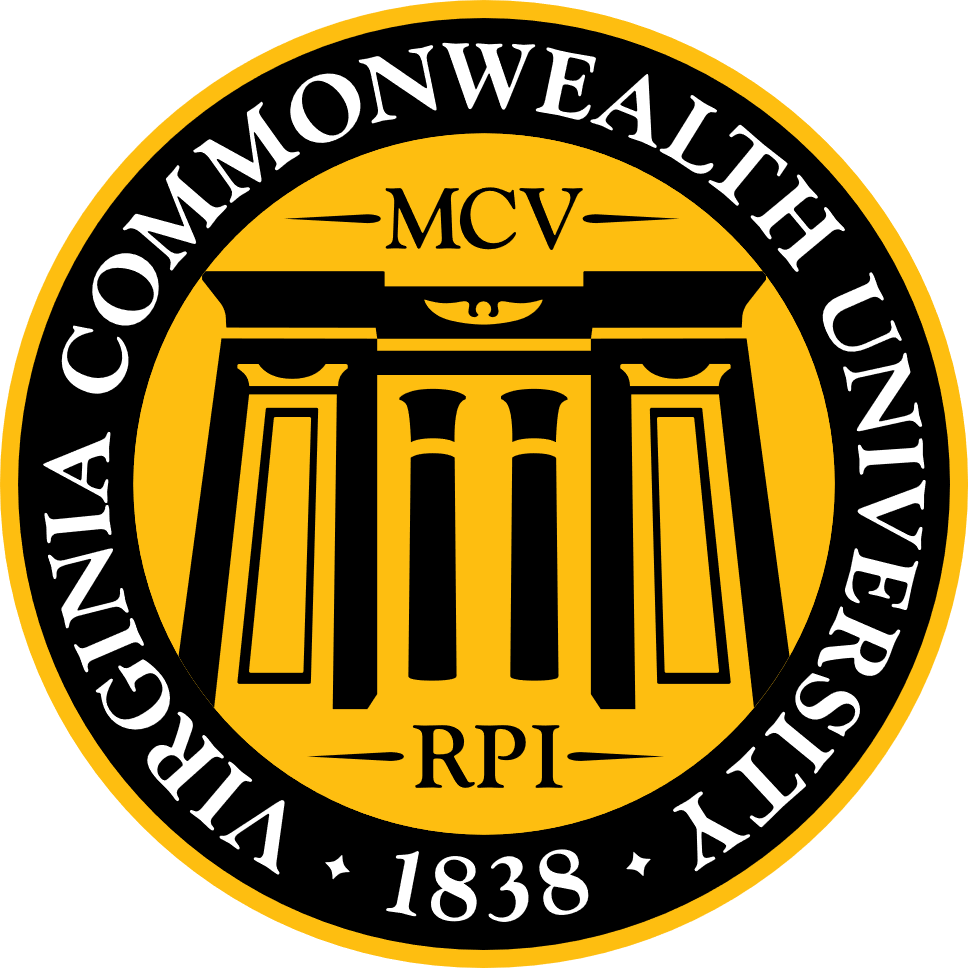
****

**VIRGINIA COMMONWEALTH UNIVERSITY**

**Statistical analysis and modelling (SCMA 632)**

**A2: Conducting a logistic regression, Confusion Matrix, ROC Curve and discussing Probit and Tobit regression**

**Arvind Srinivasan**

**V01107251**

**Date of Submission: 01-07-2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2.** | Results | **2** |
| **3.** | Interpretations | **3** |
| **4.** | Recommendations | **4** |

**Introduction**

The dataset used in this analysis comprises car evaluations with seven categorical features: buying price, maintenance cost, number of doors, number of persons, luggage boot capacity, safety, and an overall decision rating. These features provide a comprehensive view of various car attributes that influence consumer acceptability. Understanding the factors that contribute to a car's acceptability is crucial for manufacturers and marketers. By analyzing this data, we can identify the key attributes that drive consumer preferences, enabling targeted improvements in car design, feature enhancement, and more effective marketing strategies. The goal of this analysis is to build predictive models to classify car evaluations into acceptable and unacceptable categories, leveraging logistic regression with Lasso regularization and decision tree models. These models will help in understanding the relationships between different car features and their overall acceptability, providing actionable insights for business decisions.

**Objectives**

1. **Predict Car Acceptability**:
   * Develop models to classify car evaluations into acceptable and unacceptable categories.
2. **Identify Key Influential Features**:
   * Determine the most significant factors affecting car acceptability.
3. **Enhance Decision-Making**:
   * Use model insights to inform marketing, product development, and inventory management.

**Results**

**Logistic Regression with Lasso Regularization**:

* **Confusion Matrix**:

|  |  |  |
| --- | --- | --- |
|  | **Predicted: 0** | **Predicted: 1** |
| **Actual: 0** | 358 | 12 |
| **Actual: 1** | 9 | 139 |

* **AUC**: 0.9922
* **ROC Curve Analysis**:
  + The ROC curve for the logistic regression model indicates an excellent discriminative ability with an AUC of 0.9922. The curve is significantly close to the top-left corner of the plot, demonstrating the model’s high accuracy in distinguishing between acceptable and unacceptable car evaluations.
  + The model shows very few false positives (12) and false negatives (9), which indicates that it makes accurate predictions for both classes (acceptable and unacceptable).

**Decision Tree Model**:

* **Confusion Matrix**:

|  |  |  |
| --- | --- | --- |
|  | **Predicted: 0** | **Predicted: 1** |
| **Actual: 0** | 343 | 27 |
| **Actual: 1** | 2 | 146 |

* **AUC**: 0.967
* **ROC Curve Analysis**:
  + The ROC curve for the decision tree model shows good performance with an AUC of 0.967. While slightly lower than the logistic regression model, it still indicates a strong ability to correctly classify car evaluations.
  + The decision tree model has a higher number of false positives (27) compared to the logistic regression model but a lower number of false negatives (2), indicating that it is more conservative in predicting the acceptable category.

**Graphical Analysis**:

* **ROC Curves**: The ROC curves for both models illustrate the trade-off between the true positive rate and the false positive rate. The logistic regression model's curve being closer to the top-left corner suggests better overall performance compared to the decision tree model.
* **Confusion Matrices**: The matrices provide detailed insights into the models' prediction accuracies, revealing the number of true positives, true negatives, false positives, and false negatives. These metrics are essential for understanding how well each model is performing.

**Business Implications**

1. **Enhanced Marketing Strategies**: By identifying key features that influence car acceptability, businesses can create targeted marketing campaigns that emphasize these attributes (e.g., safety, maintenance cost), leading to more effective outreach and higher sales conversions.
2. **Data-Driven Product Development**: Insights from the models can guide manufacturers to focus on enhancing features that are highly valued by consumers, such as safety and luggage boot capacity, ensuring new products meet market demands and increase their competitive edge.
3. **Optimized Inventory Management**: Predictive models help in forecasting the acceptability of car models, allowing businesses to prioritize production and stocking of popular models, reducing overproduction, storage costs, and unsold inventory, thereby optimizing resource use and minimizing financial risks.

**Interpretations**

1. **Model Performance**:
   * The logistic regression model with Lasso regularization shows a high AUC of 0.9922, indicating excellent accuracy in predicting car acceptability. The decision tree model also performs well with an AUC of 0.967, though slightly less accurate than the logistic regression model.
2. **Feature Importance**:
   * The significant performance of the models suggests that attributes such as safety, maintenance cost, and luggage boot capacity play crucial roles in determining the acceptability of a car. These features are pivotal in influencing consumer preferences.
3. **Prediction Accuracy**:
   * Both models exhibit high true positive and true negative rates, with the logistic regression model demonstrating fewer false positives and false negatives compared to the decision tree. This indicates the logistic regression model's superior ability to accurately classify car evaluations.

**Recommendations**

1. **Marketing Strategies**:
   * Utilize insights from the logistic regression model to craft targeted marketing campaigns highlighting key features like safety and maintenance cost, which significantly influence car acceptability.
2. **Product Development**:
   * Focus on enhancing the features identified as most influential by the models, such as improving safety measures and luggage boot capacity, to align new car models with consumer preferences and increase market acceptance.
3. **Inventory Management**:
   * Leverage predictive models to forecast the acceptability of car models, ensuring efficient production and stocking of popular models. This will reduce overproduction, lower storage costs, and minimize the risk of unsold inventory.