

Artificial Intelligence

CSE-0408 Summer 2021

Md Mahthiul Hasan Thanvi
UG02-47-18-025
Department of Computer Science and Engineering
State University of Bangladesh (SUB)
Dhaka, Bangladesh
thanvimdmhatheulhasan@gmail.com

Abstract— The K-Nearest Neighbours method is a classification method that calculates a number to estimate the distance of two objects. We'll use Jupyter Notebook in this assignment to implement K-Nearest Neighbour.

Index Terms— Machine Learning, Supervised, Classification, Knearest neighbors.

I. INTRODUCTION

Over the course of my four years as a computer scientist, I've produced more than 80% classification models and only 15-20% regression models. These proportions can be applied across the industry in a broad sense. The rationale for this preference for classification models is that the bulk of analytical issues require making a judgment. In this assignment, we'll use K-nearest neighbors, which is a common machine learning classification technique (KNN). We'll focus on how the algorithm works and how the input parameter affects the output/prediction.

II. LITERATURE REVIEW

Piltaver, Rok, Mitja Luštrek, and Matjaž Gams. "The pathology of heuristic search in the 8-puzzle." *Journal of Experimental Theoretical Artificial Intelligence* 24.1 (2018): 65-94.

III. KNN ALGORITHM

The steps for building a KNN model are just as continues to follow:

- 1) Load the data
- 2) Initialise the value of k
- 3) For getting the predicted class, iterate from 1 to total number of training data points.
the distance between each row of training data and the test data. Because it is the most frequently used approach, we will utilize Euclidean distance as our distance metric. Chebyshev, cosine, and other metrics can also be applied. Calculate the distance between each row of training data and the test data. Because it's the most popular way, we'll adopt Euclidean distance as our distance metric. Chebyshev, cosine, and other metrics can also be

applied. Sort the estimated distances by distance values in order of increasing. Get the first k rows of a sorted order. Get the most probable class of these rows. Get the projected class.

Libraries Requirements

- *pandas*
- *sklearn*
- *matplotlib*

Pandas is used to accept input parameters, **sklearn** is used to construct and train our models, and **matplotlib** is used to graphically represent our K-Nearest Neighbors accuracy.

IV. CONCLUSION

This project is built on a graphic depiction of the quality of a KNN model. For the training and visualisation of this KNN model's accuracy, a data set is supplied.

ACKNOWLEDGMENT

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

REFERENCES

- [1] W.W. Johnson and W.E. Storey. "Notes on the '15'-Puzzle." *Amer. J. Math.* 2(1879), 397-404.
- [2] P.D.A. Schofeld, N.L. Collins, D. Michie (eds.), "Complete solution of the Eight-Puzzle" *Machine Intell.* 1, Amer. Elsevier, NY (1967), 125-133.
- [3] A. Reinefeld and T.A. Marsland. "Memory functions in iterated deepening search." *Univ. Hamburg, FB Informatik, Tech. Rep. FBIHH-M-198/91.*
- [4] Duane Merrill, Michael Garland, Andrew Grimshaw. "Scalable GPU Graph Traversal" *PPoPP'12*, February 25-29, 2012, New Orleans, Louisiana, USA, 2012 ACM 978-1-4503-1160-1/12/02
- [5] Korf, R. E., and Schultze, "Large-scale parallel breadth-first search.", *National Conference on Artificial Intelligence (AAAI)*, 1380-1385, 2005

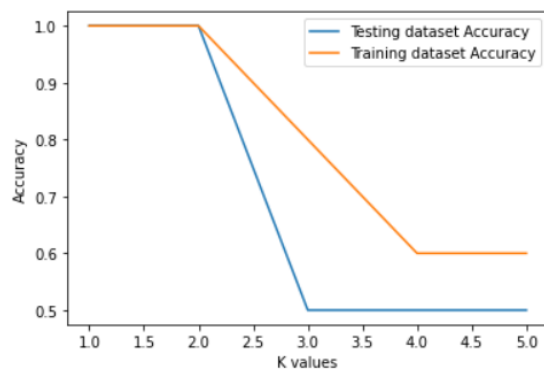


Fig. 2. Graph Accuracy

Abstract— One of the most well-known ways for modeling data classification in classifiers is decision tree classifiers. Researchers from several disciplines and backgrounds have investigated the problem of extending a decision tree using current data, including machine learning, pattern recognition, and statistics. Medical disease analysis, text categorization, user smartphone classification, images, and many more disciplines have suggested decision tree classifiers. In this Assignment, the decision trees are discussed in detail. Furthermore, the contents of the Assignment are thoroughly investigated and debated, including the algorithms/approaches used, data sets, and outcomes attained. Furthermore, all of the methodologies were examined and the best accurate classifiers were determined.

Index Terms— Artificial Intelligence, Machine Learning, Su-pervised, Classification, Decision Tree

V. INTRODUCTION

In recent years, technology has evolved substantially, particularly at the field of Machine Learning (ML), which has proven to be ineffectual in reducing human labor. Machine learning (ML) combines statistics and computer science to construct algorithms that become more efficient when given relevant data rather than specific instructions in the field of artificial intelligence. In addition to speech recognition, photo identification, and text localization, ML

involves the study of computing systems that are automatically enhanced by experience. It's considered to be a subset of artificial intelligence. In order to forecast or make judgements without being explicitly trained, ML algorithms construct a model population based on a sample, known as 'training data.' When constructing traditional algorithms to execute essential functionalities is difficult or prohibitive,

ML algorithms are employed in a wide range of applications, such as email filtering and computer vision. Machine learning has a variety of applications, the most well-known of which being predictive data mining. A

decision tree is a tree-based technique in which a data partitioning sequence is used to express any path from the root to the leaf node until a Boolean outcome is obtained. It's a node-and-connection-based hierarchical representation of knowledge relationships. When leveraging relations to classify data, nodes represent purposes. The goal of this project is to carry out a full implementation of the most recent and most productive decision tree techniques created

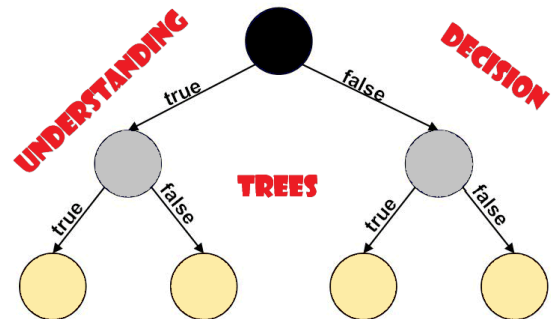
by researchers over the last three years in different domains of machine learning. A decision tree is also created as part of the process.

VI. LITERATURE REVIEW

Busato, Federico, and Nicola Bombieri. "BFS-4K: an efficient implementation of BFS for kepler GPU architectures." IEEE Transactions on Parallel and Distributed Systems 26.7 (2014): 1826-1838.

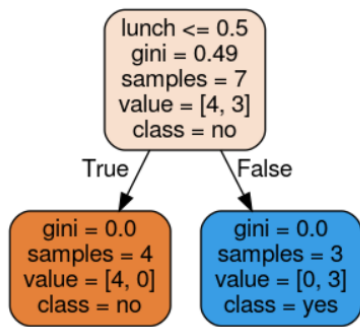
VII. DECISION TREE ALGORITHM

One of the most widely used data mining methodologies is the use of systems that generate classifiers. In data mining, classification algorithms are capable of managing massive amounts of data. It can be used to make categorical class name assumptions, categorize information using training sets and class labels, and classify newly accessible data. There are a multitude of algorithms used in machine learning classification approaches, however this study focuses on the decision tree algorithm in particular.



Decision Tree

Figure 1 represents the structure of DT. Decision trees are a powerful tool that may be applied to a wide range of fields, including machine learning, image processing, and pattern identification. DT is a sequential model that connects a series of fundamental tests by comparing a numeric property to a threshold value. The numerical weights in the neural network of node connections are significantly more complex to create than the conceptual rules. The most common application of DT is for grouping. DT is also a popular categorization model in Data Mining. The nodes and branches of each tree make it up. Each subset specifies a possible value for the node, whereas each node represents a feature in the category to be classified. Because of their ease of use and precision across a large range of data sources, decision trees have a wide range of applications. Figure 2 shows an example of DT.



Example on Decision Tree

Libraries Requirements

- *pandas*
- *sklearn*
- *IPython*
- *matplotlib*

Pandas is used to take input data sets, **sklearn** is used to develop and train our models, as well as **IPython**, **GraphViz** and **matplotlib** are used to visualize our decision trees graphically.

VIII. CONCLUSION

This project is built on a decision tree graphical display. For the training and visualisation of this decision tree, a data set is provided.

ACKNOWLEDGMENT

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

REFERENCES

- [1] A.V Aho, J.E. Hopcroft and J.B. Ullman, 77x Design and Analysis of Computer Algorithms (Aii=n-Wesley, Readin MA, 1974).
- [2] S. IB;ias;e, Computw Algorit4rrs: ir;troduction to Design and Analysis (Addison-Wesley, Reading, MA, 1983).
- [3] D. Coppersmith and S Winograd, Matrix multiplication via arithmetic progress;ons. Proc. 19th Ann. ACM S'ymp. on Theoqv of Cmwputing (May 1987) 1-6.
- [4] V. Pan, How to Mdfiply Matrices Emter (Springer, Berl 1984).
- [5] V. Strassen, Gaussian elimhtion is not optimal, Abnerische Mathematik 23 (1969) 354-3S6