# Final Report : Solving Decision Tree Algorithm and Solving K-Nearest Neighbour Algorithm

CSE-0408 Summer 2021

Khondkar Md Mufrat Tasif
*Department of Computer Science and Engineering*
*State University of Bangladesh (SUB)*
Dhaka, Bangladesh
kmmufrat08@gmail.com

***Abstract*—Code 1 :Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too**

**Code 2 :The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.**

***Index Terms*—Python, C++**

## I. Introduction

Decision Tree algorithm belongs to the family of supervised learning algorithms. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data

KNN is a lazy learning, non-parametric algorithm. It uses data with several classes to predict the classification of the new sample point.

## II. Why KNN algorithm and Decision tree algorithm is used?

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. Decision trees can handle both categorical and numerical data.

The KNN algorithm can compete with the most accurate models because it makes highly accurate predictions. The quality of the predictions depends on the distance measure.

## III. Types of Decision Tree and knn Algorithm

There are 4 popular types of decision tree algorithms:
1.ID3,
2. CART (Classification and Regression Trees),
3. Chi-Square
4.Reduction in Variance.

The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems.

## IV. Rules for Solving Decision Tree Algorithm

Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

## V. Calculated a KNN algorithm

Here is step by step on how to compute K-nearest neighbors KNN algorithm:

1. Determine parameter K = number of nearest neighbors.
2. Calculate the distance between the query-instance and all the training samples.

## VI. Report Code of Decision Tree Algorithm

```
import pandas as pd
import numpy as np

df = pd.read_csv("datadt.csv")

df

X = df.iloc[:,:-1]

X

y = df.iloc[:,3]

y

from sklearn.preprocessing import LabelEncoder

LabelEncoder_X = LabelEncoder()

X = X.apply(LabelEncoder().fit_transform)

X

from sklearn.tree import DecisionTreeClassifier
```

Fig. 1.

```
from sklearn.tree import DecisionTreeClassifier

regressor = DecisionTreeClassifier()

regressor.fit(X.iloc[:,1:4], y)

X_in = np.array([1,0,])

y_pred = regressor.predict([X_in])

y_pred
```

Fig. 2.

## VII. Report Code of KNN Algorithm Algorithm

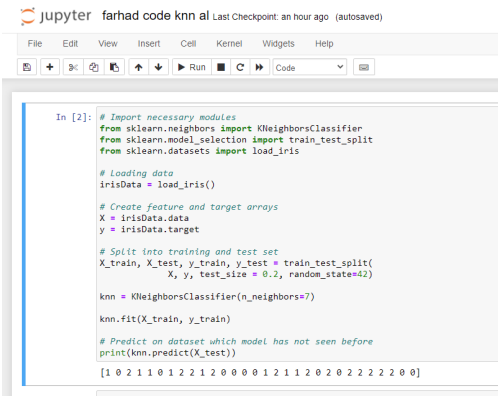


Fig. 3.

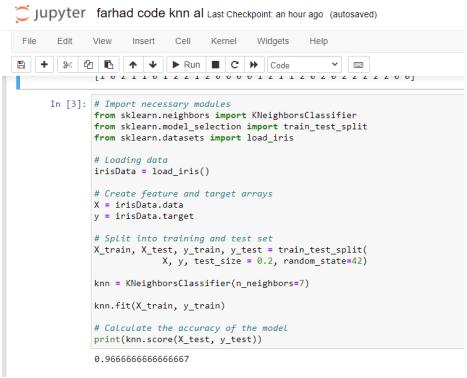


Fig. 4.

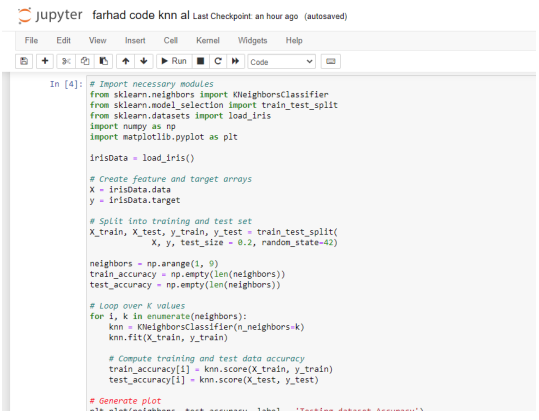


Fig. 5.

## VIII. Assignment Output ( Decision Tree Algorithm Problem Solving)

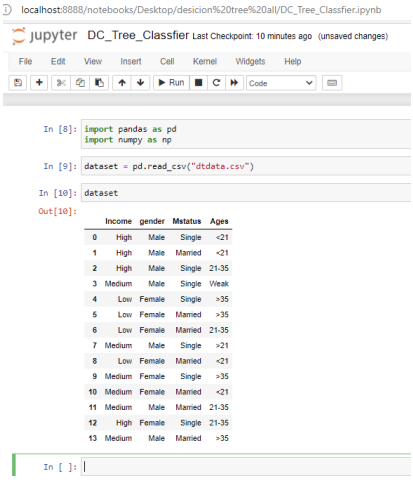


Fig. 6.

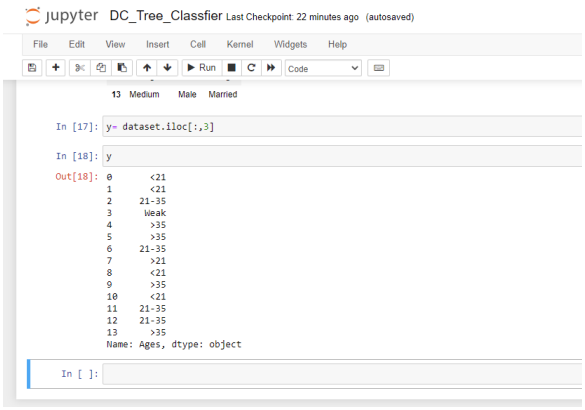


Fig. 7.

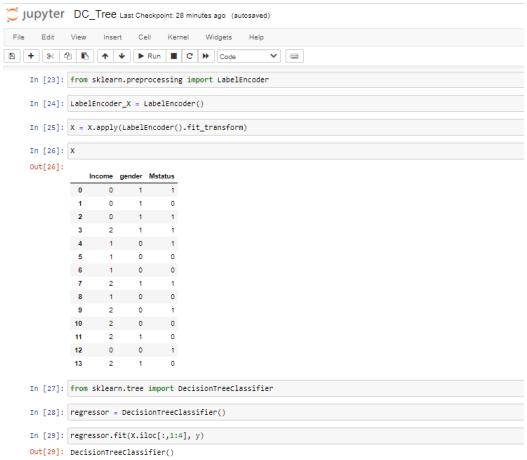


Fig. 8.

## IX. Assignment Output ( KNN algorithm Output)

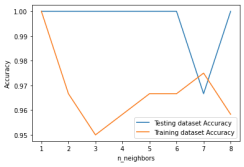


Fig. 9.


## 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.