Final Assignment Report : Decision Tree Algorithm and K-Nearest Neighbour Algorithm

CSE-0408 Summer 2021

Ahasanul Bari Rifat
Department of Computer Science and Engineering
State University of Bangladesh (SUB)
Dhaka, Bangladesh
ahsanulbari.rifat88@gmail.com

***Abstract*—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.**

An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor

***Index Terms*—Python**

I. Introduction

Decision tree are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

KNN also known as K-nearest neighbour is a supervised and pattern classification learning algorithm which helps us find which class the new input(test value) belongs to when k nearest neighbours are chosen and distance is calculated between them.

II. Why use Decision Trees?

Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.

- The logic behind the decision tree can be easily understood because it shows a tree-like structure.

III. Why or how use K-Nearest Neighbour Algorithm?

KNN algorithm can be used for both classification and regression problems. The KNN algorithm uses 'feature similarity' to predict the values of any new data points.

IV. Example of Decision Tree and knn Algorithm

An example of a decision tree can be explained using above binary tree.

KNN also known as K-nearest neighbour is a supervised and pattern classification learning algorithm which helps us find which class the new input(test value) belongs to when k nearest neighbours are chosen and distance is calculated between them.

V. Advantages and Disadvantages of Decision Tree Algorithm

Advantages :

-It is easy to grasp because it follows a constant method that somebody follows whereas creating any call-in real-life.

-It is terribly helpful for the resolution of decision-related issues.

Disadvantages :

-It may have an associate overfitting issue, which might be resolved exploitation the Random Forest formula.

-For a lot of category labels, the process quality of the choice tree could increase.

VI. Advantages and Disadvantages of KNN Algorithm

Advantages :

-New data can be added without effecting the algorithm performance or accuracy.

-Versatile – useful for regression and classification.

-High accuracy – you do not need to compare with better-supervised learning models.

Disadvantages :

-ccuracy depends on the quality of the data.

-With large data, the prediction stage might be slow.

-Sensitive to the scale of the data and irrelevant features.

-Require high memory – need to store all of the training data

VII. Report Code of Decision Tree Algorithm

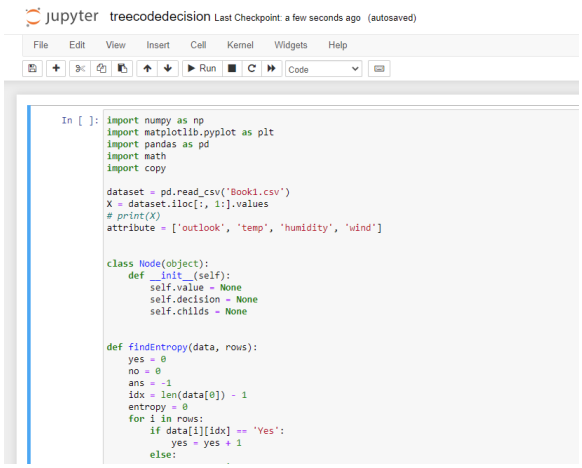

Fig. 1.

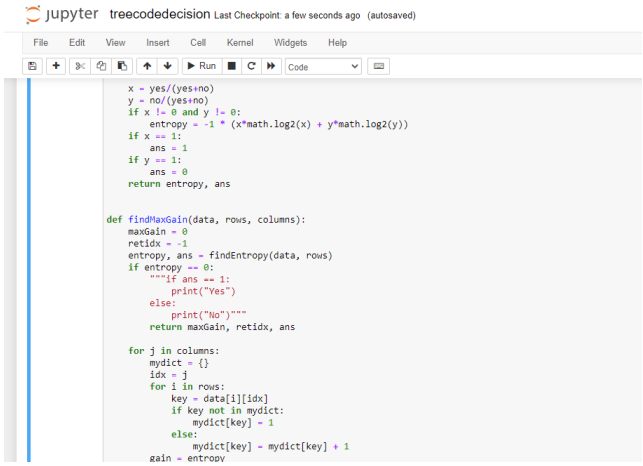

Fig. 2.

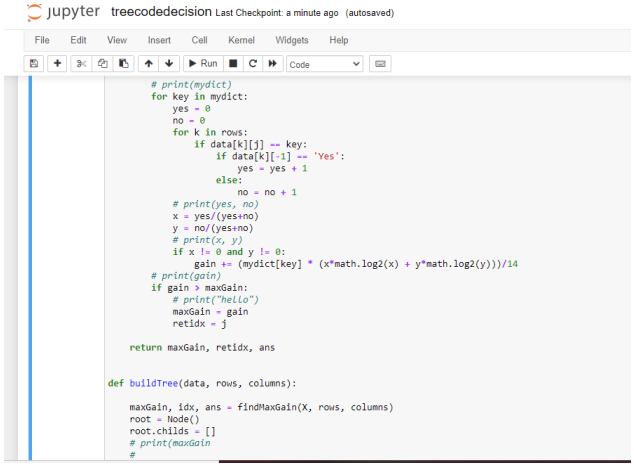

Fig. 3.

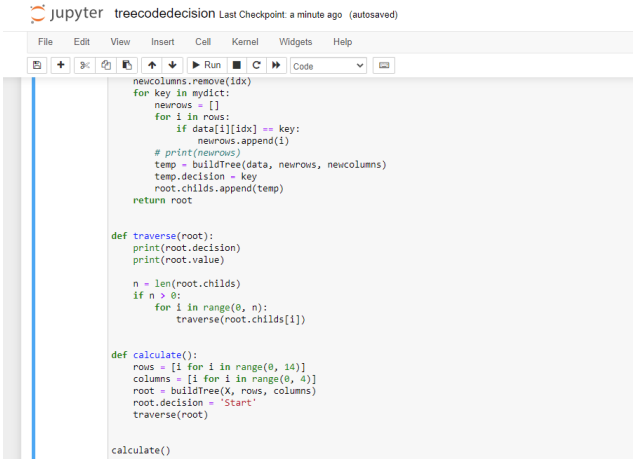

Fig. 4.

VIII. Report Code of KNN Algorithm Algorithm

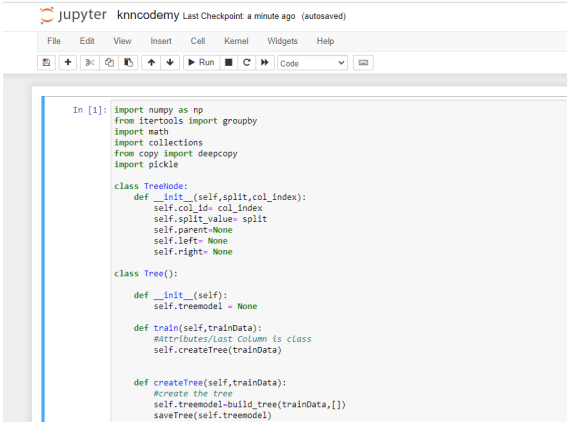

Fig. 5.

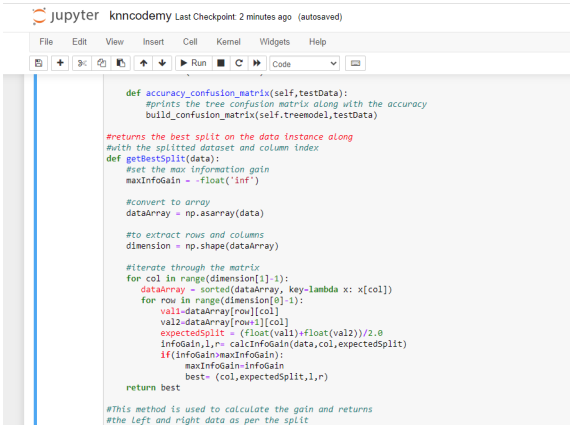

Fig. 6.

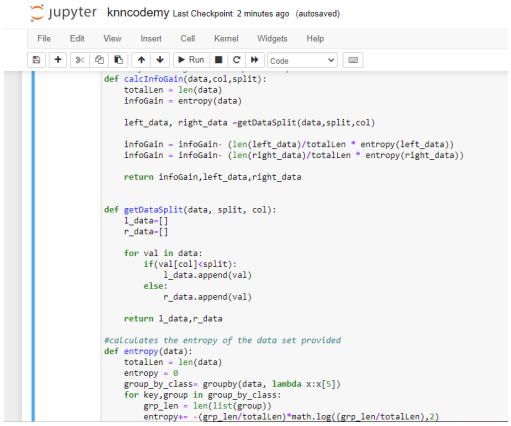

Fig. 7.

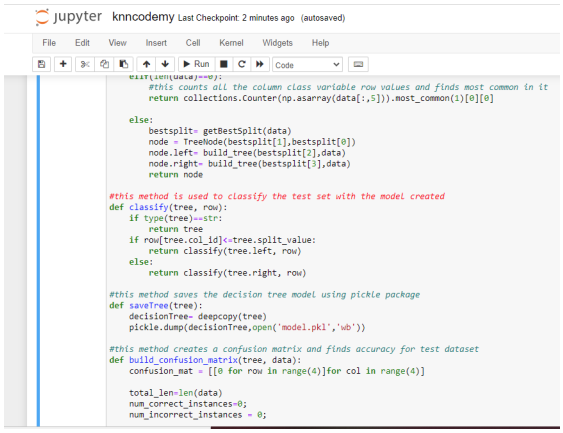

Fig. 8.

IX. Assignment Output (Decision Tree Algorithm Problem Solving)

Fig. 9.

X. Assignment Output (KNN algorithm Output)

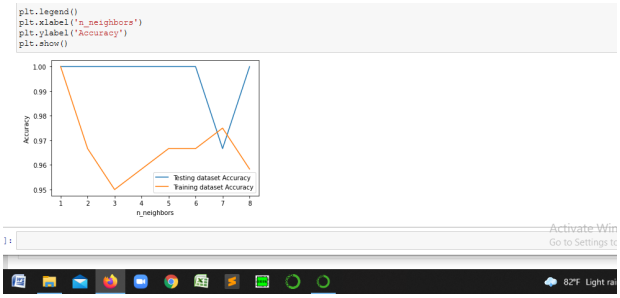

Fig. 10.

Acknowledgment

I would like to thank my honourable **Khan Md. Hasib Sir** for his time, generosity and critical insights into this project.

References

[1] Ben-Haim, Yael, and Elad Tom-Tov. "A Streaming Parallel Decision Tree Algorithm." Journal of Machine Learning Research 11.2 (2010).

[2] Freund, Yoav, and Llew Mason. "The alternating decision tree learning algorithm." icml. Vol. 99. 1999.

[3] Su, Jiang, and Harry Zhang. "A fast decision tree learning algorithm." Aaai. Vol. 6. 2006.

[4] Priyam, Anuja, et al. "Comparative analysis of decision tree classification algorithms." International Journal of current engineering and technology 3.2 (2013): 334-337.

[5] Soucy, P., Mineau, G. W. (2001, November). A simple KNN algorithm for text categorization. In Proceedings 2001 IEEE International Conference on Data Mining (pp. 647-648). IEEE.

[6] Deng, Z., Zhu, X., Cheng, D., Zong, M., Zhang, S. (2016). Efficient kNN classification algorithm for big data. Neurocomputing, 195, 143-148.

[7] Cheng, D., Zhang, S., Deng, Z., Zhu, Y., Zong, M. (2014, December). kNN algorithm with data-driven k value. In International Conference on Advanced Data Mining and Applications (pp. 499-512). Springer, Cham.