Nitte Meenakshi Institute of Technology,

Department of Computer Science and Engineering

18CSE751 Introduction to Machine

Learning Learning Activity Proposal

Obesity Predictions

Agil Srinivasan(1NT19CS400), Guruprasad k(1NT19CS407)

Course Cordinator:- Dr. Vani V

Abstract

The decision tree is one of the most widely used machine learning algorithms due to its ease of interpretation. A decision tree consists of creating different rules by which we make the prediction. For example, let's say we train an algorithm that predicts whether or not a person is obese based on their height and weight.

Introduction

In modern times, obesity has become a significant threat all over the world. Obesity means an unnatural or excessive amount of fat that is present in our bodies. People are constantly moving towards an unhealthy lifestyle, eating excessive junk food, late-night sleep, spend a long time sitting down. The purpose of this paper is to move towards a machine-learning-based pathway for predicting the obesity using machine-learning algorithms. The great thing about this paper is that people will know their obesity. We collect more than 500 data from both obesity and non-obesity. For this research, we apply Decision tree machine learning algorithms.

Data Set

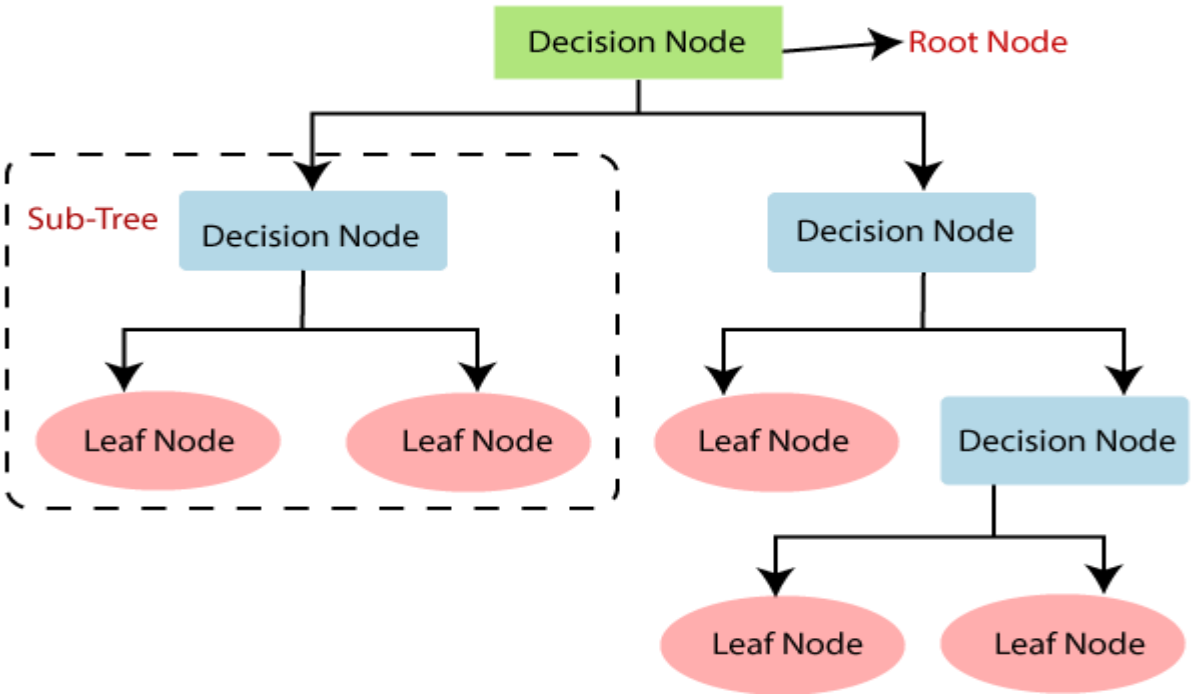
We will use the dataset from the Kaggle. The dataset is obtained from the following link – https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex
The dataset contain 500 entries of unique values.

Machine Learning Methods

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

As you can see, decision trees usually have sub-trees that serve to fine-tune the prediction of the previous node. This is so until we get to a node that does not split. This last node is known as a leaf node or leaf node. Let's see a graphic example:



Besides,a decision trees can work for both regression problems and for classification problems. In fact, we will code a decision tree from scratch that can do both.

Assessment:

1)Calculate impurity using the Gini index

2)Calculate impurity with entropy

Presentation and Visualization:-
Will be using graphs to show the presentation.

Schedule
The schedule is a table of dates and tasks that you plan to complete.

Date	Tasks to be Completed
17/01/21	Tasks completed by chosen date
18/01/22	Tasks to be completed by the final report/ presentation date

Bibliography

- https://www.researchgate.net/publication/331611719_Decision_tree_learning_to_predict_overweightobesity_based_on_body_mass_index_and_gene_polymorphisms
- https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex
- https://anderfernandez.com/en/blog/code-decision-tree-python-from-scratch/
- https://www.sciencedirect.com/science/article/pii/S2666518221000401