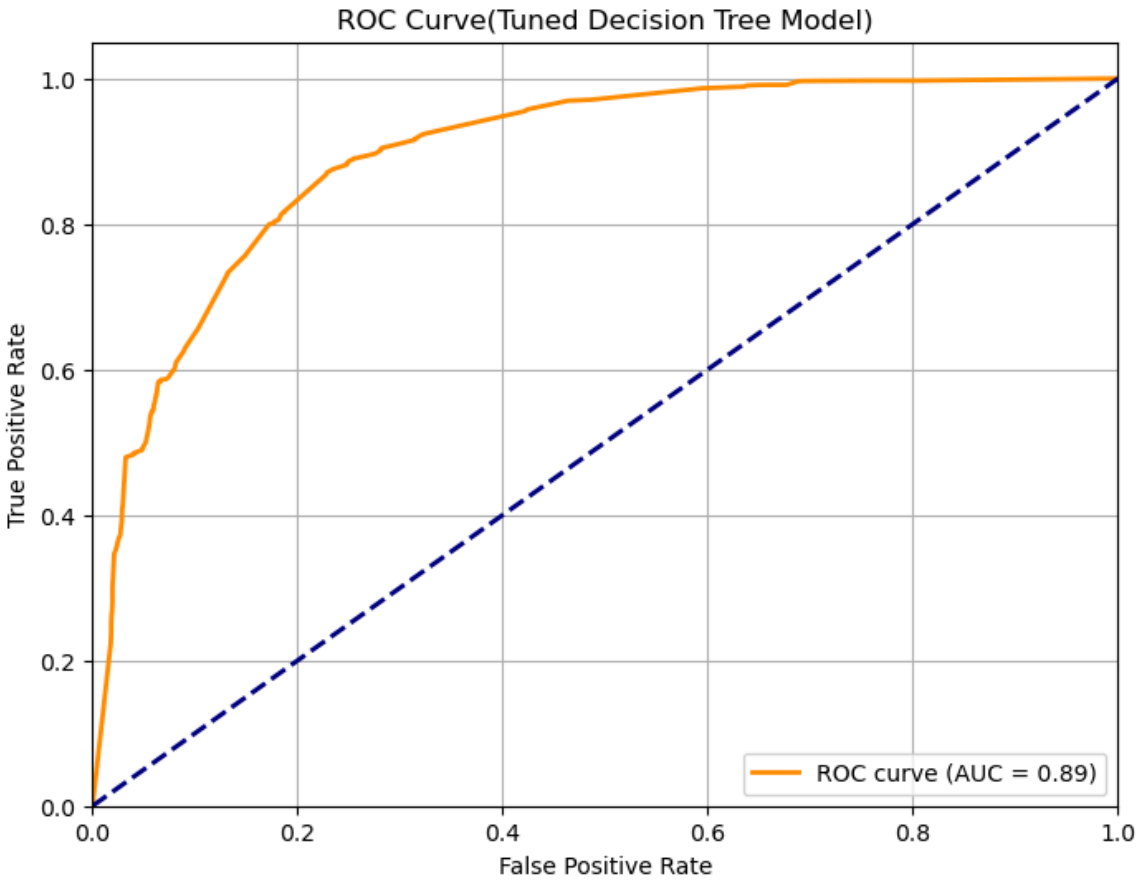
10. ROC-AUC Visualizations

After building our models, ROC Curve for each Model was plotted. The ROC Curve shows the AUC(The Area Under the Curve). Below are the ROC visualizations for each Model and the combine ROC curves for all the models for better comparison

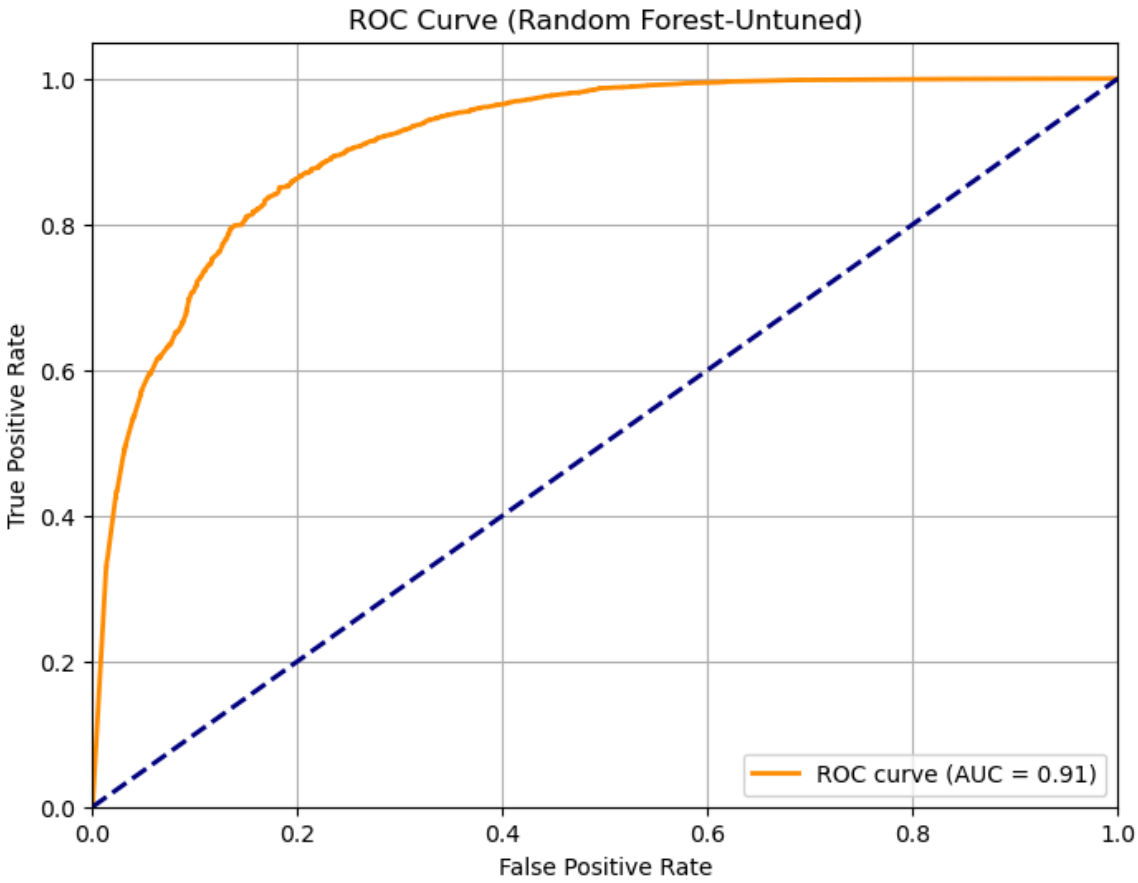
1. ROC Curve-Logistic Regression



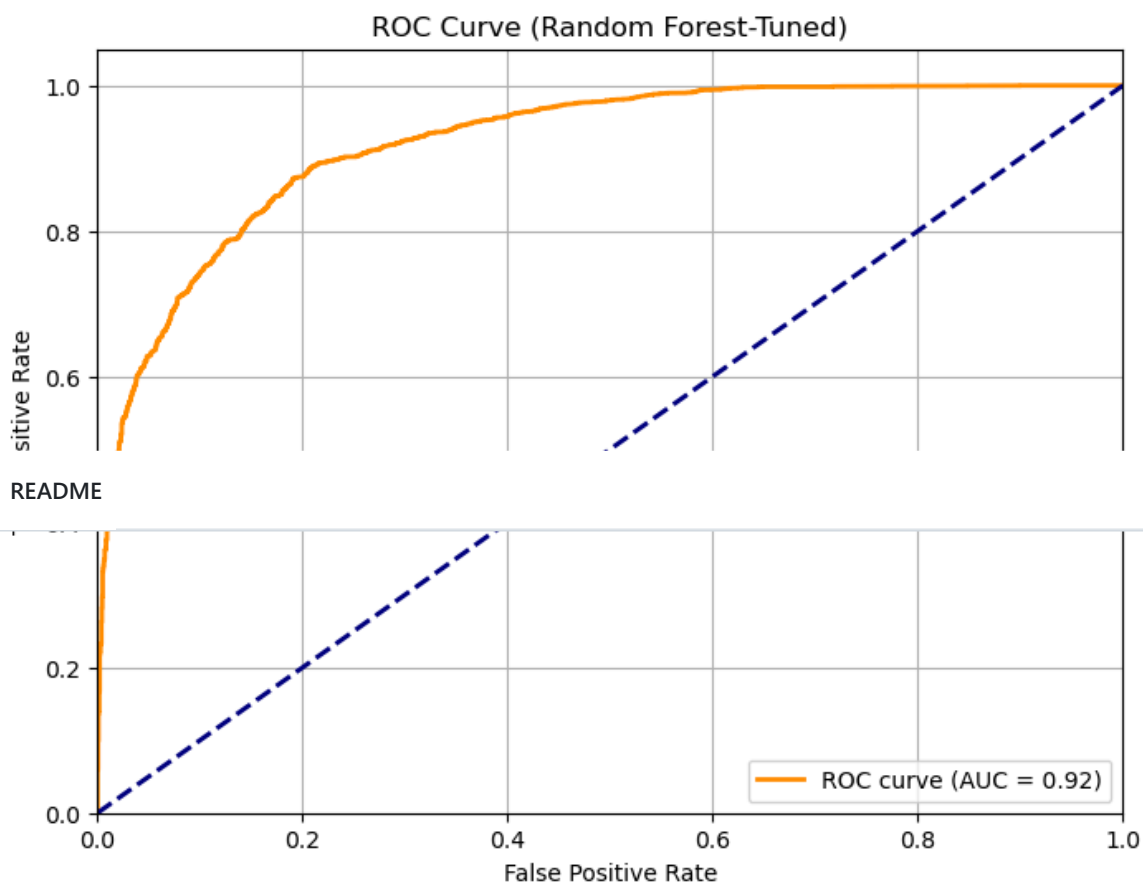
2. ROC Curve - Decision Tree(Tuned)



3. ROC Curve - Random Forest(Untuned)

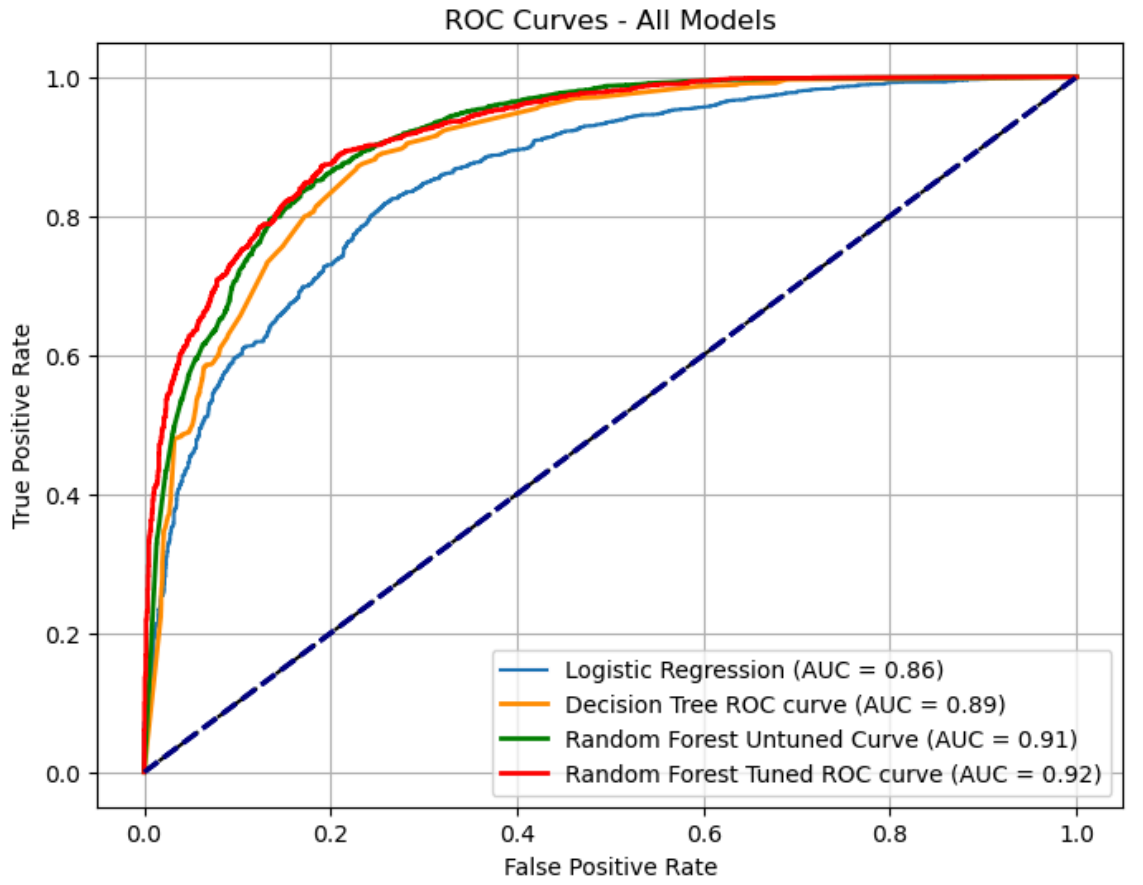


4. ROC Curve - Random Forest(Tuned)



README

5. ROC Curves - All Models



11. Findings

1. Based on the evaluation results, the tuned Random Forest Model outperforms all other models, with the highest Accuracy Score of 86% and the highest AUC of 0.92, indicating that it is the most effective in predicting hotel reservation cancellations.
2. The Untuned Random Forest Model also performed well, achieving an Accuracy Score of 85% and an AUC of 0.91, indicating strong predictive capability even without hyperparameter tuning.
3. The tuned Decision Tree Model showed a notable improvement over the baseline Logistic Regression Model, with an Accuracy Score of 84% and an AUC of 0.89, suggesting that decision trees, when optimized, are a reliable choice, though slightly less effective than Random Forest.
4. The Logistic Regression Model (Baseline), with an Accuracy Score of 79% and an AUC of 0.86, serves as a solid starting point but demonstrates lower performance compared to more complex models like Random Forest.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%