

COMP9417 FraudDetection

Team members

Grace Liu (z5207875)

Kevin Luo (z5210356)

Ryan Oo (z5206676)

Andy Tang (z5209677)

Neil Yang (z5207831)

Overview

We decided to do a research prediction competition on kaggle, where we were required to predict a probability for the isFraud variable.

This repository will contain our files which we used to examine and experiment with.

Kaggle Competition:

<https://www.kaggle.com/c/ieee-fraud-detection/overview>

File setup

Prerequisites

Installation of Graphviz may be needed for some plots.

Steps to install can be found here: <https://graphviz.org/download/>

1. First clone this repository enter the root folder

```
$ git clone https://github.com/dandruffneil/COMP9417_FraudDetection.git
$ cd COMP9417_FraudDetection
```

2. Setup virtual environment

```
$ virtualenv venv
$ . venv/bin/activate
$ pip3 install -r requirements.txt
```

3. Run selected files

```
$ python3 {filename}
# Example file
$ python3 Tree.py
```

File types

FeatureSelection.py

- Provides functions which produce heatmaps and conducts feature selection on the dataset

get_data.py

- Provides functions to preprocess the dataset
- readData(scale=False, all=False, filtered=False)
 - scale = configures whether the features are normalised
 - all = configures whether all the data was used, otherwise 10% of data was used with a 4:1 (train:test) split
 - filtered = configures whether the selected features are used

LogisticRegression.py/Tree.py/.../XGB.py/LGBM.py

- Provides the functions to conduct randomsearch and gridsearch on the dataset to find the best parameter
- Some files (Tree.py, RandomForest.py, XGB.py, LGBM.py) provide functions to produce plots and generate the csv for submitting to kaggle

Due Date: Sunday, 1st August, 11:59PM