
cognito Documentation

Release 0.0.1

Bastin Robins .J

Jun 17, 2020

Contents:

1 Collaborators	3
1.1 Installation	3
1.2 Quickstart	3
1.3 Check	3
1.4 Transform	4
1.5 Table	5
1.6 Authors	6
1.7 License	7
1.8 Changelog (Pillow)	7
1.9 Help	7
2 Indices and tables	9
2.1 Infrastructure support	9



Deployment	
Build Status	
Metrics	coverage 28%
GitHub	
License	
Community	

Cognito is an exclusive data preprocessing library that involves transforming raw data into a machine-understandable format. We at CleverInsight Open Ai Foundation took the initiative to build a better automated data preprocessing library and here it is.



CHAPTER 1

Collaborators



CHRIST (Deemed to be University) Lavasa provides developer & mentor contribution via CleverInsight Open AI Foundation Incubation Center at Christ Univeristy Lavasa since 2019.

1.1 Installation

Prerequisite

- Python3.

Install the extension by using pip.

```
$ pip install cognito
```

1.2 Quickstart

1.2.1 Usage

1.3 Check

1.3.1 Simple Usage

```
>>> from cognito.check import Check
>>> import congito
```

1.3.2 is_categorical

is_categorical is a staticmethod which checks if a column is categorical or not.

```
>> Check.is_categorical(data['Age'])
>> True
```

1.3.3 is_discrete

is_discrete is a staticmethod which takes a column as an input and checks if it is discrete or not.

```
>> Check.is_discrete(data['Age'])
>> True
```

1.3.4 is_continuous

is_continuous is a staticmethod which takes a column as an input and checks if it is continuous or not.

```
>> Check.is_continuous(data['Age'])
>> True
```

1.3.5 is_missing

is_missing is a staticmethod which checks if a column has missing values or not.

```
>> Check.is_missing(data['population'])
>> True
```

1.3.6 is_identifier

is_identifier is a staticmethod which checks if a column is an identifier or not.

```
>> Check.is_identifier(data['Name'])
>> True
```

1.4 Transform

1.4.1 encoding_categorical

```
>> Check.encoding_categorical(data['Age'])
>> False
```

1.4.2 remove_records

remove_records removes the missing data from the given dataframe row wise.


```
>> Check.remove_records(data)
>> dataframe
```

1.4.3 remove_columns

`remove_columns` removes the missing data from the given dataframe column wise.

```
>> Check.remove_columns(data)
>> dataframe
```

1.4.4 percentage_missing

`percentage_missing` returns a dictionary containing percentage of missing data in each column.

```
>> Check.percentage_missing(data)
>> dictionary
```

1.4.5 replace_mean

`replace_mean` replaces the missing value by the mean of the column.

```
>> Check.replace_mean(data)
>> series
```

1.4.6 replace_median

`replace_median` replaces the missing value by the median of the column.

```
>> Check.replace_median(data)
>> series
```

1.4.7 ignore_identifier

`ignore_identifier` remove the column from the given dataframe if the row is an identifier row.

```
>> Check.ignore_identifier(data)
>> dataframe
```

1.5 Table

1.5.1 correlation

```
>> table = Table('filename.csv')
>> table.correlation()
>> dataframe
```

1.5.2 convert_to_bin

```
>> table = Table('filename.csv')
>> table.convert_to_bin('column_name')
>> list
```

1.5.3 binning

```
>> table = Table('filename.csv')
>> table.binningn('column_name', bins)
>> dataframe
```

1.5.4 slice

```
>> table = Table('filename.csv')
>> table.slice([column_names])
>> dataframe
```

1.5.5 scale

1.5.6 covariance

```
>> table = Table('filename.csv')

>> table.scale([column_names], mode)
>> dataframe

>> table.covariance()
>> dataframe
```

1.6 Authors

1.6.1 Core Lead

- Bastin Robins J <robin@cleverinsight.co>

1.6.2 Development Leads

- Anshul Saxena <anshul.saxena@christuniversity.in>
- Vandana Bhagat <vandana.bhagat@christuniversity.in>

1.6.3 Data Science Team

- Aastik <aastikkharbanda.ak@gmail.com>
- Aishika <aishikabanik98@gmail.com>

- Nibir Paul <nibirpaul491@gmail.com>
- Rishika Kotta <rishikakotta21@gmail.com>
- Tanvi Rath <rath.t98@gmail.com>

1.7 License

BSD 3-Clause License

Copyright (c) 2020, CleverInsight All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

1.8 Changelog (Pillow)

1.8.1 0.0.1 (unreleased)

- Renamed ``modules`` directory to ``check``.
- Modified ``from cognito.modules import check`` to ``from cognito.check import check``.
- Added ``pyfiglet``, ``pyyaml``, ``prettytable`` into required dependencies.

1.9 Help

- `genindex`
- `modindex`
- `search`

2.1 Infrastructure support

- We would like to thank [GitHub](https://github.com) for providing us with [Github Cloud](https://github.com/CleverInsight/cognito) account to automatically build the documentation. <https://github.com/CleverInsight/cognito>’ _.