Classification Based on Decision Tree Algorithm for

Machine Learning

**Abstract :**

Decision tree classifiers are regarded to be a standout of the most well-known methods to data classification representation of classifiers.

Different researchers from various fields and backgrounds have considered the problem of extending a decision tree from available

data, such as machine study, pattern recognition, and statistics.

In various fields such as medical disease analysis, text classification,

user smartphone classification, images, and many more the employment of Decision tree classifiers has been proposed in many ways.

This paper provides a detailed approach to the decision trees. Furthermore, paper specifics, such as algorithms/approaches used,

datasets, and outcomes achieved, are evaluated and outlined comprehensively. In addition, all of the approaches analyzed were

discussed to illustrate the themes of the authors and identify the most accurate classifiers. As a result, the uses of different types of

datasets are discussed and their findings are analyzed

**INTRODUCTION :**

Nowadays, technology has developed a lot, especially in

the field of Machine Learning (ML), which is useful for

reducing human work. In the field of artificial intelligence, ML

integrates statistics and computer science to build algorithms

that get more efficient when they are subject to relevant data

rather than being given specific instructions. Besides speech

recognition, image detection, text localization, etc. ML is the

study of computational algorithms that are enhanced from

experience automatically. It is considered as an artificial

intelligence subset [1, 2]. Orderly to produce foretelling or

decision without being specifically programmed to do so, ML

algorithms create a model population based on a sample,

defined as 'training data' [3, 4]. In a broad area of applications,

like email filtering and computer vision, ML algorithms are

utilized where it is difficult or impractical to create traditional

algorithms to implement functions required [5]. For ML, there

are many uses, the most prominent of which is predictive data

mining. Two major mechanisms can be broken into ML

classification fulfillments; model development and model

evaluation [6, 7].

DECISION TREE ALGORITHM :

One of the widely used techniques in data mining is

systems that create classifiers [32]. In data mining,

classification algorithms are capable of handling a vast volume

of information. It can be used to make assumptions regarding

categorical class names, to classify knowledge on the basis of

training sets and class labels, and to classify newly obtainable

data [33]. Classification algorithms in machine learning

contain several algorithms, and in this work, the paper focused

on the decision tree algorithm in general. Fig. 1 illustrate a

structure of DT.

**A. Types of Decision Tree Algorithms**

There are several Types of DT algorithms such as: Iterative

Dichotomies 3 (ID3), Successor of ID3 (C4.5), Classification

And Regression Tree(CART) , CHi-squared Automatic

Interaction Detector(CHAID) , Multivariate Adaptive

Regression Splines (MARS) Generalized, Unbiased,

Interaction Detection and Estimation (GUIDE), Conditional

Inference Trees (CTREE Classification Rule with

Unbiased Interaction Selection and Estimation (CRUISE),

Quick, Unbiased and Efficient Statistical Tree (QUEST)

. Table I shown comparison between the frequently used

algorithms for the decision tree .

**B. Entropy and Information Gain**

Entropy is employed to measure a dataset's impurity or

randomness . The value of entropy always lies

between 0 and 1. Its value is better when it is equal to 0 while it

is worse when it is equal to 0, i.e. the closer its value to 0 the

better. As shown in . If the target is with different

attribute values, the entropy of the classification of set with

respect to states .

Taloba and Ismail [77] developed a new machine learning

approach for the hybrid decision tree and a genetic algorithm

known as GADT for spam detection. The most significant

algorithm for enhancing decision tree efficiency is the genetic

algorithm. Also, it is efficient and reliable for text

classification. A genetic algorithm has used the element of trust

that governs decision tree pruning to optimize and detect its

optimum value. They used the UCI Machine Learning Store

spam dataset. Besides, they used the mechanism of main

Principle Component Analysis (PCA) to delete features that are

inappropriate for email message content and process them less

frequently. The findings showed that after using PCA, the

mixed GADT approach has an accuracy of 93.4 % before using

PCA and an accuracy of 95.5 %.

This implies that the extraction of inappropriate characteristics has a great impact on the PCA

**CONCLUSION :**

CONCLUSION

Decision tree classifiers are known for their enhanced view

of performance outcomes. Because of their strong precision,

optimized splitting parameters, and enhanced tree pruning

techniques (ID3, C4.5, CART, CHAID, and QUEST) are

commonly used by all recognized data classifiers. The separate

datasets are used for training samples from a huge data set,

which in tum, affects the precision of the test set. Decision

trees have several possible concerns about robustness, an

adaptation of scalability and optimization of height. But, in

contrast to other methods of data classification, decision trees

create an efficient rule collection that is simple to understand.

This paper reviews the most recent researches that are

conducted in many areas, such as analysis of medical diseases,

classification of texts, classification of user smartphones and

images, etc. Furthermore, the details used in the

techniques/algorithms, datasets were used by the authors and

achieved outcomes related to the accuracy are summarized for

decision trees. Finally, the best accuracy achieved for the

decision tree algorithm is 99.93% when it uses a machine

learning repository as a datase