

Hereby the index of the different files with the code for the project and showed in the questions answered file, and where the different people of Enron is mention along the code files

1. Files index

PML_1_data_adq.ipynb

1. Data acquisition: import data, first view of data and curious findings
 - 1.1. Dataframe creation
 - 1.2. Dataframe overview
 - 1.3. NaN poi and non-poi
 - 1.4. Nan in Total Payment / Total Stock
 - 1.5. Nan in email address
 - 1.6. Not valid entries
 - 1.7. Data checking and correction
 - 1.8. Top Nan entries
 - 1.9. Nan per entry: who is who

PML_2_data_chc.ipynb

2. Data check: cleaning NaN, visualize data and check outliers
 - 2.1. Data set after cleaning
 - 2.2. Statistic for poi and non-poi
 - 2.3. Financial features: Data visualization, Outlier and low number of entries
 - 2.4. Email features: Data visualization and Outlier
 - 2.5. Dataset overview with and without specific entries

PML_3_Features_analysis_creation_out.ipynb

3. Features: analyses, create and chose features
 - 3.1. Features correlation overview
 - 3.2. Features detailed view with and without outliers
 - 3.3. Outlier champions
 - 3.4. Features Poi/non poi shadows
 - 3.5. Features creation
 - 3.6. Scaler with Robust scaler

PML_3a_Scaler_features.ipynb

- 3.a. Features: Create, Scale and select features
 - 3.a.1. Create new features
 - 3.a.2. Features selection before scaling
 - 3.a.3. Scaling with RobustScaler
 - 3.a.4. Features selection after Robust scaler using pipeline
 - 3.a.5. Scaling with MinMax scaler
 - 3.a.6. Features selection after MinMax scaler using pipeline
 - 3.a.7. Scaling with Normalizer scaler
 - 3.a.8. Features selection after Normalizer scaler using pipeline

PML_3b_Scaler_features_low_amount_entries.ipynb

- 3.b. Scaler behavior on features with low amount of entries
 - 3.b.1. Data frame creation with features with low amount of entries and big amount of entries
 - 3.b.2. MinMaxScaler Scaler result
 - 3.b.3. Robust Scaler result
 - 3.b.3. Normalize Scaler result

PML_4_Features_algorithm_finetune.ipynb

4. Features and algorithm to maximize the Poi identification
 - 4.0. Open file with scaled features (from PML_3) and define features lists
 - 4.1. Comparing classifier with different features
 - 4.1.1. Comparing classifiers with original features
 - 4.1.2. Comparing classifiers with original and created features
 - 4.2. Function for algorithm
 - 4.3. Random forest and original features

- 4.3.1. Random forest RState=42 and original features to choose using pipeline
- 4.3.2. Random forest and GridSearchCV to fine-tune with chosen features
- 4.3.3. Random forest fine-tune with chosen features
- 4.4. ADABOOST and new features
 - 4.4.1. ADABOOST and new features to choose
 - 4.4.2. ADABOOST and GridSearchCV to fine-tune with chosen features
- 4.5. Random forest RState=42 and all features
 - 4.5.1. Random forest RState=42 and new features to choose
 - 4.5.2. Random forest finetune to choose features
- 4.6. AdaBoost with Random forest
 - 4.6.1. AdaBoost with Random forest: original features
 - 4.6.2. AdaBoost with Random forest: all features
 - 4.6.3. ADABOOST with Random forest and GridSearchCV to tune with chosen features
- 4.7. Algorithm and features chosen for tester

PML_5_Classifier for submission

- 5. Testing classifier with tester.py tester functionFinal_for_submission_Phyton_3
 - 5.1. Comparing classifiers with different features
 - 5.2. Better classifier for submission

2. Entries described and in which file

1.4
 CHAN RONNIE
 POWERS WILLIAM
 LOCKHART EUGENE E
 1.6
 'THE TRAVEL AGENCY IN THE PARK'
 'TOTAL'
 1.7
 BELFER ROBERT
 BHATNAGAR SANJAY
 1.8.
 GRAMM WENDY L
 WHALEY DAVID A
 WROBEL BRUCE
 WAKEHAM JOHN
 1.9.
 SAVAGE FRANK
 BLAKE JR. NORMAN
 WINOKUR JR. HERBERT
 MENDELSON JOHN
 MEYER JEROME J
 PEREIRA PAULO V. FERRAZ
 URQUHART JOHN A
 CLINE KENNETH W
 WAKEHAM JOHN
 WODRASKA JOHN
 SCRIMSHAW MATTHEW

2.1 Remove of
 'THE TRAVEL AGENCY IN THE PARK'

'TOTAL'

BELFER ROBERT

BHATNAGAR SANJAY

2.3

LAY KENNETH L

2.4

KAMINSKI WINCENTY J

SHAPIRO RICHARD S

DELAINEY DAVID W

LAVORATO JOHN J

KEAN STEVEN J

BECK SALLY W

KITCHEN LOUISE

DIETRICH JANET R

FASTOW ANDREW S

KOPPER MICHAEL J

YEAGER F SCOTT

HIRKO JOSEPH

3.4

TIMOTHY BELDEN

JEFFREY KEITH SKILLING