CompareML

A comparator for machine learning algorithms libraries and services

---

User Manual

for CompareML 1.0

A. J. Fernández-García,
J. C. Preciado,
J. M. Conejero,
R. Rodríguez-Echeverria,
F. Sánchez.

This user guide is for *CompareML* version 1.0

Antonio Jesús Fernández-García, Juan Carlos Preciado, Jose María Conejero, Roberto Rodríguez-Echeverria, Fernando Sánchez.

Quercus Research Group, University of Extremadura
Applied Computing Group, University of Almeria

Welcome to the *CompareML* User Manual

Table of Contents

1. Introduction

2. Installation

3. Functionalities

The functionalities of the machine learning providers \sout{i.e., the experiments carried out using each provider's libraries, data structures and functions} are isolated in microservices~\cite{microserviceweb} making it easier to update the algorithms, maintain the code, add new providers or delete them, in line with a microservice-based architecture.

%Give a short overview of the overall software architecture.

% \subsection{Software Functionalities}

% \label{SoftwareFuncionalities}

%Present the major functionalities of the software.

% The main functionalities of \textit{CompareML} are:

% \begin{itemize}

% \item {\bf Dataset selection}. This functionality allows users to select the dataset on which they want to perform the experiments. To provide the dataset they can either upload a file directly from their computers or choose between some defaults dataset that \textit{CompareML} put at their disposal. The default dataset provided are are the \textcolor{red}{xxx, xxx, and xxx datasets}. The customized dataset uploaded by users should be given in a CSV file format.

% \item {\bf Label Selection (target feature)}. This functionality allows us to define the label feature, \textit{i.e.,} the column of the dataset which will be predicted by the models built by the machine learning algorithms. Only one label can be selected and users can select them through a drop-down menu after selecting the dataset.

% \item {\bf Providers Selection}. Through the provider selection functionality, users can choose between the variety of providers available in \textit{CompareML}. As commented in Subsection \ref{providers}, in this version the libraries and tools are \textit{Turi Graphlab Create}, \textit{Scikit-Learn} and \textit{R}.

% \item {\bf Algorithms Selection}. Through the algorithms selection functionality, users can choose between the variety of algorithms available in \textit{CompareML}. As commented in Subsection \ref{algorithms}, in this version the \textit{Random Forest}, \textit{xxx} and \textit{xxx} \texttt{Classification} algorithms and the \textit{xxxx}, \textit{xxx} and \textit{xxx} \texttt{Regression} algorithms are available. In each experiment only one type of algorithms (\texttt{Classification} or \texttt{Regression}) is allowed, due to their different predicting nature (categories and numbers respectively).

% \item {\bf Experiment Running}. This functionality compiles all the data entered by the user, ensures that it is correct and processes it to perform the experiments that users need to carry out. If the data is wrong or incomplete, the user is notified.

% \item {\bf Results Visualization}. This functionality report with the results of the experiments carried out. Where \texttt{Classification} algorithms are concerned, metrics such as \textit{accuracy}, \textit{auc}, \textit{f1_score}, \textit{precision}, \textit{recall} or a \textit{confusion matrix}\cite{ConfusionMatrix} are shown. Where \texttt{Regression} algorithms are concerned, metrics such as \textit{xxx} or \textit{auc} are shown.

% \end{itemize}

3.1. A

3.1. B

4. Illustrative Example