International Journal of Aquatic Science
ISSN: 2008-8019
Vol 14, Issue 01, 2023

A Comparative Study of Supervised Machine Learning

Ishwar Prasad Suryawanshi[1], Vishwas Victor[2], Nidhish Choubey[3], Varsha Singh[4]

[1,2,3,4]*Assistant Professor, D.P. Vipra College, Bilaspur (C.G.), India*

Abstract: Machine learning methods are an effective way to classify data. Machine learning is a broad and fascinating field. Even today, machine learning technology runs a substantial part of our life, often without our knowing it. Machine learning is also fascinating in its own right for the philosophical questions it raises about what it means to learn and succeed at tasks. Machine learning is about predicting the future based on the past. In machine learning, the aim is to fit a model to the data. In this paper, a study between different supervised machine learning algorithms: Support Vector Machine (SVM), Decision Tree, Naive Bayes (NB) and K Nearest Neighbors (k-NN). The main objective is to assess the correctness in classifying data with respect to efficiency and effectiveness of each algorithm in terms of accuracy, precision, sensitivity and specificity.

Keyword: Machine learning, Decision Tree, Naïve Bayes, K-Nearest Neighbors, Support vector machine.

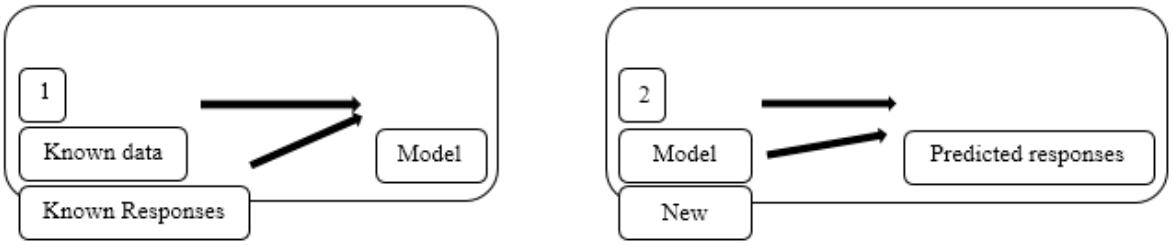
1. INTRODUCTION

This paper documents the basic concepts relating to Machine Learning. Modern day Machine Learning tasks are far more challenging due to an unprecedented increase in the amount and complexity of data. Machine learning is a broad and fascinating field. Even today, machine learning technology runs a substantial part of our life, often without our knowing it. Machine learning is also fascinating in its own right for the philosophical questions it raises about what it means to learn and succeed at Machine learning is about predicting the future based on the past. Some of the biggest transformations in our lives in the last half century are due to computing and digital technology. The tools, devices, and services we had invented and developed in the centuries before, have been increasingly replaced by their computerized "e-" versions, and we in turn have been continuously adapting to this new digital environment.
Machine learning is not just a database or programming problem; it is also a requirement for artificial intelligence. In machine learning, the aim is to fit a model to the data. Machine learning also helps us to find solutions too many problems in vision, speech recognition, and robotics. Application of machine learning methods to large databases is called data mining. Machine learning is programming computers to optimize a performance criterion using example data or past experience.
Machine Learning Model or Algorithm can classify two main subcategories:

1. Supervised Machine Learning Algorithm 2. Unsupervised Machine Learning Algorithm

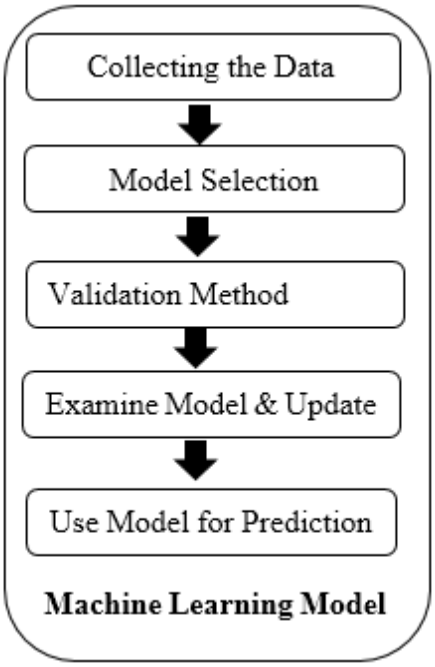
The term algorithm is used in computer science to describe a finite, deterministic, and effective problem solving method suitable for implementation as a computer program.

Supervised Machine Learning:
A supervised learning algorithm takes a known set of input data and known responses to the data (output), and trains a model to generate reasonable predictions for the response to new data.



Supervised learning splits into two broad categories: classification and regression. In classification, the goal is to assign a class (or label) from a finite set of classes to an observation. In regression, the goal is to predict a continuous measurement for an observation. That is, the responses variables are real numbers.

Machine Learning Model:



Collecting the Data:
All supervised machine learning methods start with an input data matrix or dataset.

2. Model Selection:
Supervised learning models are the models where there is a clear distinction between explanatory and dependent variables. The models are trained to explain dependent variables using explanatory variables. In other words, the model output attributes are known beforehand. There are several selection criteria of above models, such as:

1. Speed of Training
2. Memory Usage
3. Predictive accuracy of new data
4. Transparency or interpretability

There are several model or algorithms, such as:

1. Decision Tree 2. Naïve Bayes 3. Support Vector Machine 4. K-Nearest Neighbor

2.1. Decision Tree:

A Decision Tree is a flowchart-like tree structure, where internal node (Non-leaf test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) denotes a holds a class label. The topmost node in a tree is the root node.

The Learning and classification steps of decision tree are simple and fast. Decision Tree algorithms have been used for classification in many application areas such as medicine, financial analysis, astronomy and molecular biology.

2.2. Naïve Bayes

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. Simple Bayesian classifier known as naïve Bayesian classifier to be comparable in performance with decision tree and selected neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class-conditional independence. It is made to simplify the computations involved and, in this sense, is considered "naïve".

2.3. Support Vector Machine

A Support vector machine transforms training data into a higher dimension, where it finds a hyperplane that separates the data by class using essential training tuples called support vector. SVMs, a method for the classification of both linear and nonlinear data.

SVMs can be used for numeric prediction as well as classification. SVMs algorithms have been used for classification in many application areas such as handwritten digit recognition, object recognition, and speaker identification.

2.4. k-NN

Lazy learners or instance-based method of classification, which store all of the training tuples in pattern space and wait until presented with a test tuple before performing generalization. Nearest neighbour classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n attributes. Each tuple represents a point in an n-dimensional space. k-Nearest Neighbor classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k "nearest neighbors" of the unknown tuple.

Validation Method:

The three main methods to examine the accuracy of the resulting model are:

1. Examine the re-substitution error.
2. Examine the Cross-validation error.
3. Examine the out-of-bag error for bagged decision trees.

Examine Model and Update Until Satisfied:
After validating the model, we change it for better accuracy, better speed, or to use less memory.
* Change Parameters to try to get a more accurate model.
* Change parameters to try to get a smaller model. This sometimes gives a model with more accuracy.
*Try a different algorithm

Use Model for Prediction:
When satisfied with a model of some type, removes training data and other properties not required for prediction, from the model to reduce memory consumption.

3. CONCLUSION

The classification methods decision tree, Bayesian classification, support vector machines are example of eager learners. Eager learners, when given a set of training tuples, will construct a generalization (i.e. classification) model before receiving new (e.g. test) tuples to classify. The classification method k-Nearest Neighbor are example of lazy learner. Lazy learners simply store training tuple or does only a little minor processing and waits until it is given a test tuple. Unlike eager learning methods, lazy learners do less work when a training tuple is presented and more work when making a classification or numeric prediction.

4. REFERENCES

[1] Alpaydin, E. (2016). Machine Learning: The new AI, MIT Press.
[2] Han, J. & Kamber, M. & Pei, J. (2012). Data Mining Concepts and Techniques. 3/e. MK Publishers.
[3] Brawnlee, J. (2016). Master Machine Learning Algorithms. Edition v1.1.
[4] Burges, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Bell Laboratories and Lucent Technologies.
[5] Daume III, H. (2017). A Course on Machine Learning. Second printing, Self-Piblished.