Best R Libraries for Machine Learning

Paulo Henrique S. Guimarães

ICET/DES - UFLA

paulo.guimaraes@ufla.br

Introduction

Machine learning (ML) algorithms have shown great promise in providing solutions to a number of complex problems.

Introduction

Machine learning (ML) algorithms have shown great promise in providing solutions to a number of complex problems.

ML is the application of artificial intelligence (AI) to statistics and statistical analysis.

Introduction

Machine learning (ML) algorithms have shown great promise in providing solutions to a number of complex problems.

ML is the application of artificial intelligence (AI) to statistics and statistical analysis.

The techniques used in ML allows you to automate the search for patterns in data.

Introduction

Some applications:

- Credit risk analysis;
- Medical diagnostics;
- Marketing;
- Credit card fraud detection;
- Natural language processing;
- Recommendation systems;
- And many more!

Types of problems and tasks

- Supervised learning
- Unsupervised learning
- Reinforcement learning.

Supervised learning

In supervised learning, the training data you feed the algorithm includes the target values.

Supervised learning

In supervised learning, the training data you feed the algorithm includes the target values.

In this way, the solutions can be used to help supervise the training process to find the optimal algorithm parameters.

- Regression problem
- Classification problem.

Supervised learning

Predictive analytics methods typically provide for two phases which are:

- **Construction (or training) of the model** - input and output values are known;

Supervised learning

Predictive analytics methods typically provide for two phases which are:

- **Construction (or training) of the model** - input and output values are known;
- **Prediction** - Classification for nominal data (categorical variable) and regression for numerical variable.

Supervised learning - main models

There are many models used in machine learning (ML), of which we can highlight:

Supervised learning - main models

There are many models used in machine learning (ML), of which we can highlight:

- Classic linear and nonlinear functional models;
- Decision trees;
- Random forests;
- Artificial neural networks (ANN);
- Probabilistic methods (Bayesian);
- K nearest neighbors - KNN;
- Support Vector Machines (SVM);
- ...

Supervised learning - some packages (best)

There are many packages currently used in machine learning. Among these packages, we can mention:

- **caret**;

Supervised learning - some packages (best)

There are many packages currently used in machine learning. Among these packages, we can mention:

- **caret**;
- **caTools**;

Supervised learning - some packages (best)

There are many packages currently used in machine learning. Among these packages, we can mention:

- **caret**;
- **caTools**;
- dplyr;

Supervised learning - some packages (best)

There are many packages currently used in machine learning. Among these packages, we can mention:

- **caret**;
- **caTools**;
- dplyr;
- **e1071**;

Supervised learning - some packages (best)

There are many packages currently used in machine learning. Among these packages, we can mention:

- **caret**;
- **caTools**;
- dplyr;
- **e1071**;
- ggplot2;

Supervised learning - some packages (best)

There are many packages currently used in machine learning. Among these packages, we can mention:

- **caret**;
- **caTools**;
- dplyr;
- **e1071**;
- ggplot2;
- **randomForest**;

Supervised learning - some packages (best)

There are many packages currently used in machine learning. Among these packages, we can mention:

- **caret**;
- **caTools**;
- dplyr;
- **e1071**;
- ggplot2;
- **randomForest**;
- **rpart**.

Application example

	A	B	C	D	E
1	Sepal Length	Sepal Width	Petal Length	Petal Width	Class
2	5.1	3.5	1.4	0.2	Iris-setosa
3	4.9	3	1.4	0.2	Iris-setosa
4	4.7	3.2	1.3	0.2	Iris-setosa
5	4.6	3.1	1.5	0.2	Iris-setosa
6	5	3.6	1.4	0.2	Iris-setosa
7	5.4	3.9	1.7	0.4	Iris-setosa
8	4.6	3.4	1.4	0.3	Iris-setosa
9	5	3.4	1.5	0.2	Iris-setosa
10	4.4	2.9	1.4	0.2	Iris-setosa
11	4.9	3.1	1.5	0.1	Iris-setosa
12	5.4	3.7	1.5	0.2	Iris-setosa
13	4.8	3.4	1.6	0.2	Iris-setosa
14	4.8	3	1.4	0.1	Iris-setosa
15	4.3	3	1.1	0.1	Iris-setosa
16	5.8	4	1.2	0.2	Iris-setosa
17	5.7	4.4	1.5	0.4	Iris-setosa
18	5.4	3.9	1.3	0.4	Iris-setosa
19	5.1	3.5	1.4	0.3	Iris-setosa
20	5.7	3.8	1.7	0.3	Iris-setosa

Final considerations

- **H2O**: an open source, memory, distributed, ML and predictive analytics platform, allowing you to build and produce ML models.

Final considerations

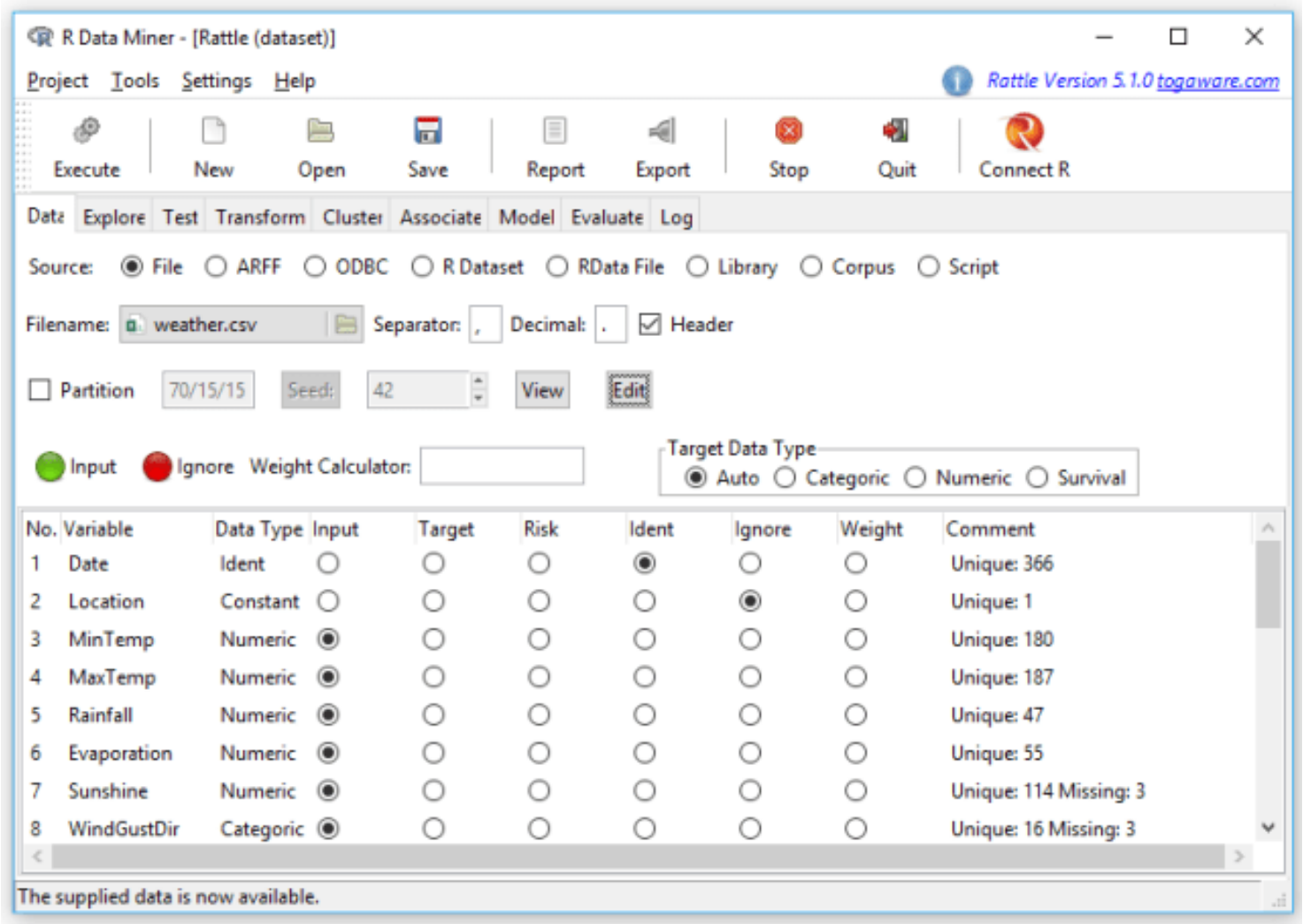
- **H2O**: an open source, memory, distributed, ML and predictive analytics platform, allowing you to build and produce ML models.
- **KernLab**: Kernel-Based Machine Learning Lab is a package for classification, regression, clustering, novelty detection, quantile regression and dimensionality reduction.

Final considerations

- **H2O**: an open source, memory, distributed, ML and predictive analytics platform, allowing you to build and produce ML models.
- **KernLab**: Kernel-Based Machine Learning Lab is a package for classification, regression, clustering, novelty detection, quantile regression and dimensionality reduction.
- **nnet**: Feed - Forward Neural Networks and Multinomial Log-Linear Models.

Final considerations

- **rattle**: a free graphical user interface for Data Science.



References

1) Bruce, P.; Bruce, A; Gedeck,P. **Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python**. 2nd edition. O'Reilly., 2020.

2) Gareth, J., Daniela Witten, Trevor Hastie, Robert Tibshirani. **An Introduction to Statistical Learning : with Applications in R**. New York :Springer, 2013.

3) Hastie, T., Tibshirani, R. Friedman, J. (2001). **The Elements of Statistical Learning**. New York, NY, USA, Springer New York Inc.

4) Schmuller, J. **R Projects for Dummies**. Hoboken, New Jersey : John Wiley Sons, Inc., 2018.