Report on Machine Learning Classification for Laptop Type Prediction

1. Introduction

This report presents the results of a machine learning project aimed at predicting laptop types based on key hardware features. The project involved building and evaluating two classification models: Random Forest and Decision Tree classifiers. The focus was on determining which model performs better based on evaluation metrics such as Accuracy, Precision, Recall, and F1-Score.

The hardware features used for the prediction task were:

- Inches : Screen size in inches.
- RAM : The amount of random-access memory in GB.
- PrimaryStorage : Storage capacity in GB.
- Weight : Weight of the laptop in kilograms.

2. Methodology

2.1 Data Preprocessing

The data consisted of various laptop specifications. The following steps were taken to prepare the data for model training:

Feature Selection : The key features used were `Inches`, `RAM`, `PrimaryStorage`, and `Weight`.

Data Splitting : The data was split into a training set (to train the model) and a testing set (to evaluate model performance).

2.2 Modeling Approaches

Two machine learning classification models were built:

1. Random Forest Classifier

Model Description : The Random Forest classifier is an ensemble method that builds multiple decision trees and aggregates their predictions to improve accuracy. It reduces overfitting and improves generalization by averaging the predictions of individual trees.

Training : The Random Forest was trained on the training data, and feature importance was calculated to understand which features contributed the most to the prediction task.

2. Decision Tree Classifier

Model Description : The Decision Tree classifier creates a tree-like structure where the data is split into branches based on feature values. Each internal node represents a decision based on a feature, and each leaf node represents a predicted class.

Training : The Decision Tree classifier was trained using the same data. Similar to the Random Forest, feature importance was computed to evaluate the contribution of each feature.

2.3 Evaluation Metrics

To evaluate the performance of each model, the following metrics were used:

- Accuracy : The percentage of correctly classified instances out of the total instances.
- Precision : The ratio of true positives to the sum of true positives and false positives (how many selected items are relevant).
- Recall : The ratio of true positives to the sum of true positives and false negatives (how many relevant items are selected).
- F1-Score : The harmonic mean of Precision and Recall, balancing both metrics.

3. Results

3.1 Feature Importance

The importance of each feature in predicting the laptop type was analyzed for both models:

1. Random Forest Classifier :

- Weight : 53.44%
- RAM : 18.28%
- Inches : 17.82%
- PrimaryStorage : 10.45%

2. Decision Tree Classifier :

- Weight : 63.47%
- RAM : 16.29%
- Inches : 10.98%
- PrimaryStorage : 9.26%

Across both models, Weight was consistently the most important feature, followed by RAM and Inches.

3.2 Model Performance Evaluation

The performance of both models was evaluated using the test set, and the following results were obtained:

1. Random Forest Classifier :

- Accuracy : 0.74 (74%)
- Precision : 0.75
- Recall : 0.76
- F1-Score : 0.75

2. Decision Tree Classifier :

- Accuracy : 0.76 (76%)
- Precision : 0.74
- Recall : 0.76
- F1-Score : 0.75

3.3 Model Comparison

Both models showed similar performance across all metrics. However, the Decision Tree Classifier outperformed the Random Forest Classifier slightly in terms of accuracy.

The Decision Tree Classifier had an accuracy of 76%, whereas the Random Forest Classifier achieved an accuracy of 74%. Both models had comparable Precision, Recall, and F1-Scores, but the Decision Tree model was marginally better.

The Decision Tree model also had a higher feature importance for the Weight attribute, which was consistent with the model's slightly better performance.

4. Conclusion

From the results, the Decision Tree Classifier emerged as the better model for predicting laptop types, achieving higher accuracy compared to the Random Forest Classifier. While both models demonstrated good performance, the simplicity and interpretability of the Decision Tree make it an attractive option for this task. The analysis also revealed that Weight and RAM were the most important features in predicting the laptop type.

4.1 Recommendations

Future improvements could include:

Exploring other machine learning models, such as Support Vector Machines or Gradient Boosting Machines, to further improve the prediction accuracy.

Incorporating additional hardware features (e.g., processor type, GPU) to increase the predictive power of the models.

Using hyperparameter tuning techniques (such as grid search or randomized search) to optimize model performance.

4.2 Future Work

Further experiments could be conducted to test the model on a larger dataset or with more diverse types of laptops. Additionally, feature engineering techniques could be applied to transform the input data and extract more meaningful patterns, leading to potentially better classification performance.