Package 'ResIndex'

July 21, 2025

Type Package

Title Generate Simple yet Effective Metric of Feature Importance for Classification Problems

Version 0.1.0

Description An intuitive and explainable metric of Feature Importance for Classification Problems. Resolution Index measures the extent to which a Feature clusters different classes when data is sorted on it. User provides a DataFrame, column name of the Class, sample size and number of iterations used for calculation. Resolution Index for each Feature is returned, which can be effectively used to rank Features and reduce Dimensionality of Training data. For more details on Feature Selection see Theng and Bhoyar (2023) <doi:10.1007/s10115-023-02010-5>.

License LGPL-2

Encoding UTF-8

NeedsCompilation no

Author Anand Jha [aut, cre]

Maintainer Anand Jha <anandorjha18@gmail.com>

Repository CRAN

Date/Publication 2025-03-18 15:10:15 UTC

Contents

ResIndex 1

Index **3**

`ResIndex` *RESOLUTION INDEX: A SIMPLE YET EFFECTIVE METRIC OF FEATURE IMPORTANCE FOR CLASSIFICATION PROBLEMS*

Description

Provides an intuitive and explainable metric of Feature Importance for Classification Problems. Resolution Index measures the extent to which a Feature clusters different classes when data is sorted on it. User provides a DataFrame, column name of the Class, sample size and number of iterations used for calculation. Resolution Index for each Feature is returned, which can be effectively used to rank Features and reduce Dimensionality of Training data for ehnanced accuracy of Supervised Learning Models.

Usage

```
ResIndex(df, class, f=0.7, N=5, seed=NULL)
```

Arguments

df	R dataframe used for Classification
class	Column Name of Class
f	Sample size as fraction of total data size
N	Number of iterations
seed	seed if desired

Details

Resolution Index internally utilizes a metric called Transition. In a Classification Problem, it is count of instances where the Class changes between consecutive records. An important Feature should reduce Transition much higher than an irrelevant Feature, when data is sorted on it. This Resolution ability of a Feature to segregate different Classes is calculated as a non-dimensional metric and can be used to rank and filter Features in high dimensional data.

Value

Maximum Transition (between 0 and 1) Random Transition (between 0 and 1) Resolution Index (for each Feature)

Author(s)

Anand Jha, Senior Data Scientist

References

D. Theng and K. Bhoyar (2023) "Feature selection techniques for machine learning: a survey of more than two decades of research" <https://doi.org/10.1007/s10115-023-02010-5> J. Tang, S. Alelyani and H. Liu (2014) "Feature selection for classification: A review" <https://doi.org/10.1201/b17320>

Examples

```
ResIndex(iris, Species)
ResIndex(iris, Species, 0.8, 10)
ResIndex(iris, Species, 0.8, 10, seed=123)
```

Index

ResIndex, 1