



Chhatrapati Shahu Maharaj Shikshan Sanstha's

CHH. SHAHU COLLEGE OF ENGINEERING

Kanchanwadi, Paithan Road, Aurangabad.

INDEX

Date :

List of Experiments in :

S.N.	Title	Page No.	Date	Remarks
1.	Implement and demonstrate the find-s algo for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a csv file.			
2.	Implement and demonstrate the candidate elimination algorithm.			
3.	Implement & demonstrate the supervised machine learning algorithm decision trees.			
4.	Implementation of k-nearest neighbours (KNNs) on synthetic data using python.			
5.	Implementation of simple and multiple linear regression using python.			

This is to Certify that, the experiments mentioned above were executed within the four walls of Institute by _____

Lecture-in-Charge

Head of Department

Principal



Chhatrapati Shahu Maharaj Shikshan Sanstha's

CHH. SHAHU COLLEGE OF ENGINEERING

Kanchanwadi, Palthan Road, Aurangabad.

Date :

Experiment No. 1.

Aim :- Implement and demonstrate the FIND-5 algorithm for finding the most specific hypothesis based on a given set of training data samples.
Read the training data from a .csv file.

Theory :-

Introduction to find-5 algorithm :-

The find-5 algorithm is a search algorithm designed to efficiently search through a sorted list of data elements to find the index of a specific target element. The acronym FINDS stands for "Fast Index-based non-recursive discrete search." This algorithm works by recursively dividing the search space into smaller and smaller subspaces until the target element is found or it is determined that the element does not exist in the list.

The FINDS algorithm is particularly useful when searching through large sorted datasets, where linear search algorithms would be too slow. This algorithm has a time complexity of $O(\lg n)$, which means that it can efficiently search through datasets with millions or even billions of elements in a reasonable amount of time.



The FINDS algorithm is widely used in computer science and is particularly important in the fields of database management and information retrieval, where efficient search algorithms are essential for managing and querying large datasets.

Dataset Description :-

The FINDS algorithm is not a dataset, but rather a search algorithm designed to search through sorted datasets. The algorithm works by recursively dividing the search space into smaller and smaller subspaces until the target element is found or it is determined the element does not exist in the list.

To use the FINDS algorithm, you would need a sorted dataset to search through, as it is designed specially for searching through a sorted list of data elements. The dataset can be any type of data that can be sorted, such as numerical data, text data, or any other type of structured data.



conclusion:-

FINDS algorithm is a fast and efficient search algorithm that is designed to search through sorted lists of data elements. By dividing the search space into smaller and smaller subspaces, the algorithm can quickly locate the index of a target element, even in very large datasets. Overall, understanding the FINDS algorithm can be very beneficial for computer science students and professionals, as it can be used to optimize and speed up a wide range of search and retrieval tasks.



Experiment No. 02

Aim :- Implement and demonstrate the decision candidate elimination algorithm.

Theory :-

Introduction to candidate elimination algorithm :-

Candidate elimination is a process used in machine learning to narrow down the set of potential algorithms that can be used to solve a particular problem. It involves evaluating each candidate algorithm against a set of criteria, and eliminating those that do not meet the criteria.

The criteria used for candidate elimination may vary depending on the specific problem being addressed, but typically include factors such as the algorithms' accuracy, computational efficiency, scalability, interpretability, and ease of implementation. These factors are often prioritized based on the specific requirements of the problem and the available data.

To begin the candidate elimination process, a set of potential algorithms is identified based on their ability to solve similar problems or their suitability for the specific problem being addressed. Each algorithm is then evaluated against the criteria, and those that do not meet the criteria are eliminated from consideration.



The remaining candidate algorithms further evaluated and compared using performance metrics such as accuracy, precision, recall and F1 score. The algorithm with the best performance on the training and testing data then chosen as the final model.

Candidate elimination is an important step in the machine learning process, as it helps to narrow down the set of potential algorithms and identify the most suitable one for a particular problem. By carefully evaluating and comparing the available options, it is possible to choose an algorithm that can effectively solve the problem at hand.

Dataset Description :-

A dataset is a collection of data is used for analysis, modelling, and machine learning. In machine learning, a dataset typically consists of a set of input features and a corresponding target variable, which is the variable that the algorithm is trying to predict.



Conclusion :- candidate elimination is ~~a~~ an essential step in the machine learning process, as it helps to narrow down the set of potential algorithms and identify the most suitable option for a particular problem.



Experiment No. 3

Aim :- Implement and demonstrate the supervised machine learning algorithm decision-trees.

Theory :- 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. 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 and each leaf node represent the outcome.

In a decision tree, there 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.

The decisions on the test are performed on the basis of features of the given dataset.



It is a graphical representation getting all the possible solutions to a problem decision based on given conditions.

It is called a decision tree because similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

In order to build a tree, we use CART algorithm, which stands for classification and regression tree algorithm.

A decision tree simply asks a question based on the answer (yes/no), it further divides the tree into subtrees.

Entropy :-

Entropy is nothing but the uncertainty in our dataset or measure of disorder. Let us try to explain this with the help of an example.

Suppose you have a group of friends who decides which movie they can watch together on Sunday. There are 2 choices for movies, the first is "Lucy" and the second is "Titanic" and everyone has to tell their choice. After everyone gives their answer we see that "Lucy" gets 7 votes and "Titanic" gets 5 votes.



which movie do we watch now? isn't it hard to choose a movie now? because the votes for both the movies are somewhat equal.

This is exactly what we call disorderness, there is an equal number of votes for both the ~~movies~~ movies, and we can't really decide which movie we should watch. It would have been much easier if the votes for "Lucy" were 8 and for "Titanic" it was 2. Here we could easily say that the majority of votes are for "Lucy" hence everyone will be watching this movie.

Information gain :-

Information gain measures the reduction of uncertainty given some features and it is also a deciding factor for which attribute should be selected as a decision node or root node.



conclusion :- Hence, we have successfully
~~also~~ implemented and demonstrated the
machine learning algorithm decision tree.