International Scholars Journals

African Journal of Library and Information Science ISSN 2756-3383 Vol. 7 (2), p. 001, September, 2021.
Available online at www.internationalscholarsjournals.com © International Scholars Journals



*Editorial*

# Techniques to make the data effective for machine learning

**Greg Ching***

Department of Computer Information, Lunghwa University of Science and Technology, Taoyuan City, Taiwan.

Accepted 16 September , 2021

## EDITORIAL NOTE

Health The process of automating the tasks of applying machine learning to real-world issues is known as Automated Machine Learning (AutoML). From the raw dataset to the deployable machine learning model, AutoML covers the entire pipeline. It was offered as an AI-based answer to the ever-increasing issue of machine learning application. AutoML's high level of automation enables non-experts to employ machine learning models and techniques without having to become machine learning professionals.

Practitioners in a typical machine learning application have a collection of input data points to train with. The raw data may not be in a format that can be used by all algorithms. An expert may need to use proper data pre-processing, feature engineering, feature extraction, and feature selection procedures to make the data suitable for machine learning.

Data preprocessing is a crucial phase in the data mining process that involves manipulating or removing data before it is utilised to ensure or improve performance. Analyzing data that hasn't been thoroughly checked for such issues can lead to false conclusions. As a result, before doing any analysis, the representation and quality of data must come first. In many cases, especially in computational biology, data preprocessing is the most crucial aspect of a machine learning project.

Feature engineering is the process of extracting features (characteristics, qualities, and attributes) from raw data using domain expertise. A feature is a quality shared by independent units that can be used for analysis or prediction. Predictive models use them, and the results are influenced by them. In machine learning initiatives, feature engineering has been used.

Feature extraction in machine learning starts with a set of measured data and builds derived values (features) that are meant to be useful and non-redundant, easing the learning and generalization phases and, in some situations, leading to improved human interpretations. Dimensionality reduction is linked to feature extraction. When an algorithm's input data is too vast to analyse and is suspected of being redundant, it can be reduced to a smaller collection of features also called as feature vector. Feature selection is the process of determining a subset of the initial features. The selected features should contain the necessary information from the input data, allowing the intended task to be completed using this reduced representation rather of the entire initial data.

Feature selection, also known as variable selection, attribute selection, or variable subset selection in machine learning and statistics, is the process of choosing a subset of relevant features (variables, predictors) for use in model creation.

Following these above stages, practitioners must choose an algorithm and optimise hyperparameters to improve their model's prediction performance. The task of selecting a set of ideal hyperparameters for a learning algorithm is known as hyperparameter optimization or tuning. A hyperparameter is a value for a parameter that is used to influence the learning process. Other factors, such as node weights, are, on the other hand, learned. Each of these phases might be difficult, making machine learning difficult to implement.

Automating the entire machine learning process has the added benefit of improving simpler solutions, faster generation of those solutions, and models that frequently outperform hand-designed models. AutoML was used in a prediction model to compare the relative importance of each factor.

*Corresponding author. Greg Ching, E-mail: ggcg.12@gmail.com