
pipeline Documentation

Release 0.1.0

DSaPP Researchers

May 05, 2017

Contents

1	Johnson County Early Intervention System	3
2	Installation	5
2.1	Setting up the Virtual Environment	5
3	Configuration	7
3.1	Credentials	7
3.2	Constants	7
3.3	Experiment configuration	7
3.4	Evaluation configuration	8
4	Data extraction and cleaning	9
4.1	Extract raw data	9
4.2	Clean relevant tables	9
4.3	Deduplicate identities	9
4.4	Further processing	10
5	The Pipeline: Feature building, modeling, and evaluation	11
5.1	Preprocessing: Feature building	11
5.2	Modeling	11
5.3	Evaluation	11
5.4	Module contents	12
6	Indices and tables	25
	Python Module Index	27

Contents:

Johnson County Early Intervention System

DSSG has partnered with Johnson County and Salt Lake County to build a prototype early intervention system (EIS) for individuals who repeatedly cycle through multiple systems, including jails, EMS, mental health services. Currently, there is little coordination between systems to address each person's underlying needs. An accurate Early Intervention System will quickly identify individuals at risk of contact with any or all systems so our partners can provide appropriate services and interventions to them.

To achieve this goal, we developed models that assign risk scores to individuals making contact with one system of making future contact with another system. These models produce ranked lists of individuals at risk who may receive follow-up care or interventions. The models provide proactive risk warnings at points of contact (e.g., EMS dispatch, jail bookings).

See [this blog post](#) for a broad overview of the project. The paper [Reducing Incarceration through Prioritized Interventions](#) is under review and provides a more in-depth description of the project's implementation and its results. More specific documentation of the code itself can be built with Sphinx by `make -C doc html` or the public documentation may be viewed online at johnson-county-ddj.readthedocs.io.

Use Git to clone the repository.

Setting up the Virtual Environment

The scripts, notebooks, and other tools in this repository rely on a specific Python environment combining Python 2.7 and a set of package versions specified in `requirements.txt`. To ensure that the code runs on your machine, follow the steps outlined below to set up and activate a Python virtual environment with the required configuration:

ONE: Ensure that you have Python 2.7 installed on your machine and that you know the directory where it is installed.

TWO: If you do not have virtual environment installed, install it using:

```
$ pip install virtualenv
```

THREE: Create the virtual environment to use with this software. First change your working directory to the directory where you would like to install the virtual environment. Then, create a virtual environment with the following command, replacing `/usr/bin/python2.7` with the location of your Python 2.7 installation and `venv` with the name you would like to give to the environment:

```
$ virtualenv -p /usr/bin/python2.7 venv
```

FOUR: Activate the virtual environment, replacing `venv` with the directory you just created:

```
$ source venv/bin/activate
```

To make activating the virtual environment in the future easier, consider adding an alias to your `.bashrc` or `.bash_profile`:

```
alias venv="source /PATH/TO/VIRTUAL/ENVIRONMENT/venv/bin/activate"
```

FIVE: To configure the virtual environment to use the correct packages and versions, run the following commands, pointing to the `requirements.txt` file in the repository for the final one:

```
$ pip install numpy==1.11.2 $ pip install scipy==0.18.1 $ pip install -r requirements.txt
```

If this fails, you may need to open `requirements.txt` and install each package individually. For example:

```
$ pip install collate==0.1.0
```

SIX: To set up the virtual environment for use within Jupyter Notebooks, run the following command:

```
$ ipython kernel install --user --name=python2 --install-name=python2 in
/home/USER/.local/share/jupyter/kernels/python2
```

Copy the kernelspec to a directory where ipython will find it and give it a name you will recognize as your virtual environment (*venv* in this example):

```
$ mkdir -p ~/.ipython/kernels $ mv ~/.local/share/jupyter/kernels/python2 ~/.ipython/kernels/venv
```

Then, edit the *kernel.json* file in the directory you just created, changing the JSON key called *display_name* to the name of your virtual environment (e.g., *venv*).

SEVEN: When you are finished working with the tools in this repository, deactivate your virtual environment with:

```
$ deactivate
```

CHAPTER 3

Configuration

A number of configuration files are used to setup database credentials and specify various names and parameters.

Credentials

There are several configuration files that contain confidential information (like credentials) that cannot be committed to the repository. Example files are provided in the `config` directory for both database and s3 credentials. Simply remove the `example_` prefix and populate each appropriately.

Constants

The file `pipeline/default_profile.yaml` contains a number of constants that are used throughout the pipeline codebase, including the paths to the credential files above as well as specific table and column names. It generally does not need to be modified for use with Johnson County's data.

Experiment configuration

The `yamls` folder contains the experimental configuration that gets passed to the pipeline preprocessing and modeling scripts. Each `yaml` file within that folder specifies a very specific experimental configuration. See the `yamls/default_sample.yaml` file for an example configuration. There are several very important categories that are broken out as block comments in that sample configuration:

- **Type of Experiment** rarely changes; the unit entity is a 'person'.
- **Temporal parameters** specify the time blocking, including start dates, the prediction window (in days), and the "fake today" – this specifies the date to simulate the experiment.
- **Labeling details** specify which labels to consider as outcomes. The current set of labels are all based upon interactions with a given data provider. The label names are underscore-delimited, separating the data providers. All but the final component specify the population of interest and are currently *exclusive* of any of the other

providers (this is something that we want to change). The final component specifies the outcome of interest. For example: `jims_mh_jims` labels the population who has previously interacted with criminal justice and mental health (but not EMS) who have a new interaction with the criminal justice system as positive labels.

- **Feature Selection** specifies the feature groups that should be included in the models.
- **Model selection** specifies the models to run and the parameters over which they should be parameterized. All model parameters may be lists and get combined with the cross-product so all possible combinations get tested.
- **Output file details** specifies where the outputs should go.

Evaluation configuration

The variables of interest for evaluation may be specified in the `pipeline/evaluation/eval_profile.yaml` file. These determine which metrics get calculated for each model that was run.

Data extraction and cleaning

ETL is handled by `drake`, which is expected to be called from the top (or root) directory of this repository. All paths assume that `drake` was invoked from the root of the repository.

There are four conceptual steps: extract, clean, deduplicate, and process.

Extract raw data

The scripts here expect to find a zipped data dump in a `data` directory in the repository root. It will extract that, and restore into the database. The database configuration is specified within a `config` directory in the repository root.

Clean relevant tables

The raw tables, as restored from the database dump, get placed into the public schema of the database. Before being used by the pipeline, all tables go through a cleaning and normalization process. All scripts contained within the `etl/data_cleaning` directory are executed at this point.

Deduplicate identities

Deduplication is handled by `superdeduper`. A SQL script creates a master “entries” table by combining all relevant columns from all the data sources, and then `superdeduper` is called with the saved configuration in `etl/dedupe/config.yaml`. After a `dedupe_id` is appended to the entries table, the `apply_results.py` script goes back to the specified tables in the `clean` schema to append a `dedupe_id` column.

Further processing

Finally, a few SQL scripts are used to create computed tables for convenience. These include things like a timeline of events. All scripts contained within the `etl/data_processing` directory are executed at this point.

The Pipeline: Feature building, modeling, and evaluation

There are three major steps to the pipeline: features building, modeling, and model evaluation. Each step is a sub-module of the pipeline and has its own run command-line interface, designed to be run from the repository root as `python -m pipeline.component.run` with command-line arguments. Or all three steps may be run with a single invocation of `python -m pipeline.run`. The `-h` flag will show the help for each command. A broad overview of each component is provided here, with more specific inline documentation in the code and exposed as module documentation below.

Preprocessing: Feature building

The command `python -m pipeline.preprocessing.run yamls/default_sample.yaml` will use the sample experiment configuration to build the required feature table. The feature tables are timestamped with the time at which the command was run.

Modeling

The command `python -m pipeline.modeling.run yamls/default_sample.yaml` will use the sample experiment configuration and the most recently created feature tables in order to train all the models specified in the files at the given splits.

Evaluation

The command `python -m pipeline.evaluation.run` will evaluate all unprocessed models it finds in the database and compute the metrics found in the default evaluation configuration file.

Module contents

pipeline.preprocessing package

Subpackages

pipeline.preprocessing.features package

Submodules

pipeline.preprocessing.features.abstract module

class pipeline.preprocessing.features.abstract.**SimpleFeature**

class pipeline.preprocessing.features.abstract.**TimeBoundedFeature** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.SimpleFeature*

pipeline.preprocessing.features.class_map module

exception pipeline.preprocessing.features.class_map.**UnknownFeatureError** (feature)
Bases: *exceptions.Exception*

pipeline.preprocessing.features.class_map.**lookup** (feature, **kwargs)

pipeline.preprocessing.features.emsfeatures module

class pipeline.preprocessing.features.emsfeatures.**CountOfEms** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**Destination** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**DifferentResidenceOfCityRecorded** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**EverHomelessness** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**Homelessness** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**LastMonthEmsCount** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**LastWeekEmsCount** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**LastYearEmsCount** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**NoTreamentRequiredCount** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*

class pipeline.preprocessing.features.emsfeatures.**PrimaryImpression** (**kwargs)
Bases: *pipeline.preprocessing.features.abstract.TimeBoundedFeature*


```

class pipeline.preprocessing.features.emsfeatures.RefusedCareCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TransportedSum (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TreatedRefusedTransportCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TreatedSum (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TreatedTransferredCareCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TreatedTransportedALSCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TreatedTransportedBLSCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TreatedTransportedByLawEnforcementCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.emsfeatures.TriageOfEms (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

```

pipeline.preprocessing.features.jimsfeatures module

```

class pipeline.preprocessing.features.jimsfeatures.ArrestingAgencyCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.AvgBailAmount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.BailTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.BailedOutCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.CaseTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.CountOfJims (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.CurrentChargesCoarseFinding
    Bases: pipeline.preprocessing.features.abstract.SimpleFeature

class pipeline.preprocessing.features.jimsfeatures.CurrentChargesDrugOffense
    Bases: pipeline.preprocessing.features.abstract.SimpleFeature

class pipeline.preprocessing.features.jimsfeatures.CurrentChargesFelonyOrMisdemeanor
    Bases: pipeline.preprocessing.features.abstract.SimpleFeature

class pipeline.preprocessing.features.jimsfeatures.CurrentChargesFindingTrialOccurred
    Bases: pipeline.preprocessing.features.abstract.SimpleFeature

class pipeline.preprocessing.features.jimsfeatures.CurrentChargesFoundOrPleadGuilty
    Bases: pipeline.preprocessing.features.abstract.SimpleFeature

```

```

class pipeline.preprocessing.features.jimsfeatures.LastMonthAvgBailAmount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastMonthBailTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastMonthCaseTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastMonthJimsCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastWeekAvgBailAmount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastWeekBailTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastWeekCaseTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastWeekJimsCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastYearAvgBailAmount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastYearBailTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastYearCaseTypeCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastYearJailDaysAvg (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastYearJailDaysStddev (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastYearJailDaysSum (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.LastYearJimsCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.jimsfeatures.ProbationType (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

```

pipeline.preprocessing.features.mentalhealthfeatures module

```

class pipeline.preprocessing.features.mentalhealthfeatures.CountOfMentalHealth (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.Diagnoses (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.Discharge (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.ImportantDiagnoses (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

```

```
class pipeline.preprocessing.features.mentalhealthfeatures.LastMonthMhCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.LastWeekMhCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.LastYearMhCount (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.LastYearMhDaysAvg (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.LastYearMhDaysStddev (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.LastYearMhDaysSum (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.MostCommonTherapistNumber (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.NumOfMHServices (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.NumOfUniqueMHServices (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.NumberOfTherapists (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.Program (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.ProgramsDischarges (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.Referral (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.mentalhealthfeatures.ServicesRecieved (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature
```

pipeline.preprocessing.features.miscfeatures module

```
class pipeline.preprocessing.features.miscfeatures.AvgDaysBetweenEvents (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.miscfeatures.IntersectionsPublicService (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.miscfeatures.StdDaysBetweenEvents (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature
```

pipeline.preprocessing.features.person module

```
class pipeline.preprocessing.features.person.Age (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.AgeDiscrete (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature
```

```
class pipeline.preprocessing.features.person.AgeFirstInteractionPublicService (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.AgeFirstInteractionPublicServiceDiscrete (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.AgeFirstInteractionPublicServiceInYears (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.AgeLastInteractionPublicService (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.AgeLastInteractionPublicServiceDiscrete (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.AgeLastInteractionPublicServiceInYears (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.AgeYears (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.person.Gender
    Bases: pipeline.preprocessing.features.abstract.SimpleFeature

class pipeline.preprocessing.features.person.Race (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.SimpleFeature
```

pipeline.preprocessing.features.rsifeatures module

```
class pipeline.preprocessing.features.rsifeatures.AvgIntervalFromInAndDisposition (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.rsifeatures.CountOfRSI (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.rsifeatures.ResidencyRecordedMost (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.rsifeatures.TransportedBy (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature
```

pipeline.preprocessing.features.seqfeatures module

```
class pipeline.preprocessing.features.seqfeatures.BookingBooking182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.BookingBooking365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.BookingBooking730 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.BookingBooking90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.BookingBookingBooking182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.BookingBookingBooking365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature
```

[illegible]


```

class pipeline.preprocessing.features.seqfeatures.EmsBooking730 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsBookingBooking365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsBookingEms182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsBookingEms365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEms182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEms30 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEms365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEms7 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEms730 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEms90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsBooking365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsEms182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsEms30 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsEms365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsEms90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsEmsEms182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsEmsEms365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsEmsEms90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsEmsMh90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsMh182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.EmsMh365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

```

```

class pipeline.preprocessing.features.seqfeatures.EmsMh90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhBooking182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhBooking365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhBooking730 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhBooking90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhBookingBooking365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhEms182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhEms365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhEms90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhMh182 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhMh365 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

class pipeline.preprocessing.features.seqfeatures.MhMh90 (**kwargs)
    Bases: pipeline.preprocessing.features.abstract.TimeBoundedFeature

```

Module contents

Submodules

pipeline.preprocessing.feature_processor module

```

class pipeline.preprocessing.feature_processor.FeatureGrabber (end_date, engine,
                                                             config_db, con)

    getFeature (feature_to_load)

pipeline.preprocessing.feature_processor.convert_categorical (df)
pipeline.preprocessing.feature_processor.feature_name_grabber (df)
pipeline.preprocessing.feature_processor.imputation_zero (df)
pipeline.preprocessing.feature_processor.numerical_column_clean (df)

```

pipeline.preprocessing.feature_table_builder module

```

class pipeline.preprocessing.feature_table_builder.Labeller (start_date, end_date, labels)

```

`get_labels()`

`pipeline.preprocessing.feature_table_builder.chunker(seq, size)`

`pipeline.preprocessing.feature_table_builder.dataframe_merge(d1, d2)`

`pipeline.preprocessing.feature_table_builder.generate_fake_todays(fake_today, prediction_window, start_date)`

Given a final prediction window start date, the length of the prediction windows, and a training start date, return the start and end dates for all prediction windows as a dictionary.

Parameters

- **fake_today** (*datetime*) – start date for the final prediction window
- **prediction_window** (*int*) – length of the prediction windows in days
- **start_date** (*datetime*) – start date for the training period

Returns start and end dates for all prediction windows

Return type dict

`pipeline.preprocessing.feature_table_builder.generate_feature_list(config)`

`pipeline.preprocessing.feature_table_builder.generate_feature_table(config, fake_today, prediction_window, start_date, feature_timestamp)`

`pipeline.preprocessing.feature_table_builder.label_feature_producer(start_date, end_date, features, labels)`

`pipeline.preprocessing.feature_table_builder.merge_feature_dictionaries(d1, d2)`

`pipeline.preprocessing.feature_table_builder.write_dataframe_to_sql(df_name, df, schema)`

pipeline.preprocessing.run module

`pipeline.preprocessing.run.main(config_file_name)`

Module contents

pipeline.evaluation package

Submodules

pipeline.evaluation.eval_old_models module

pipeline.evaluation.evaluation module

pipeline.evaluation.make_precision_recall_at_k_graphs module

pipeline.evaluation.run module

`pipeline.evaluation.run.main` (*eval_config_file*)

Runs evaluation code to generate metrics for models in the models table, stash them in a csv, and upload them in bulk to the metrics table.

Parameters `config_file_name` (*str*) – path to evaluation configuration file

Returns None – always returns None as default

Return type None

pipeline.evaluation.user_timeline module

pipeline.evaluation.utils module

Module contents

pipeline.modeling package

Submodules

pipeline.modeling.feature_model_grabber module

`class pipeline.modeling.feature_model_grabber.FeatureModelGrabber` (*fake_today*,
prediction_window,
config, *feature_timestamp*,
s3_profile, *discard_model*)

`add_labels_to_feature_sets` (*feature_sets*, *labels*)

`combine_models_labels_features` (*models*, *labelled_features*)

`compile_results` (*res*, *bulk_model_list*, *force_write=False*)

After a model is run, compile the model information and the predictions. Temporarily stash them in csvs. If more than 49 models have been stashed or this is the last model to be run, copy the csvs to the models and predictions table, remove the csvs, and return an empty list.

Parameters

- **self** (`FeatureModelGrabber`) – inherit object properties
- **res** (*dict*) – dictionary of model information
- **bulk_model_list** (*list*) – list of model info to be saved to database
- **force_write** (*bool*) – should the stashed info be saved to the database regardless of length?

Returns list of models run since last write

Return type list

connect_to_s3()

Open a connection to s3 and return the resource objects and a dictionary of s3 configuration details.

Returns s3 resource and s3_config

Return type boto3 resource and dict

csv_to_database (*file_name*, *table_name*)

Given a csv name and a database table name, append the contents of the csv to the database table and remove the csv.

Parameters

- **file_name** (*str*) – name of csv to save to database
- **table_name** (*str*) – name of the database table to copy to

Returns None

Return type None

export_data_table (*table*, *end_date*, *label*, *feature_names*)

Save a data set as an HDF table for later reuse.

Parameters

- **table** (*pandas DataFrame*) – the DataFrame to save
- **end_date** (*a date format of some kind*) – end of labeling period
- **label** (*str*) – name of the column containing labels
- **feature_names** (*list*) – names of the columns containing features

Returns the prefix of the HDF filename

Return type str

export_metadata (*end_date*, *label*, *feature_names*)

Construct and export metadata for a matrix. Return a unique identifier based on this metadata to used as a filename.

Parameters

- **end_date** (*str*) – the end date of the labeling period for the matrix
- **label** (*str*) – name of the column containing labels
- **feature_names** (*list*) – names of the columns containing features

Returns unique identifier for the matrix

Return type str

extract_train_x (*feature_set*, *full_feature_table*)

generate_feature_group (*feature_sets*)

generate_feature_group_combinations (*feature_groups*)

generate_model_parameter_list ()

generate_uuid (*metadata*)

Generate a unique identifier given a dictionary of matrix metadata.

Parameters **metadata** (*dict*) – metadata for the matrix

Returns unique name for the file

Return type str

get_feature_sets (*feature_names_dict*)

iterate_train_test (*iterable*)

Iterate over prediction window start dates, returning the start dates for train and test data for the current model.

Parameters **prediction_window_start_dates** – list of prediction window start dates

Type list

Returns train date and test date

Return type

load_feature_name_dictionary ()

load_table (*train_or_test*, *feature_timestamp*)

load_test_table ()

load_train_table ()

parameter_generator (*params_lst*)

pickle_results (*res_dict*, *clf*)

Pickle the model object locally, upload to s3, and delete local copy

Parameters

- **self** (*FeatureModelGrabber*) – inherit object properties
- **res_dict** (*dict*) – dictionary of model information
- **clf** (*model*) – model object

Returns path to pickle file

Return type str

run (*labels*)

upload_file_to_s3 (*key_name*, *bucket*, *local_file_path*)

write_matrix_pairs (*train_test_combos*)

Given a list of train-test pairs, write them locally, check s3 for an existing set, combine the sets, remove duplicates, and upload new copy to s3.

Parameters **train_test_combos** (*list*) – list of dictionaries with keys ‘train’ and ‘test’ with filenames of HDF matrices as values

Returns None

Return type None

write_to_csv (*df*, *column_order*, *file_name*)

Given a dataframe, a specific column order, and a csv filename, enforce the column order on the dataframe and then append the data to the specified csv file.

Parameters

- **self** (*FeatureModelGrabber*) – inherit object properties
- **df** – data output by the modeling process containing either model information or predictions
- **column_order** (*list*) – the order of columns in the relevant database table

- **file_name** (*str*) – the name of the csv file to write to

Returns None

Return type None

`pipeline.modeling.feature_model_grabber.chunker(seq, size)`

`pipeline.modeling.feature_model_grabber.write_dataframe_chunks(df_name, df)`

pipeline.modeling.models module

class `pipeline.modeling.models.ConfigError`

class `pipeline.modeling.models.Model(model_name, model_params, label, training_data, testing_data, cols_to_use, config)`

`compute_confusion_matrix(predicted_labels, labels)`

`define_model(model, parameters, n_cores=0)`

`get_data(df, undersample=False)`

`get_feature_importance(clf, model_name)`

`get_test_data()`

`get_training_data()`

`run()`

pipeline.modeling.run module

`pipeline.modeling.run.main(config_file_name, feature_timestamp, discard_models)`

Replaces template placeholder with values.

Parameters `config_file_name` (*str*) – path to config yaml file

Returns None – always returns None as default

Return type None

Module contents

CHAPTER 6

Indices and tables

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

p

- `pipeline.evaluation`, [21](#)
- `pipeline.evaluation.run`, [21](#)
- `pipeline.modeling`, [24](#)
- `pipeline.modeling.feature_model_grabber`,
[21](#)
- `pipeline.modeling.models`, [24](#)
- `pipeline.modeling.run`, [24](#)
- `pipeline.preprocessing`, [20](#)
- `pipeline.preprocessing.feature_processor`,
[19](#)
- `pipeline.preprocessing.feature_table_builder`,
[19](#)
- `pipeline.preprocessing.features`, [19](#)
- `pipeline.preprocessing.features.abstract`,
[12](#)
- `pipeline.preprocessing.features.class_map`,
[12](#)
- `pipeline.preprocessing.features.emsfeatures`,
[12](#)
- `pipeline.preprocessing.features.jimsfeatures`,
[13](#)
- `pipeline.preprocessing.features.mentalhealthfeatures`,
[14](#)
- `pipeline.preprocessing.features.miscfeatures`,
[15](#)
- `pipeline.preprocessing.features.person`,
[15](#)
- `pipeline.preprocessing.features.rsifeatures`,
[16](#)
- `pipeline.preprocessing.features.seqfeatures`,
[16](#)
- `pipeline.preprocessing.run`, [20](#)

A

add_labels_to_feature_sets()
 (pipeline.modeling.feature_model_grabber.FeatureModelGrabber
 method), 21

Age (class in pipeline.preprocessing.features.person), 15

AgeDiscrete (class in pipeline.preprocessing.features.person),
 15

AgeFirstInteractionPublicService (class in
 pipeline.preprocessing.features.person), 15

AgeFirstInteractionPublicServiceDiscrete (class in
 pipeline.preprocessing.features.person), 16

AgeFirstInteractionPublicServiceInYears (class in
 pipeline.preprocessing.features.person), 16

AgeLastInteractionPublicService (class in
 pipeline.preprocessing.features.person), 16

AgeLastInteractionPublicServiceDiscrete (class in
 pipeline.preprocessing.features.person), 16

AgeLastInteractionPublicServiceInYears (class in
 pipeline.preprocessing.features.person), 16

AgeYears (class in pipeline.preprocessing.features.person),
 16

ArrestingAgencyCount (class in
 pipeline.preprocessing.features.jimsfeatures),
 13

AvgBailAmount (class in
 pipeline.preprocessing.features.jimsfeatures),
 13

AvgDaysBetweenEvents (class in
 pipeline.preprocessing.features.miscfeatures),
 15

AvgIntervalFromInAndDisposition (class in
 pipeline.preprocessing.features.rsifeatures),
 16

pipeline.preprocessing.features.jimsfeatures),
 13

BookingBooking182 (class in
 pipeline.preprocessing.features.seqfeatures),
 16

BookingBooking365 (class in
 pipeline.preprocessing.features.seqfeatures),
 16

BookingBooking730 (class in
 pipeline.preprocessing.features.seqfeatures),
 16

BookingBooking90 (class in
 pipeline.preprocessing.features.seqfeatures),
 16

BookingBookingBooking182 (class in
 pipeline.preprocessing.features.seqfeatures),
 16

BookingBookingBooking365 (class in
 pipeline.preprocessing.features.seqfeatures),
 16

BookingBookingBooking730 (class in
 pipeline.preprocessing.features.seqfeatures),
 16

BookingBookingBookingBooking365 (class in
 pipeline.preprocessing.features.seqfeatures),
 17

BookingBookingBookingBooking730 (class in
 pipeline.preprocessing.features.seqfeatures),
 17

BookingBookingEms365 (class in
 pipeline.preprocessing.features.seqfeatures),
 17

BookingBookingMh365 (class in
 pipeline.preprocessing.features.seqfeatures),
 17

B

BailedOutCount (class in
 pipeline.preprocessing.features.jimsfeatures),
 13

BailTypeCount (class in
 BookingEms182 (class in
 pipeline.preprocessing.features.seqfeatures),
 17

BookingEms30 (class in
 pipeline.preprocessing.features.seqfeatures),

[17](#)
 BookingEms365 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingEms7 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingEms730 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingEms90 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingEmsBooking182 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingEmsBooking365 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingEmsBooking730 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingEmsEms365 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingMh182 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingMh730 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 BookingMhBooking365 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)

C

CaseTypeCount (class in [pipeline.preprocessing.features.jimsfeatures](#)),
[13](#)
 chunker() (in module [pipeline.modeling.feature_model_grabber](#)),
[24](#)
 chunker() (in module [pipeline.preprocessing.feature_table_builder](#)),
[20](#)
 combine_models_labels_features()
 (pipeline.modeling.feature_model_grabber.FeatureModelGrabber
 method), [21](#)
 compile_results() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber
 method), [21](#)
 compute_confusion_matrix()
 (pipeline.modeling.models.Model method), [24](#)
 ConfigError (class in [pipeline.modeling.models](#)), [24](#)
 connect_to_s3() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber
 method), [21](#)
 convert_categorical() (in module

[pipeline.preprocessing.feature_processor](#)),
[19](#)
 CountOfEms (class in [pipeline.preprocessing.features.emsfeatures](#)),
[12](#)
 CountOfJims (class in [pipeline.preprocessing.features.jimsfeatures](#)),
[13](#)
 CountOfMentalHealth (class in [pipeline.preprocessing.features.mentalhealthfeatures](#)),
[14](#)
 CountOfRSI (class in [pipeline.preprocessing.features.rsifeatures](#)),
[16](#)
 csv_to_database() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber
 method), [22](#)
 CurrentChargesCoarseFinding (class in [pipeline.preprocessing.features.jimsfeatures](#)),
[13](#)
 CurrentChargesDrugOffense (class in [pipeline.preprocessing.features.jimsfeatures](#)),
[13](#)
 CurrentChargesFelonyOrMisdemeanor (class in [pipeline.preprocessing.features.jimsfeatures](#)),
[13](#)
 CurrentChargesFindingTrialOccurred (class in [pipeline.preprocessing.features.jimsfeatures](#)),
[13](#)
 CurrentChargesFoundOrPleadGuilty (class in [pipeline.preprocessing.features.jimsfeatures](#)),
[13](#)

D

dataframe_merge() (in module [pipeline.preprocessing.feature_table_builder](#)),
[20](#)
 define_model() (pipeline.modeling.models.Model
 method), [24](#)
 Destination (class in [pipeline.preprocessing.features.emsfeatures](#)),
[12](#)
 Diagnoses (class in [pipeline.preprocessing.features.mentalhealthfeatures](#)),
[14](#)
 DifferentResidenceOfCityRecorded (class in [pipeline.preprocessing.features.emsfeatures](#)),
[12](#)
 Diversion (class in [pipeline.preprocessing.features.mentalhealthfeatures](#)),
[14](#)

E

EmsBooking182 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 EmsBooking365 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)
 EmsBooking730 (class in [pipeline.preprocessing.features.seqfeatures](#)),
[17](#)

EmsBooking365 (class in pipeline.preprocessing.features.seqfeatures),
17
EmsBooking730 (class in pipeline.preprocessing.features.seqfeatures),
17
EmsBookingBooking365 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsBookingEms182 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsBookingEms365 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEms182 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEms30 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEms365 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEms7 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEms730 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEms90 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsBooking365 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsEms182 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsEms30 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsEms365 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsEms90 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsEmsEms182 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsEmsEms365 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsEmsEms90 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsEmsMh90 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsMh182 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsMh365 (class in pipeline.preprocessing.features.seqfeatures),
18
EmsMh90 (class in pipeline.preprocessing.features.seqfeatures),
18
EverHomelessness (class in pipeline.preprocessing.features.emsfeatures),
12
export_data_table() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 22
export_metadata() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 22
extract_train_x() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 22
FeatureNameGrabber (class in pipeline.preprocessing.feature_processor),
19
FeatureGrabber (class in pipeline.preprocessing.feature_processor),
19
FeatureModelGrabber (class in pipeline.modeling.feature_model_grabber),
21
F
G
Gender (class in pipeline.preprocessing.features.person),
16
generate_fake_todays() (in module pipeline.preprocessing.feature_table_builder),
20
generate_feature_group() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 22
generate_feature_group_combinations() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 22
generate_feature_list() (in module pipeline.preprocessing.feature_table_builder),
20
generate_feature_table() (in module pipeline.preprocessing.feature_table_builder),
20
generate_model_parameter_list() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 22
generate_uuid() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 22
get_data() (pipeline.modeling.models.Model method), 24
get_feature_importance() (pipeline.modeling.models.Model method), 24

[get_feature_sets\(\)](#) (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23
[get_labels\(\)](#) (pipeline.preprocessing.feature_table_builder.Labeler method), 19
[get_test_data\(\)](#) (pipeline.modeling.models.Model method), 24
[get_training_data\(\)](#) (pipeline.modeling.models.Model method), 24
[getFeature\(\)](#) (pipeline.preprocessing.feature_processor.FeatureGrabber method), 19

H

[Homelessness](#) (class in pipeline.preprocessing.features.emsfeatures), 12

I

[ImportantDiagnoses](#) (class in pipeline.preprocessing.features.mentalhealthfeatures), 14
[imputation_zero\(\)](#) (in module pipeline.preprocessing.feature_processor), 19
[IntersectionsPublicService](#) (class in pipeline.preprocessing.features.miscfeatures), 15
[iterate_train_test\(\)](#) (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

L

[label_feature_producer\(\)](#) (in module pipeline.preprocessing.feature_table_builder), 20
[Labeller](#) (class in pipeline.preprocessing.feature_table_builder), 19
[LastMonthAvgBailAmount](#) (class in pipeline.preprocessing.features.jimsfeatures), 13
[LastMonthBailTypeCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastMonthCaseTypeCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastMonthEmsCount](#) (class in pipeline.preprocessing.features.emsfeatures), 12
[LastMonthJimsCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastMonthMhCount](#) (class in pipeline.preprocessing.features.mentalhealthfeatures), 14
[LastWeekAvgBailAmount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastWeekBailTypeCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastWeekCaseTypeCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastWeekEmsCount](#) (class in pipeline.preprocessing.features.emsfeatures), 12
[LastWeekJimsCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastWeekMhCount](#) (class in pipeline.preprocessing.features.mentalhealthfeatures), 15
[LastYearAvgBailAmount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastYearBailTypeCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastYearCaseTypeCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastYearEmsCount](#) (class in pipeline.preprocessing.features.emsfeatures), 12
[LastYearJailDaysAvg](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastYearJailDaysStddev](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastYearJailDaysSum](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastYearJimsCount](#) (class in pipeline.preprocessing.features.jimsfeatures), 14
[LastYearMhCount](#) (class in pipeline.preprocessing.features.mentalhealthfeatures), 15
[LastYearMhDaysAvg](#) (class in pipeline.preprocessing.features.mentalhealthfeatures), 15
[LastYearMhDaysStddev](#) (class in pipeline.preprocessing.features.mentalhealthfeatures), 15
[LastYearMhDaysSum](#) (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

load_feature_name_dictionary() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

load_table() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

load_test_table() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

load_train_table() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

lookup() (in module pipeline.preprocessing.features.class_map), 12

NumOfUniqueMHServices (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

M

main() (in module pipeline.evaluation.run), 21

main() (in module pipeline.modeling.run), 24

main() (in module pipeline.preprocessing.run), 20

merge_feature_dictionaries() (in module pipeline.preprocessing.feature_table_builder), 20

MhBooking182 (class in pipeline.preprocessing.features.seqfeatures), 19

MhBooking365 (class in pipeline.preprocessing.features.seqfeatures), 19

MhBooking730 (class in pipeline.preprocessing.features.seqfeatures), 19

MhBooking90 (class in pipeline.preprocessing.features.seqfeatures), 19

MhBookingBooking365 (class in pipeline.preprocessing.features.seqfeatures), 19

MhEms182 (class in pipeline.preprocessing.features.seqfeatures), 19

MhEms365 (class in pipeline.preprocessing.features.seqfeatures), 19

MhEms90 (class in pipeline.preprocessing.features.seqfeatures), 19

MhMh182 (class in pipeline.preprocessing.features.seqfeatures), 19

MhMh365 (class in pipeline.preprocessing.features.seqfeatures), 19

MhMh90 (class in pipeline.preprocessing.features.seqfeatures), 19

Model (class in pipeline.modeling.models), 24

MostCommonTherapistNumber (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

N

NoTreatmentRequiredCount (class in pipeline.preprocessing.features.emsfeatures), 15

ModelOfTherapists (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

ModelGrabber (class in pipeline.preprocessing.feature_processor), 19

NumOfUniqueMHServices (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

P

parameter_generator() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

pickle_results() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

pipeline.evaluation (module), 21

pipeline.evaluation.run (module), 21

pipeline.modeling (module), 24

pipeline.modeling.feature_model_grabber (module), 21

pipeline.modeling.models (module), 24

pipeline.modeling.run (module), 24

pipeline.preprocessing (module), 20

pipeline.preprocessing.feature_processor (module), 19

pipeline.preprocessing.feature_table_builder (module), 19

pipeline.preprocessing.features (module), 19

pipeline.preprocessing.features.abstract (module), 12

pipeline.preprocessing.features.class_map (module), 12

pipeline.preprocessing.features.emsfeatures (module), 12

pipeline.preprocessing.features.jimsfeatures (module), 13

pipeline.preprocessing.features.mentalhealthfeatures (module), 14

pipeline.preprocessing.features.miscfeatures (module), 15

pipeline.preprocessing.features.person (module), 15

pipeline.preprocessing.features.rsifeatures (module), 16

pipeline.preprocessing.features.seqfeatures (module), 16

pipeline.preprocessing.run (module), 20

PrimaryImpression (class in pipeline.preprocessing.features.emsfeatures), 12

ProbationType (class in pipeline.preprocessing.features.jimsfeatures), 14

Program (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

ProgramsDischarges (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

R

13

Race (class in pipeline.preprocessing.features.person), 16

Referral (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

RefusedCareCount (class in pipeline.preprocessing.features.emsfeatures), 12

ResidencyRecordedMost (class in pipeline.preprocessing.features.rsifeatures), 16

run() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

run() (pipeline.modeling.models.Model method), 24

S

ServicesRecieved (class in pipeline.preprocessing.features.mentalhealthfeatures), 15

SimpleFeature (class in pipeline.preprocessing.features.abstract), 12

StdDaysBetweenEvents (class in pipeline.preprocessing.features.miscfeatures), 15

T

TimeBoundedFeature (class in pipeline.preprocessing.features.abstract), 12

TransportedBy (class in pipeline.preprocessing.features.rsifeatures), 16

TransportedSum (class in pipeline.preprocessing.features.emsfeatures), 13

TreatedRefusedTransportCount (class in pipeline.preprocessing.features.emsfeatures), 13

TreatedSum (class in pipeline.preprocessing.features.emsfeatures), 13

TreatedTransferredCareCount (class in pipeline.preprocessing.features.emsfeatures), 13

TreatedTransportedALSCount (class in pipeline.preprocessing.features.emsfeatures), 13

TreatedTransportedBLSCount (class in pipeline.preprocessing.features.emsfeatures), 13

TreatedTransportedByLawEnforcementCount (class in pipeline.preprocessing.features.emsfeatures), 13

TriageOfEms (class in pipeline.preprocessing.features.emsfeatures),

U

UnknownFeatureError, 12

upload_file_to_s3() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

W

write_dataframe_chunks() (in module pipeline.modeling.feature_model_grabber),

write_dataframe_to_sql() (in module pipeline.preprocessing.feature_table_builder), 20

write_matrix_pairs() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23

write_to_csv() (pipeline.modeling.feature_model_grabber.FeatureModelGrabber method), 23