**Report on Car Evaluation Prediction**

**Title**: Car Evaluation Prediction Using Machine Learning Techniques

**Student Name**: Priyankesh

**Enrollment Number:** 11519011622

**Email ID:** priyankeshom@gmail.com

**Contact Number**: 9599242344

**Google Drive Link:** https://drive.google.com/drive/folders/1BiE3mZGLtY6dCxL5g9mV4aFb8x\_zlpG?usp=sharing

**Abstract:**

This report describes a machine learning project to predict car acceptability based on features like price, maintenance cost, number of doors, capacity, luggage boot size, and safety rating. The data used for this project is the car\_evaluation.csv dataset. The project explores various classification algorithms including Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors Classifier, and Logistic Regression. The report discusses the performance of each model and identifies the model with the best accuracy.

Keywords: Car evaluation, Machine learning, Classification, Decision Tree, Random Forest, KNN, Logistic Regression

**1. Introduction**

This project aims to develop a machine learning model to predict car acceptability (unacceptable, acceptable) based on various features of the car. The model can be used to help potential car buyers in their decision-making process.

**2. Proposed Methodology**

The proposed methodology involves the following steps:

Data Acquisition: The car\_evaluation.csv dataset was used for this project.

Data Preprocessing: The data was preprocessed to handle missing values and convert categorical variables into numerical ones using ordinal encoding.

Exploratory Data Analysis (EDA): This step involved visualizing the data to understand the distribution of features and identify any potential relationships between features.

Feature Engineering: No additional feature engineering was performed in this project.

Model Building: Several classification models were trained and evaluated, including:

\* Decision Tree Classifier

\* Random Forest Classifier

\* K-Nearest Neighbors Classifier

\* Logistic Regression

Model Evaluation: The performance of each model was evaluated using accuracy scores and confusion matrices.

Comparison: The performance of all the models was compared to identify the model with the best accuracy.

**2.1 Explanation of Each Sub Module**

Data Loading: This module loads the car\_evaluation.csv dataset.

Data Preprocessing: This module handles missing values and encodes categorical variables.

Exploratory Data Analysis (EDA): This module visualizes the data to understand its distribution.

Feature Engineering: This module (optional) can be used to create new features from existing ones.

Model Building: This module trains the different classification models.

Model Evaluation: This module evaluates the performance of each model using accuracy scores and confusion matrices.

Comparison: This module compares the performance of all the models.

**3. Results and Discussion**

3.1 Data Exploration

The initial exploration of the data provides insights into the distribution of features and helps identify potential relationships between them. This can be achieved through techniques like creating histograms, boxplots, and scatter plots.

3.2 Feature Engineering

In this project, no additional feature engineering was performed. However, depending on the dataset, feature engineering techniques can be used to create new features that might improve model performance.

3.3 Model Building

Several classification models were trained on the preprocessed data. The models were:

Decision Tree Classifier: A decision tree model was trained with a maximum depth of 3.

Random Forest Classifier: A random forest model was trained with default parameters.

K-Nearest Neighbors Classifier: A KNN model was trained with 5 neighbors.

Logistic Regression: A logistic regression model was trained.

3.4 Model Evaluation

The performance of each model was evaluated using training and testing accuracy scores and confusion matrices. Confusion matrices provide a visualization of how well the model classified the data points.

3.5 Comparison

The models were compared based on their testing accuracy scores. The model with the highest testing accuracy is considered the best performing model.

**4. Conclusion and Future Work**

This project successfully explored various machine learning classification algorithms for car evaluation prediction. The Random Forest achieved the highest testing accuracy of 1.0 among the models explored in this project.

Future work could involve:

\* Fine-tuning the hyperparameters of the models to potentially improve performance.

\* Exploring the use of feature engineering techniques to create new features.

\* Using more advanced classification algorithms like Support Vector Machines (SVMs) or deep learning models.

\* Implementing the model into a web application or mobile app for real-world use.