

General Architecture for Automation of Machine Learning Practices

Authors : U. Borasi, Amit Kr. Jain, Rakesh, Piyush Jain

Abstract : Data collection, data preparation, model training, model evaluation, and deployment are all processes in a typical machine learning workflow. Training data needs to be gathered and organised. This often entails collecting a sizable dataset and cleaning it to remove or correct any inaccurate or missing information. Preparing the data for use in the machine learning model requires pre-processing it after it has been acquired. This often entails actions like scaling or normalising the data, handling outliers, selecting appropriate features, reducing dimensionality, etc. This pre-processed data is then used to train a model on some machine learning algorithm. After the model has been trained, it needs to be assessed by determining metrics like accuracy, precision, and recall, utilising a test dataset. Every time a new model is built, both data pre-processing and model training—two crucial processes in the Machine learning (ML) workflow—must be carried out. Thus, there are various Machine Learning algorithms that can be employed for every single approach to data pre-processing, generating a large set of combinations to choose from. Example: for every method to handle missing values (dropping records, replacing with mean, etc.), for every scaling technique, and for every combination of features selected, a different algorithm can be used. As a result, in order to get the optimum outcomes, these tasks are frequently repeated in different combinations. This paper suggests a simple architecture for organizing this largely produced “combination set of pre-processing steps and algorithms” into an automated workflow which simplifies the task of carrying out all possibilities.

Keywords : machine learning, automation, AUTOML, architecture, operator pool, configuration, scheduler

Conference Title : ICAIDM 2024 : International Conference on Artificial Intelligence and Data Mining

Conference Location : New Delhi, India

Conference Dates : February 19-20, 2024