Machine Learning Model

Classification

1. Logistic Regression (Recommand)

The logistic regression model will generate a probability output for a label. And then it can set a threshold by comparing the probability output (Through the sigmoid function).

A picture containing text, diagram, screenshot, line

Description automatically generated

Figure 1. The principle for logistic regression.

This is the link I find for simple logistic regression in python: <https://towardsdatascience.com/simple-logistic-regression-using-python-scikit-learn-86bf984f61f1>

1. Decision Tree Model

It uses if-else statement for each branch. Each branch classifies the dataset into subsets based on the most significant features.

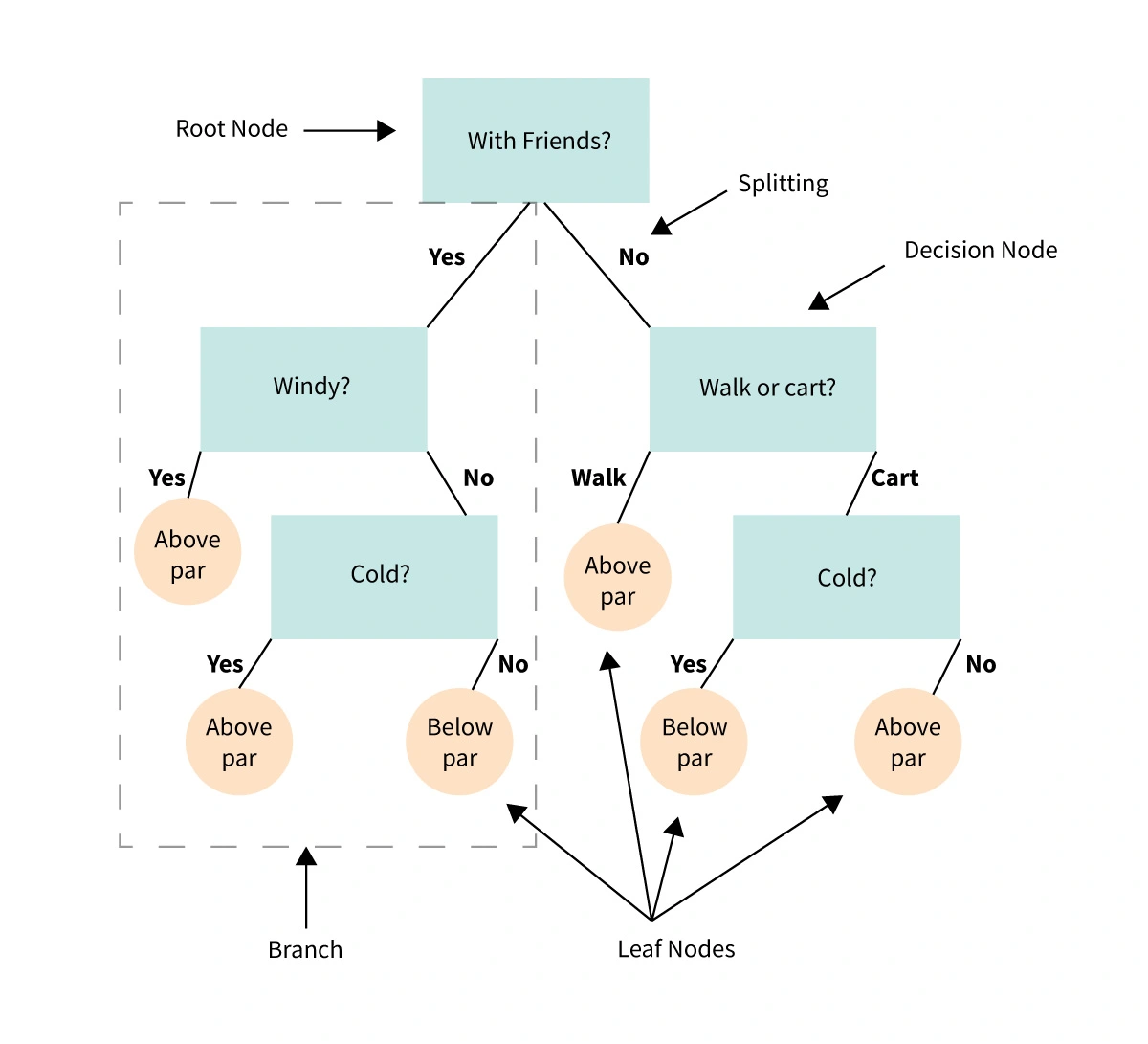


Figure 2. The principle for decision tree model.

1. Random Forest

This is a collection of the decision tree, it allows each tree trained on a random sampling of original dataset and takes majority vote from trees.

Compared with decision tree model, it has better generalization but less interpretable (because of more layers).



Figure 3. The principle for Random Forest model.

1. Support Vector Machine (Recommended)

The best way to classify the databased on the position in relation to a board between positive class and negative class. The board is the hyperplane which maximize the distance between data points from different classes.

This is the link I find the resource for SVM:

<https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>



Figure 3. The principle for support vector machine model.

1. K-Nearest Neighbour (KNN) Quiet complex (Not recommend to use)

It calculates the distance between on point to another, and then assign the label of unobserved data based on the labels of nearest observed data points. It is used for building recommendation system.

A picture containing line, diagram, circle

Description automatically generated

Figure 4. The principle KNN model.

1. Naïve Bayes (This may be used when the data become larger)

It can calculate the condition probability based on prior knowledge, each feature is independent with each other. This is suit for the large amount of data.

A picture containing line, diagram, plot, design

Description automatically generated

Figure 6. The principle naïve bayes model.