Patterns Recognition

Project

Faculty of Computers and Artificial Intelligence

Cairo University

Under Supervision of Dr: Mona Souleiman

Mobile Price Classification report

Prepared for:

Dr: Mona Souleiman

Prepared by:

Islam Abd-Alglill Mahmoud

20th May

Fall 2023

Mobile Price Classification

Table of contents:

1-Audience……...4

2-Abstract……4

3-Objective……..….4

4-Dataset…….…….5

5-Classifier …….…….....5

5.1-Introduction…….…….…….5

5.2-How Decision Trees Work……..…….…….5

5.3-Advantages of Decision Trees……...……. ….6

5.4-Disadvantages of Decision Trees…….…….6

5.5-Examples of Decision Tree Use……..……..7

6-Result…….……7

7-Conclusion…….……..7

Audience:

1. Researchers: Those who are interested in the latest developments and advancements in Pattern Recognition research.

2. Engineers: Those who design and develop systems that use Pattern Recognition technologies.

3. Data Scientists: Those who work with large datasets and use Pattern Recognition techniques to extract meaningful insights.

4. Academics: Those who teach Pattern Recognition or related subjects at universities or other educational institutions.

5. General Public: Those who are interested in how Pattern Recognition is being used in various fields and how it may impact their lives.

Abstract:

This project focuses on the classification of mobile phone prices based on given features. The aim is to accurately predict the class of a mobile phone from a given set of input features. The dataset consists of various features such as battery power, RAM, screen size, and other relevant specifications. The dataset is divided into four classes representing different price ranges. The project employs machine learning algorithms to train a model that can effectively classify a mobile phone into one of the four price classes.

Objective:

To develop and implement a machine learning model that accurately classifies mobile phones into one of four price categories based on a set of input features. The model should achieve a high level of accuracy while also being scalable and efficient enough to handle large amounts of data in real-time.

Dataset:

The dataset is made of a feature vector of 20 features like (battery power, ram, clock speed and Wi-Fi etc..) so no need to do any feature extraction technique, the length of the dataset is 2000 rows, and there are 4 price range classes the program should classify each phone price range We split the dataset into two categories:

- train category we used 1600 samples to train our program.
- test category we used 400 samples to test our program.

Decision tree classifier:

Introduction:

The decision tree algorithm is a popular machine learning technique that has been used in many applications for both classification and regression tasks. The algorithm works by recursively partitioning the data into subsets based on the values of the input features until a stopping criterion is met. This report will explain how the decision tree algorithm works, including its advantages and disadvantages, and some examples of its use.

How Decision Trees Work:

The decision tree algorithm starts with all the training data at the root node of the tree. For each node in the tree, the algorithm selects the feature that best separates the data into different classes or groups. This is typically done using a metric such as Gini index, entropy, or information gain. The goal is to select the feature that maximizes the separation between the different classes while minimizing the variation within each class.

Once the feature has been selected, the algorithm splits the data into two or more subsets based on the selected feature. Each subset will correspond to a new child node in the tree. The process of selecting the best feature and splitting the data is repeated for each child node until a stopping criterion is met. This could be a maximum tree depth, a minimum number of samples per leaf node, or some other criteria.

When the decision tree has been built using the above steps, it can be used for making predictions on new data by traversing the tree from the root node down to a leaf node based on

the values of the input features. At each node, the algorithm evaluates the value of the selected feature and takes the appropriate path down the tree until it reaches a leaf node. The final prediction is made based on the majority class or average value of the samples in the leaf node.

Advantages of Decision Trees:

Decision trees have several advantages over other machine learning algorithms, including:

1. Easy to understand: Decision trees are easy to interpret and visualize, making them ideal for explaining the reasoning behind the predictions.

2. Fast and efficient: Decision trees can handle large datasets with many features and are able to make predictions quickly.

3. Can handle non-linear relationships: Decision trees can capture non-linear relationships between the input variables and the output variable.

4. Robust to noise: Decision trees are resistant to noise and can handle missing data without affecting the overall performance of the algorithm.

Disadvantages of Decision Trees:

Despite their advantages, decision trees have some limitations, such as:

1. Overfitting: Decision trees have a tendency to overfit the training data, which can lead to poor generalization on new data.

2. Instability: Small changes in the training data can lead to significant changes in the structure of the decision tree, making them unstable.

3. Bias: Decision trees tend to favor features with many possible values or high cardinality, leading to bias towards these features.

Examples of Decision Tree Use:

Decision trees have been used in many applications, including:

1. Healthcare: Decision trees have been used to predict the risk of heart disease based on patient demographic data and medical history.

2. Finance: Decision trees have been used to predict stock prices based on financial market data.

3. Marketing: Decision trees have been used to predict customer behavior based on historical purchase data.

Result:

After training 1600 sample of the dataset and testing 400 sample on Decision tree algorithm with hyper parameter (maximum depth=7 and minimum samples split=3) the predicted accuracy result for price range for testing samples was 81.5%

Conclusion:

The decision tree algorithm is a popular machine learning technique that can be used for classification and regression tasks. The algorithm works by recursively partitioning the data into subsets based on the values of the input features until a stopping criterion is met. Decision trees have several advantages, including ease of interpretation and fast performance, but also have some limitations, such as instability and bias. Despite these limitations, decision trees have been successfully used in many applications, including healthcare, finance, and marketing.