
ml_belt Documentation

Release 0.1

Adriano Freitas

May 17, 2019

Contents:

1	Getting started	3
1.1	Installation	3
1.2	Preprocessing:	3
2	Prep Class	5
3	Source:	7
4	Indices and tables	9

The machine learning and data science utility belt.

We recommend that you have Anaconda installed: <https://anaconda.org/>

1.1 Installation

You can install **ML Belt** using pip:

```
pip install ml_belt
```

1.2 Preprocessing:

All you need to do is import `Prep` class, create a new object passing a pandas `DataFrame`, so you can chain some steps of pre-processing that will be applied over the result of antecessor.

At the end your modified `DataFrame` will be in `data` attribute.

```
from ml_belt.prep import Prep
import pandas as pd

def custom_fn(df):
    '''Do some modifications on df and return it.'''
    return df

# Prep needs a pandas dataframe
prep_df = Prep(pd.read_csv('data.csv')) \
    .drop_nulls(['col1', 'col2']) \
    .bool_to_int(['col3']) \
    .drop_cols(['nosense_col1', 'nosense_col2']) \
    .fill_null_with('mean', ['col_with_nulls', 'another_coll_with_nulls']) \
    .fill_null_with(-1, ['col_with_nulls2', 'col_with_nulls3']) \
    .apply_custom(custom_fn) \
    .drop_nulls()

# Now you have a modified df
```

(continues on next page)

(continued from previous page)

```
# with all steps above
df = prep_df.df
df.to_csv('data_prep.csv')
```


CHAPTER 2

Prep Class

Come back later #)

CHAPTER 3

Source:

https://github.com/adrianogfreitas/ml_belt

CHAPTER 4

Indices and tables

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