

# Summary: An Introduction to Machine Learning for Students in Secondary Education

*Steven D. Essinger, Gail L. Rosen*

**Summary** by *Abhishek Arya*

The paper elucidates a proposed machine learning (ML) lab for high school students, helping them to come up with machine learning based solutions that they can implement in their daily life. These are applicable in almost any aspect of life, in any type of scenario. The primary motivation behind the lab is that just simple math and engineering concepts are required to work with most ML algorithms today.

It describes machine learning as automating the process of coming up with a solution to a problem and talks about pattern recognition and identifying patterns in data to decide output based on that pattern.

Neural networks are mentioned as one of the most extensively used ML models around. Inspired by the neurons from the brain, they are very imprecise, yet they have shown “tremendous success in several ML applications”. A few examples of trending ML applications such as – Speech Recognition aka Natural Language Processing, Image Processing, DNA Sequence Classification, Financial Analysis, Sports Predictions, and Search Engine Algorithms are cited.

How simple maths can be used to exploit the benefits of ML and how this enables a high school student to work with ML is delineated using Euclidean distance and the mean as exemplars. The paper justifies how ML is multi-disciplinary and can expand further according to the research area that it is being applied to.

Steps in a typical ML solution are elaborated – Problem Formulation, Feature Extraction, Model Selection, Model Implementation, and Evaluation.

The following terms are introduced:

1. Unsupervised Solution – solution dealing with unknown input data
2. Class – Type or Category
3. Feature – information used to distinguish between classes
4. Feature Extraction – the process in which the user provides features

The paper also emphasizes that the problem formulation, feature selection and choosing an algorithm is crucial for the overall correctness and accuracy of the final result. If poor features are chosen, then the performance will be low, this is known as garbage-in, garbage-out theorem.

Furthermore, various problems belonging to different disciplines and bailiwick are discussed throughout the paper. The K-means algorithm based on the concept of Euclidean Distance is chosen to solve the problem (lab modules) in the rest of the paper. We are open to choose any evaluation metric, but the rand index is chosen for evaluation which checks the similarity of the output set (X) with the truth set (Y).

The K-means algorithm is briefly explained as an algorithm that groups the input data into a number of groups specified by the user. We initially provide the algorithm with some features as we must provide one free parameter in all ML techniques, this is called no free lunch theorem.

Some class exercises are also discussed in brief:

1. Coin Sorting [three features – diameter, luster, weight, and bacterial composition]
2. Recycling Containers [K-means, rand index, two features – opacity and weight,  $k = 3$ ]
3. Bacterial Classification [K-means, rand index, three features – shape, size, and mobility,  $k = 2$ ]

Some class discussion topics are mentioned such as experimenting by changing the features or the number of features, visualization of data, developing techniques to automatically extract the features, and trying the model with different datasets but with the same parameters.

Finally, it expands upon the implementation of the lab. The author found out that the briefing of the students on the lab and background and a short discussion afterward will lead to a better response from students. To be able to successfully implement the lab as discussed in the paper, the students must have an active participation and discussions should be of prime focus.

The lab has a great scope for enhancement because of the scalable nature of ML. From using neural networks and advanced ML techniques as well as algorithms inspired from biological beings. Activities can be developed for students from any discipline including but not limited to creative arts.

I believe that the proposed lab will not only act as an introduction to ML, but also to engineering, problem analysis, and decision making for students in secondary education and more and more problems can be solved using ML enabling us to be more efficient and accurate in solving problems from virtually any domain.