FLSentencingTreeOutputs

December 8, 2022

```python
import pandas as pd
import numpy as np
import os

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.tree import DecisionTreeClassifier, DecisionTreeRegressor,
 ↪plot_tree
from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor
from sklearn.metrics import classification_report, confusion_matrix,
 ↪accuracy_score, r2_score, mean_squared_error
from xgboost import XGBRegressor, XGBClassifier
import shap
from functools import reduce
```

```python
#Read-in custom function
import CleanSAO
from CleanFDOC import clean_fdoc
from CreateCCMaster import create_ccm
from CreateModelDF import model_df
from TreeModels import get_tree
```

```python
dir = os.getcwd()
path = dir+'/data/'
os.chdir(path)

#Clean charges(prosecutor specific actions) datasets (drug, theft)
drug_sa_clean, theft_sa_clean = CleanSAO.clean_sao('CjdtSAOCase_00000.csv')

#Clean sentencing (offenses) datasets (drug, theft)
drug_offenses_clean, theft_offenses_clean = clean_fdoc(['Active_Offenses_PRPR.
 ↪csv', 'Active_Offenses_CPS.csv',
                                                        'Release_Offenses_PRPR.
 ↪csv', 'Release_Offenses_CPS.csv',
                                                        'Active_Root.csv',
 ↪'Release_Root.csv'])
```

```python
#Circuit-county-year political/State Attorney dataset
ccm = create_ccm(safile = "SA_Political_Leanings.csv",
                 housefile = "clean_house.csv",
                 senatefile = "clean_senate.csv",
                 presfile = "clean_pres.csv",
                 circuitcountyfile="circuit_county_crosswalk.csv")

# Compile modeling datasets
drug_sa_df = model_df(drug_sa_clean, ccm, obstype = 'action', crimetype =
 ↪'drug')
theft_sa_df = model_df(theft_sa_clean, ccm, obstype = 'action', crimetype =
 ↪'theft')
drug_off_df = model_df(drug_offenses_clean, ccm, obstype = 'offense', crimetype
 ↪= 'drug')
theft_off_df = model_df(theft_offenses_clean, ccm, obstype = 'offense',
 ↪crimetype = 'theft')
```

0.1 CART Models

0.1.1 CART Classifier - Drug Charges

```python
[ ]: cart = {'max_depth':[3, 6, 10],
        'min_samples_split':[8, 20],
        'min_samples_leaf': [10, 20, 100]}

drug_sa_tree = get_tree(drug_sa_df, target = 'FINAL_ACTION_DESC', paramdict =
 ↪cart, model = DecisionTreeClassifier,  seed = 42)
```

```
Confusion matrix:
 [[  48  189]
 [   9 1212]] Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.20      0.33       237
           1       0.87      0.99      0.92      1221

    accuracy                           0.86      1458
   macro avg       0.85      0.60      0.63      1458
weighted avg       0.86      0.86      0.83      1458
 Accuracy: 0.8641975308641975
Feature Importance Table

    Importance                                        Features
```
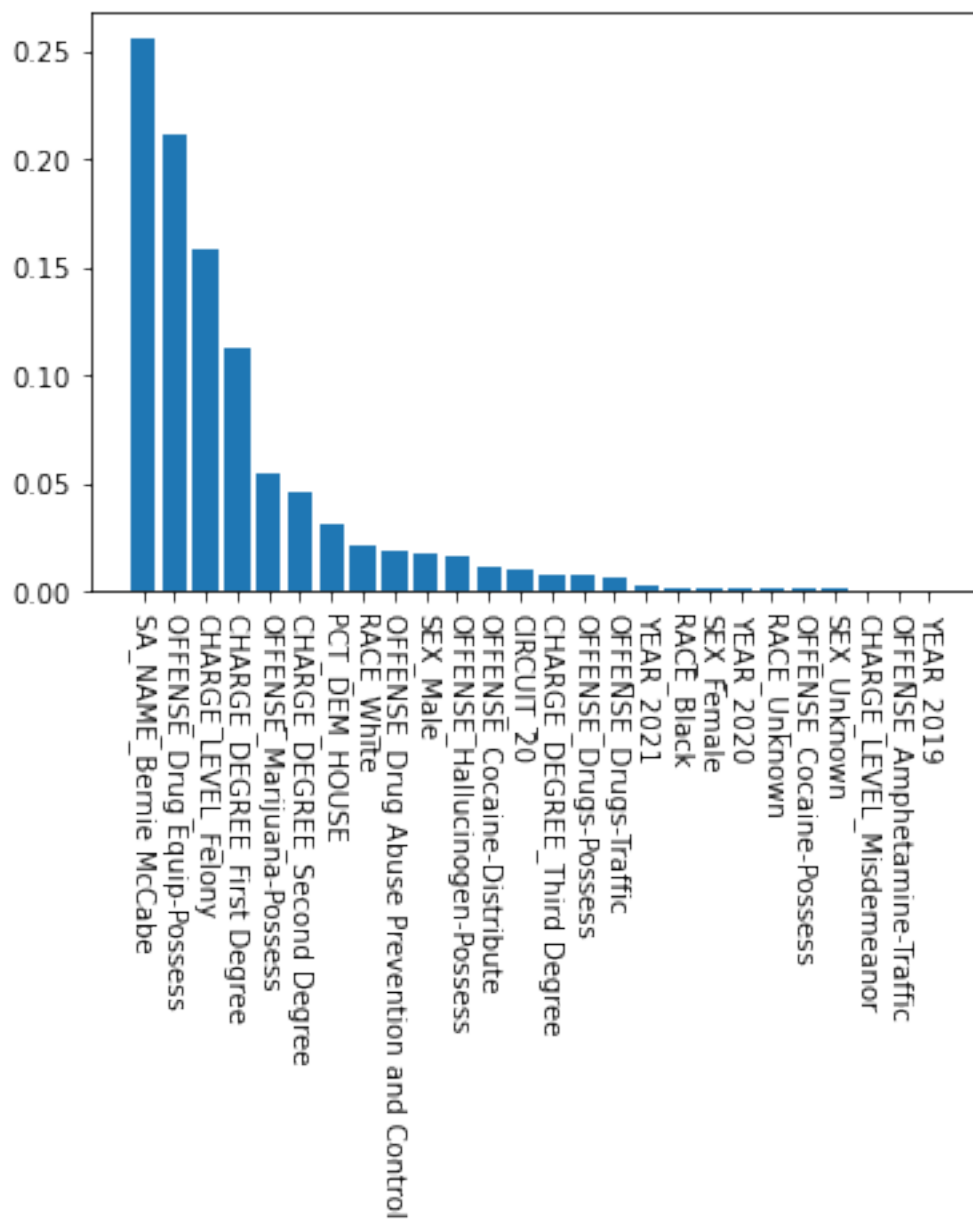
```
81      0.256043                       SA_NAME_Bernie McCabe
27      0.211825                 OFFENSE_Drug Equip-Possess
7       0.158817                        CHARGE_LEVEL_Felony
0       0.112998                  CHARGE_DEGREE_First Degree
42      0.054556                  OFFENSE_Marijuana-Possess
4       0.046151                 CHARGE_DEGREE_Second Degree
47      0.030667                             PCT_DEM_HOUSE
14      0.021412                                RACE_White
26      0.019173  OFFENSE_Drug Abuse Prevention and Control
16      0.017370                                  SEX_Male
37      0.016797               OFFENSE_Hallucinogen-Possess
21      0.011390                  OFFENSE_Cocaine-Distribute
78      0.009708                                 CIRCUIT_20
5       0.007655                  CHARGE_DEGREE_Third Degree
31      0.007269                      OFFENSE_Drugs-Possess
34      0.006398                       OFFENSE_Drugs-Traffic
72      0.002513                                 YEAR_2021
12      0.002011                                RACE_Black
15      0.001558                                SEX_Female
71      0.001462                                 YEAR_2020
13      0.001249                              RACE_Unknown
22      0.001092                    OFFENSE_Cocaine-Possess
17      0.000968                               SEX_Unknown
8       0.000522                   CHARGE_LEVEL_Misdemeanor
19      0.000250                 OFFENSE_Amphetamine-Traffic
70      0.000144                                 YEAR_2019
```
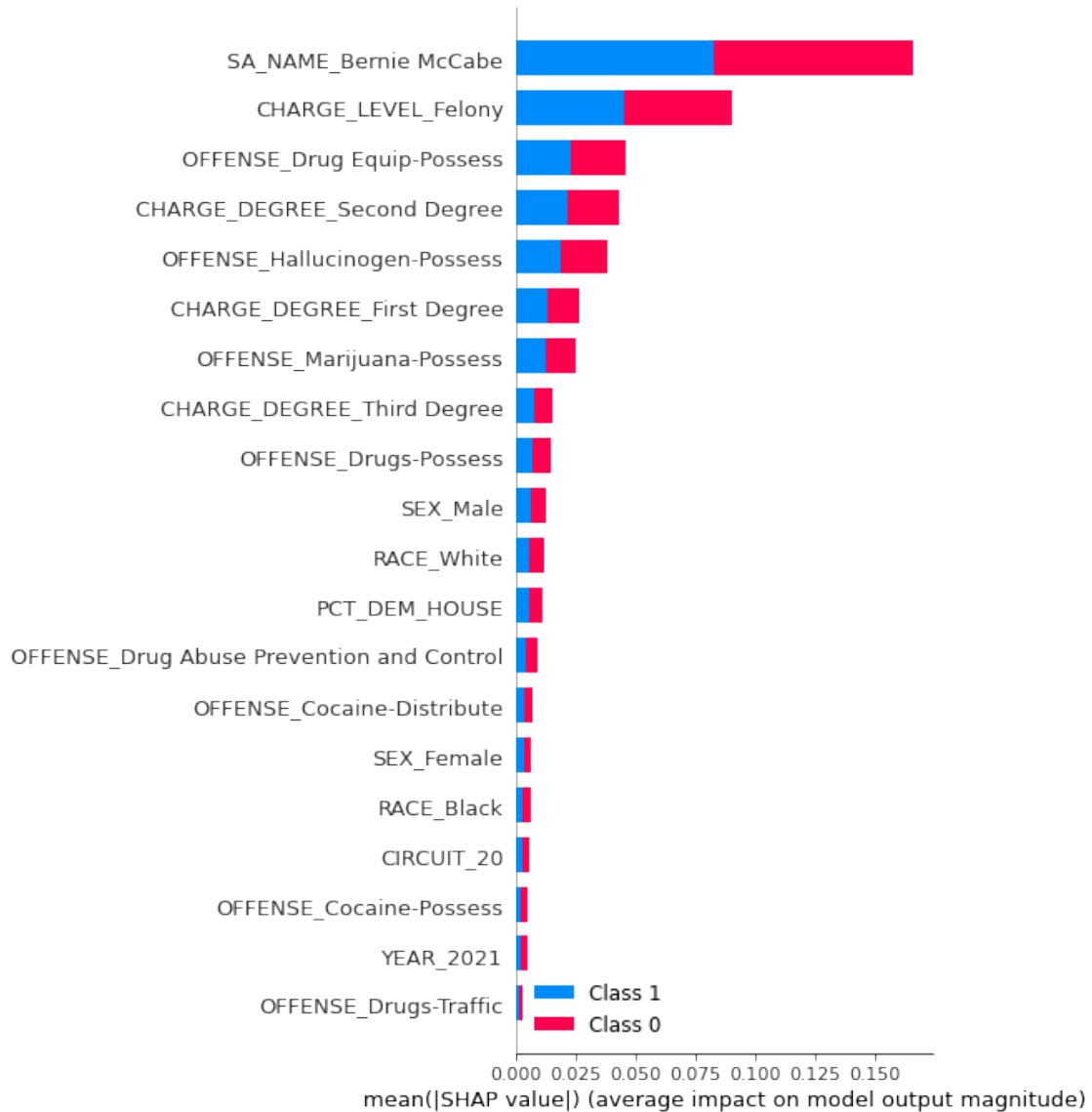
Histogram of Feature Importance

Feature Beeswarm Plot

SA_NAME_Bernie McCabe
CHARGE_LEVEL_Felony
OFFENSE_Drug Equip-Possess
CHARGE_DEGREE_Second Degree
OFFENSE_Hallucinogen-Possess
CHARGE_DEGREE_First Degree
OFFENSE_Marijuana-Possess
CHARGE_DEGREE_Third Degree
OFFENSE_Drugs-Possess
SEX_Male
RACE_White
PCT_DEM_HOUSE
OFFENSE_Drug Abuse Prevention and Control
OFFENSE_Cocaine-Distribute
SEX_Female
RACE_Black
CIRCUIT_20
OFFENSE_Cocaine-Possess
YEAR_2021
OFFENSE_Drugs-Traffic

Class 1
Class 0

0.000 0.025 0.050 0.075 0.100 0.125 0.150
mean(|SHAP value|) (average impact on model output magnitude)

0.1.2 CART Classifier - Theft Charges

```
theft_sa_tree = get_tree(theft_sa_df, target = 'FINAL_ACTION_DESC', paramdict =␣
 ↪cart, model = DecisionTreeClassifier, seed = 42)
```

```
Confusion matrix:
 [[  5  98]
 [  2 739]] Classification Report:
              precision    recall  f1-score   support

           0       0.71      0.05      0.09       103
           1       0.88      1.00      0.94       741
```

```
       accuracy                                0.88       844
      macro avg        0.80       0.52         0.51       844
   weighted avg        0.86       0.88         0.83       844
 Accuracy: 0.8815165876777251
```
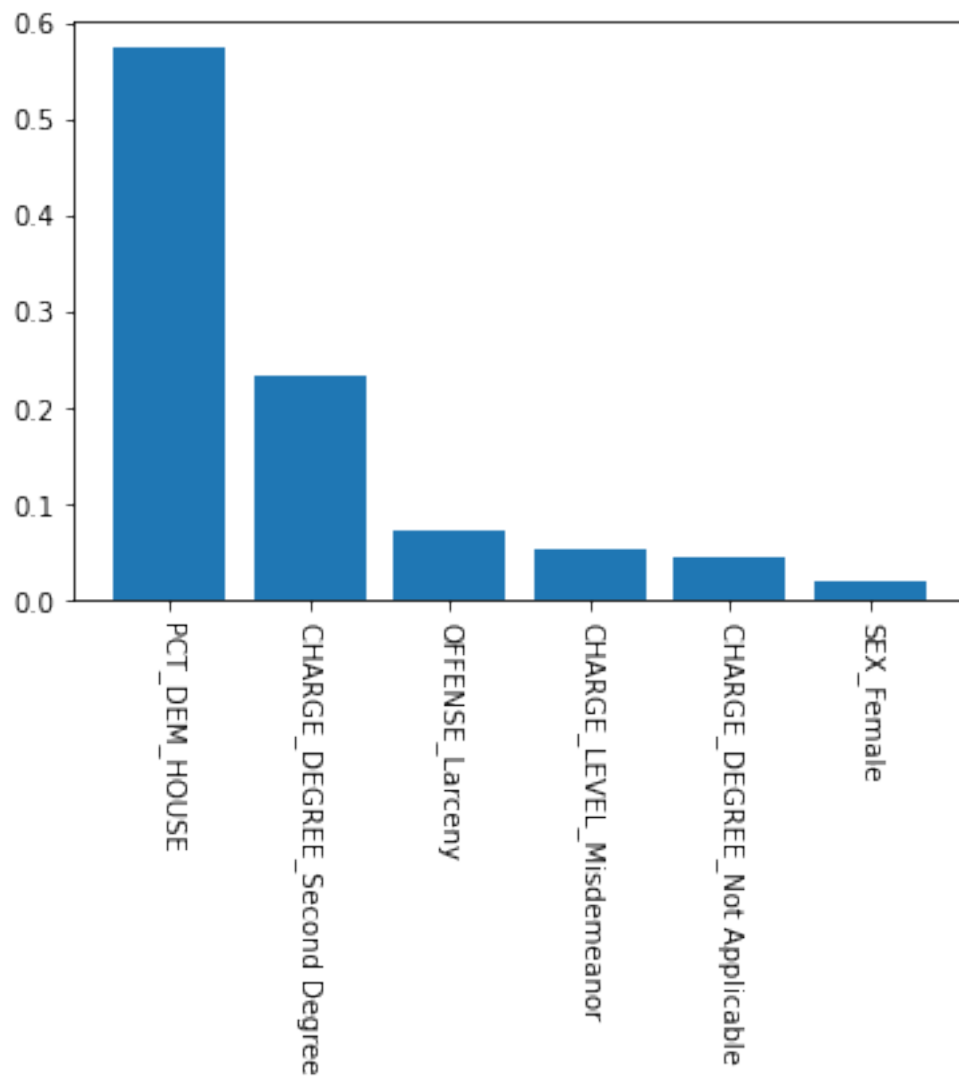
Feature Importance Table

```
      Importance                        Features
30     0.573790                    PCT_DEM_HOUSE
4      0.235240     CHARGE_DEGREE_Second Degree
22     0.071874                  OFFENSE_Larceny
8      0.054824        CHARGE_LEVEL_Misdemeanor
3      0.044836  CHARGE_DEGREE_Not Applicable
16     0.019436                      SEX_Female
```
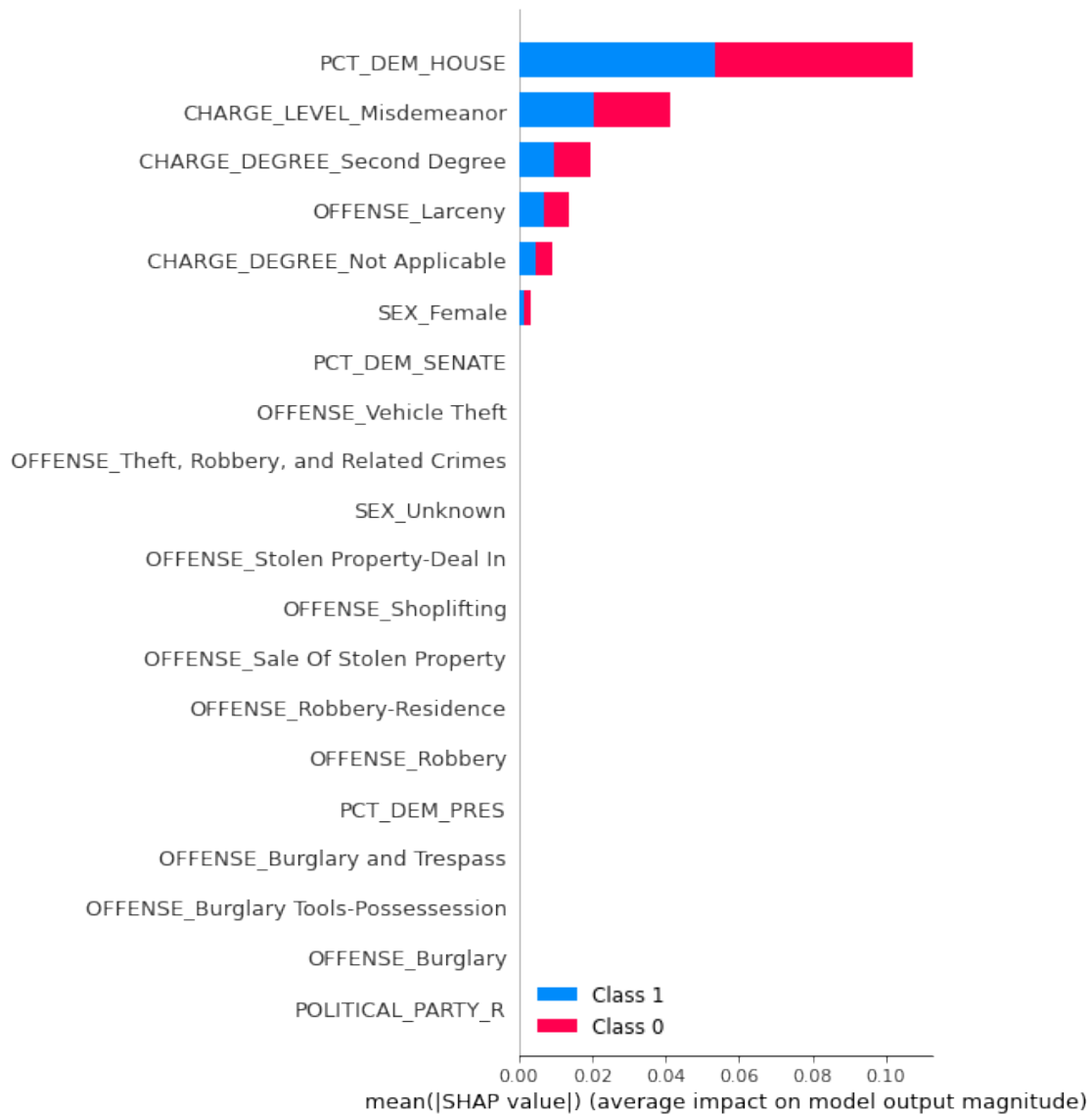
Histogram of Feature Importance

Feature Beeswarm Plot

0.1.3 CART Regressor - Drug Sentencing

```
drug_off_tree = get_tree(drug_off_df, target = 'TERM_YEARS', paramdict = cart,␣
↪model = DecisionTreeRegressor, seed = 42)
```

train RMSE: 3.095397064432317
test RMSE: 3.1896629709326603
Feature Importance Table

      Importance   Features

3	0.375679	POSSESSION
7	0.118662	NON-PRESCRIPTION_OPIOIDS
2	0.104234	AGE
4	0.063408	MARIJUANA
21	0.049927	SEX_M
24	0.048151	PCT_DEM_PRES
22	0.029917	PCT_DEM_HOUSE
23	0.025253	PCT_DEM_SENATE
119	0.018972	CIRCUIT_13.0
94	0.018060	COUNTY_VOLUSIA
5	0.016801	COCAINE
0	0.016068	CHARGE_COUNT
20	0.013773	SEX_F
116	0.011996	CIRCUIT_9.0
1	0.011821	PRIOR_ADJ
15	0.008819	RACE_B
115	0.007049	CIRCUIT_8.0
134	0.006995	SA_NAME_Brad King
111	0.005518	CIRCUIT_4.0
104	0.004770	YEAR_2018
19	0.003860	RACE_W
110	0.003852	CIRCUIT_3.0
132	0.003684	SA_NAME_Bill Cervone
131	0.003264	SA_NAME_Bernie McCabe
114	0.003075	CIRCUIT_7.0
83	0.002909	COUNTY_POLK
78	0.002434	COUNTY_OSCEOLA
87	0.002334	COUNTY_SEMINOLE
117	0.002286	CIRCUIT_10.0
126	0.002246	CIRCUIT_20.0
27	0.002173	DRUG_POSSESSION
125	0.002090	CIRCUIT_19.0
157	0.001581	SA_NAME_Mike Satz
143	0.001463	SA_NAME_Ed Brodsky
107	0.001230	YEAR_2021
56	0.001229	COUNTY_HILLSBOROUGH
33	0.000971	COUNTY_BREVARD
113	0.000913	CIRCUIT_6.0
8	0.000852	METH
31	0.000633	COUNTY_BAY
163	0.000584	SA_NAME_Steve Russell
64	0.000260	COUNTY_LEE
98	0.000202	YEAR_2012

Histogram of Feature Importance

Feature Beeswarm Plot
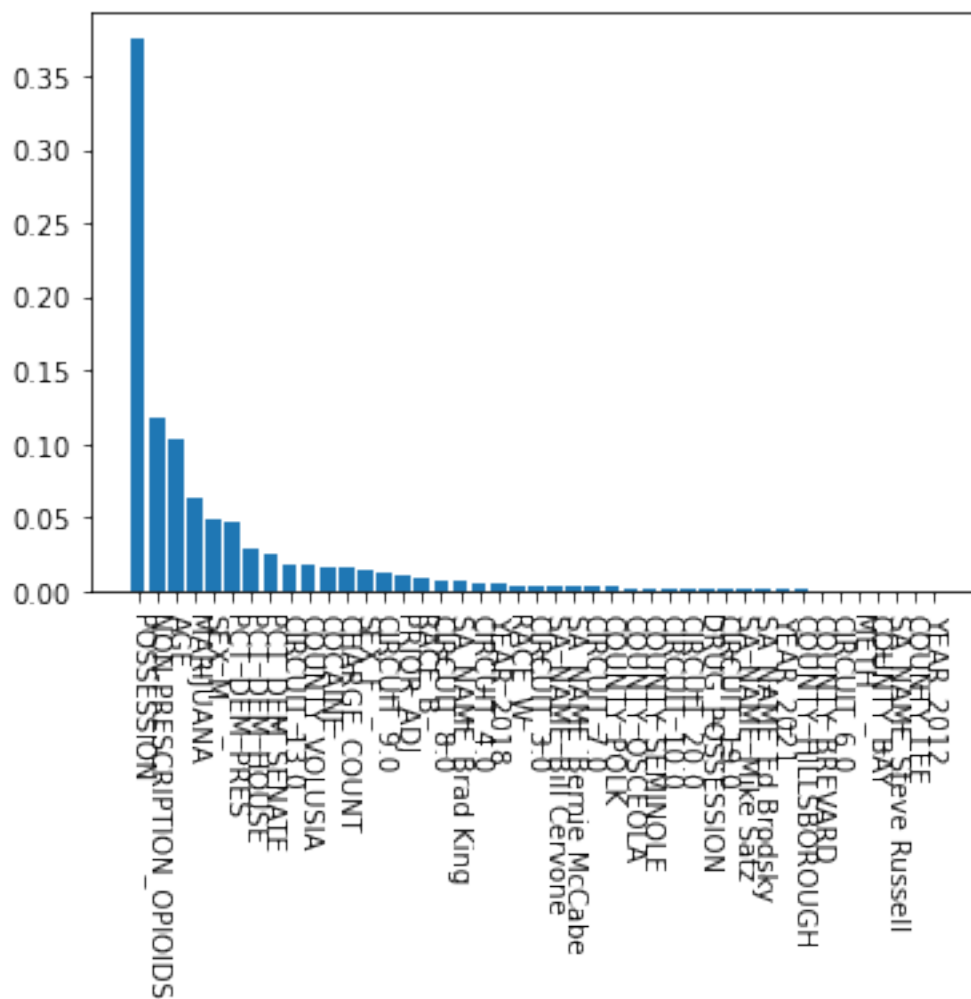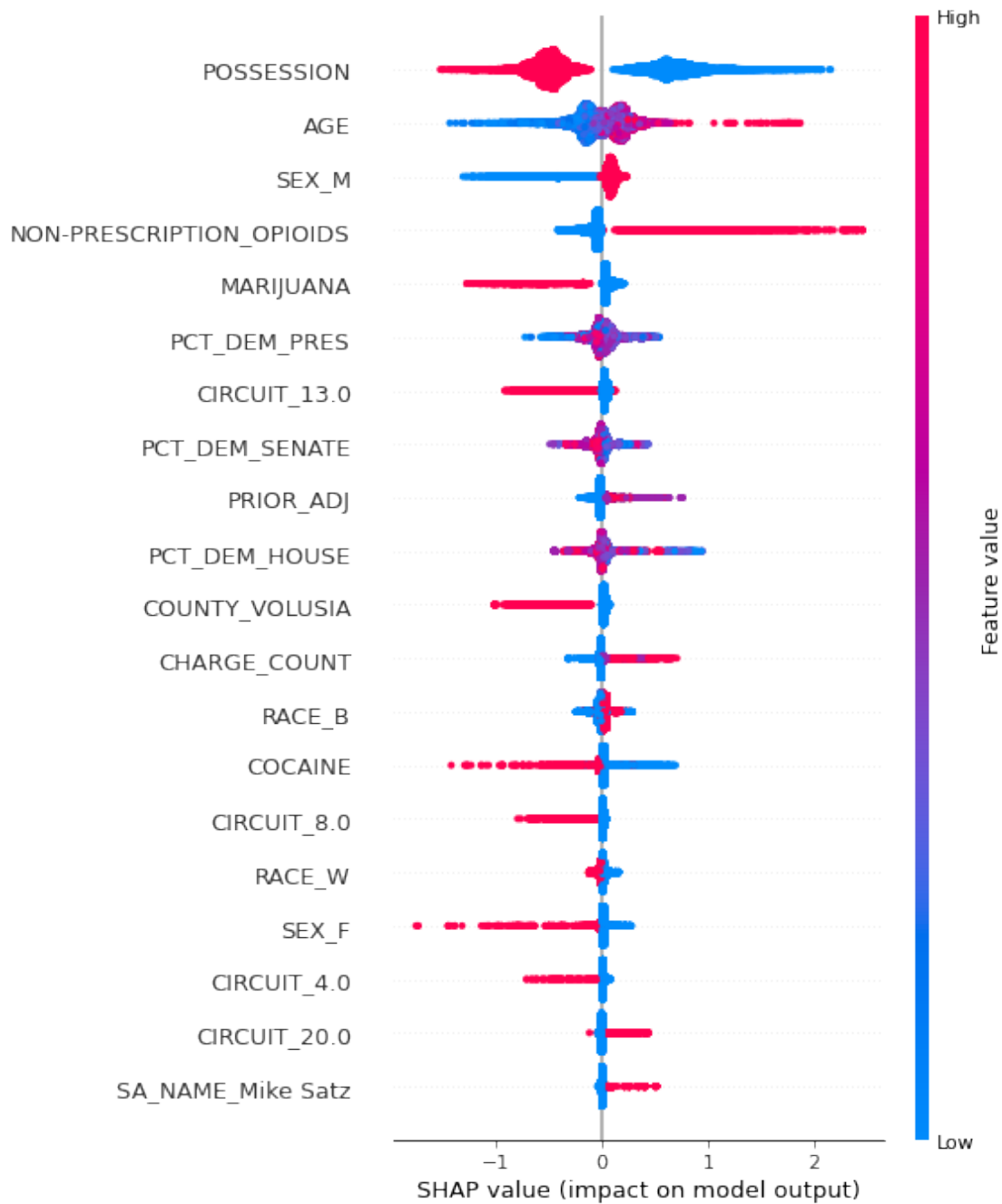
0.1.4 CART Regressor - Theft Sentencing

```
theft_off_tree = get_tree(theft_off_df, 'TERM_YEARS', paramdict = cart, model =␣
↪DecisionTreeRegressor, seed = 42)
```

train RMSE: 3.9170554956791563
test RMSE: 3.985657882448129

```
Feature Importance Table

     Importance   Features
8   0.363025   ROBBERY
16   0.146809   NON-DEADLY_WEAPON
2   0.110107   AGE
15   0.093605   NO_WEAPON
14   0.085255   DEADLY_WEAPON
0   0.042273   CHARGE_COUNT
27   0.023322   PCT_DEM_PRES
3   0.018416   BURGLARY
26   0.016028   PCT_DEM_SENATE
23   0.015574   SEX_F
1   0.012695   PRIOR_ADJ
25   0.012039   PCT_DEM_HOUSE
22   0.012002   RACE_W
9   0.007724   PROPERTY_THEFT
24   0.006086   SEX_M
18   0.005867   RACE_B
59   0.005625   COUNTY_HILLSBOROUGH
46   0.004007   COUNTY_DUVAL
6   0.003697   PETIT_THEFT
29   0.002288   PETTY_THEFT
37   0.001969   COUNTY_BROWARD
117   0.001389   CIRCUIT_7.0
126   0.001385   CIRCUIT_17.0
122   0.000935   CIRCUIT_13.0
4   0.000908   GRAND THEFT
116   0.000889   CIRCUIT_6.0
164   0.000821   SA_NAME_R. J. Larizza
172   0.000795   POLITICAL_PARTY_R
118   0.000747   CIRCUIT_8.0
28   0.000654   DEMOCRAT_YN
131   0.000509   SA_NAME_Andrew Warren
108   0.000429   YEAR_2019
102   0.000365   YEAR_2013
31   0.000349   NONVIOLENT_CRIME
160   0.000269   SA_NAME_Mike Satz
107   0.000261   YEAR_2018
83   0.000234   COUNTY_PALM BEACH
120   0.000230   CIRCUIT_10.0
135   0.000206   SA_NAME_Bill Cervone
105   0.000140   YEAR_2016
104   0.000074   YEAR_2015

Histogram of Feature Importance
```

Feature Beeswarm Plot
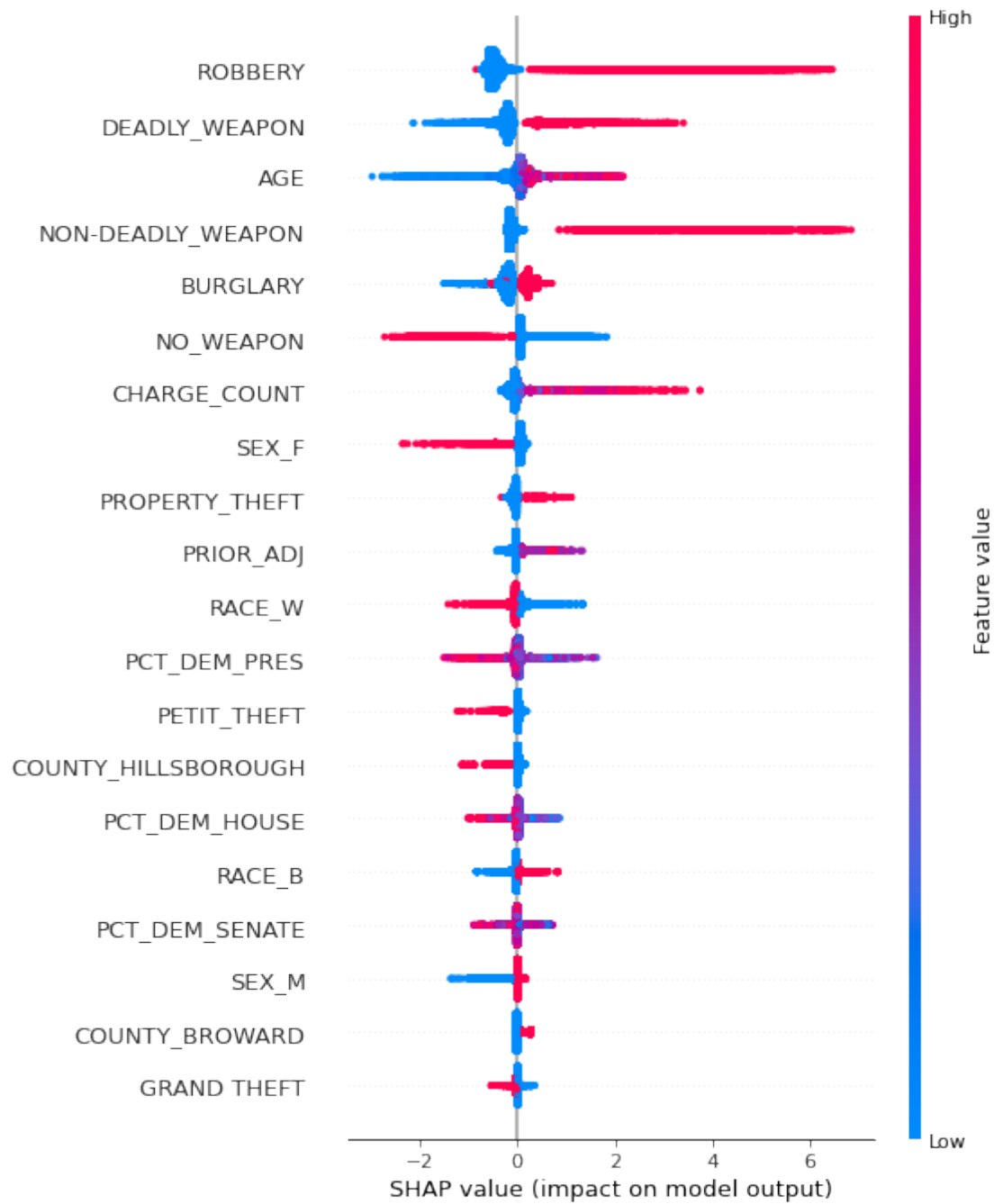
## 0.2 Random Forest Models

### 0.2.1 RF Classifier - Drug Charges

```python
#RF hyperarameter space
bag = {'max_depth':[10, 15],
       'min_samples_leaf': [100, 200],
       'max_features': [40, 80],
       'n_estimators':[15, 20]}

drug_sa_tree = get_tree(drug_sa_df, target = 'FINAL_ACTION_DESC', paramdict =
 ↪bag, model = RandomForestClassifier, seed = 42)
```

```
Confusion matrix:
 [[  45  192]
 [  16 1205]] Classification Report:
              precision    recall  f1-score   support

           0       0.74      0.19      0.30       237
           1       0.86      0.99      0.92      1221

    accuracy                           0.86      1458
   macro avg       0.80      0.59      0.61      1458
weighted avg       0.84      0.86      0.82      1458
 Accuracy: 0.8573388203017832
Feature Importance Table

     Importance                          Features
7      0.276330            CHARGE_LEVEL_Felony
27     0.202035      OFFENSE_Drug Equip-Possess
62     0.127718               COUNTY_PINELLAS
75     0.100969                      CIRCUIT_6
81     0.067022          SA_NAME_Bernie McCabe
8      0.066094        CHARGE_LEVEL_Misdemeanor
42     0.044568        OFFENSE_Marijuana-Possess
49     0.023584                   PCT_DEM_PRES
37     0.019727  OFFENSE_Hallucinogen-Possess
14     0.018944                     RACE_White
51     0.014862                    PETTY_THEFT
31     0.014425           OFFENSE_Drugs-Possess
47     0.004961                  PCT_DEM_HOUSE
72     0.004166                      YEAR_2021
5      0.003419    CHARGE_DEGREE_Third Degree
22     0.003287        OFFENSE_Cocaine-Possess
17     0.002905                    SEX_Unknown
16     0.001576                       SEX_Male
0      0.001502    CHARGE_DEGREE_First Degree
12     0.000910                     RACE_Black
52     0.000754                DRUG_POSSESSION
```
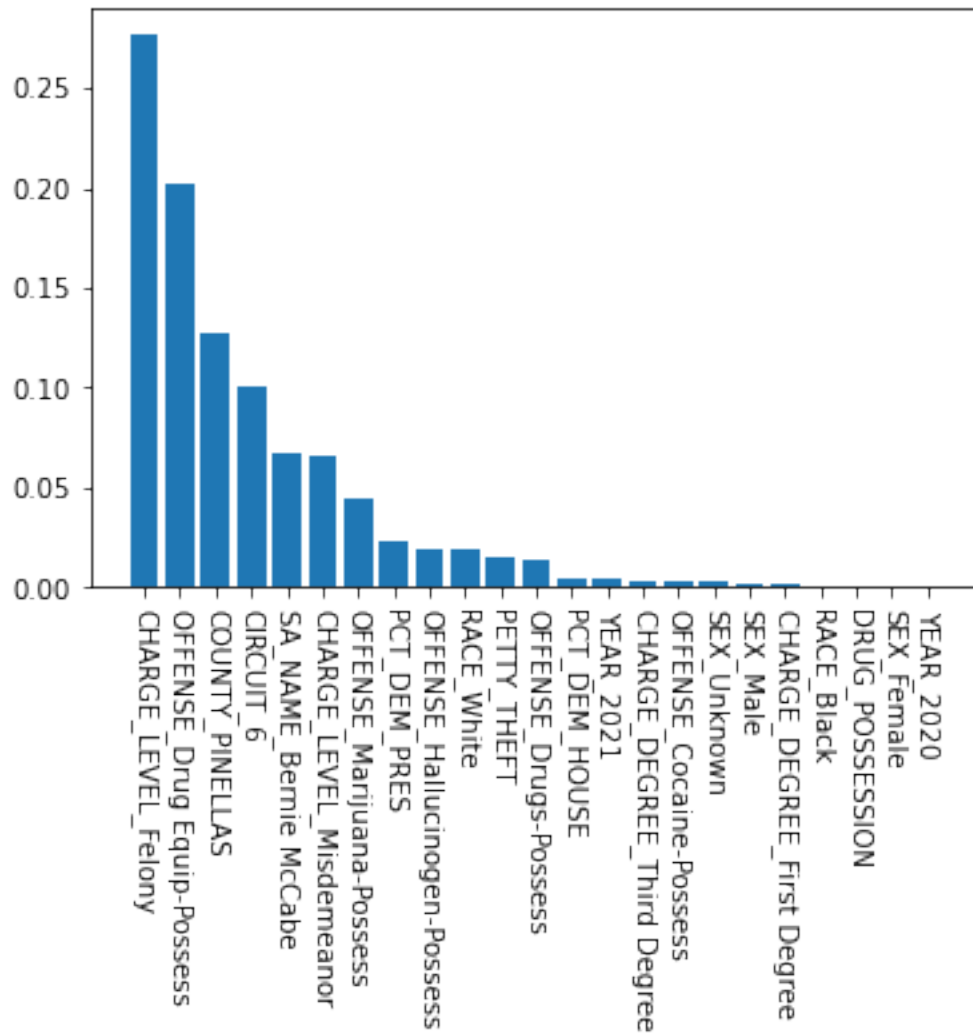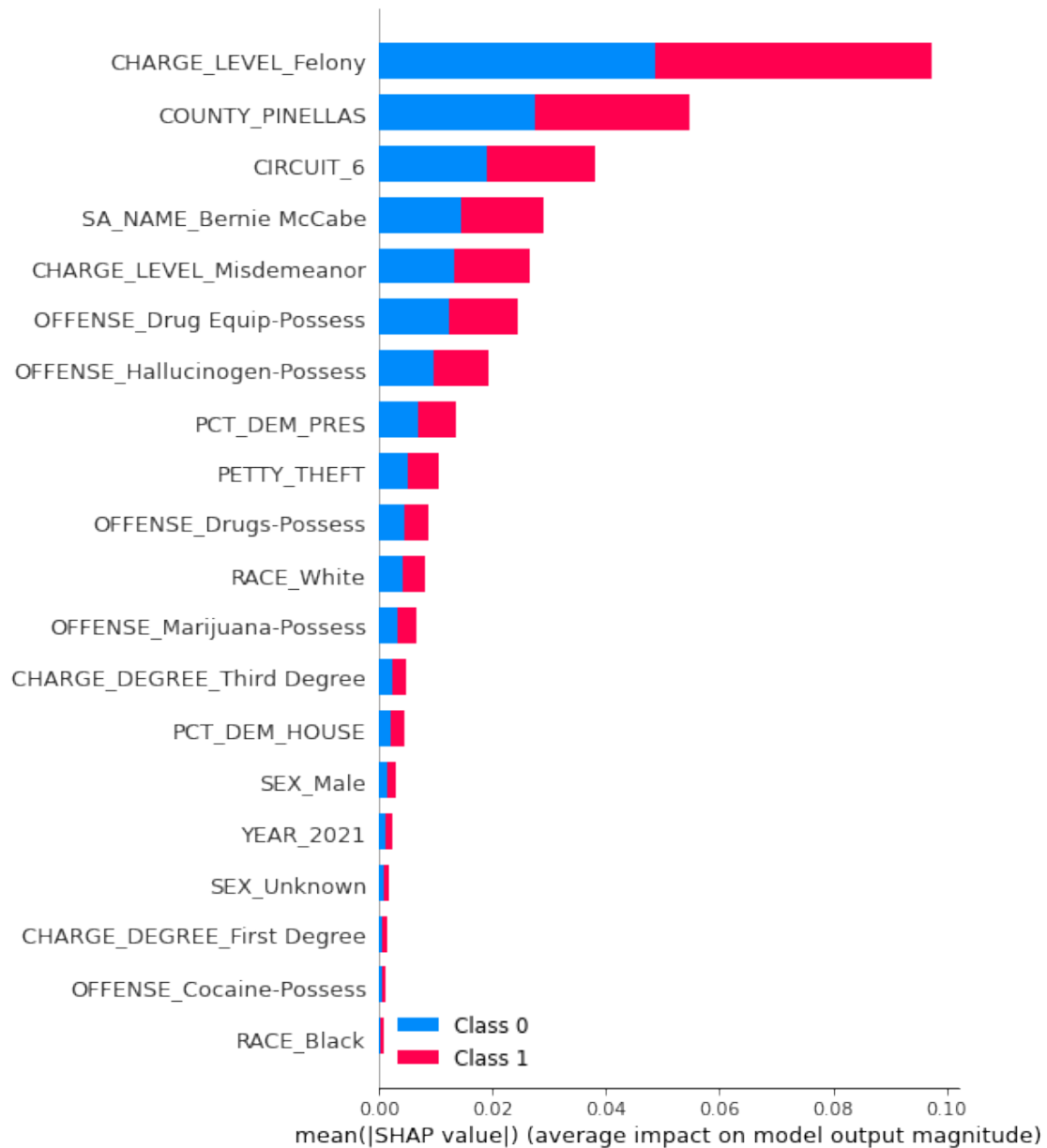
```
15    0.000132                    SEX_Female
71    0.000109                    YEAR_2020
```

Histogram of Feature Importance

Feature Beeswarm Plot

Legend:
- Class 0 (blue)
- Class 1 (red)

x-axis: mean(|SHAP value|) (average impact on model output magnitude)

Features (top to bottom):
CHARGE_LEVEL_Felony
COUNTY_PINELLAS
CIRCUIT_6
SA_NAME_Bernie McCabe
CHARGE_LEVEL_Misdemeanor
OFFENSE_Drug Equip-Possess
OFFENSE_Hallucinogen-Possess
PCT_DEM_PRES
PETTY_THEFT
OFFENSE_Drugs-Possess
RACE_White
OFFENSE_Marijuana-Possess
CHARGE_DEGREE_Third Degree
PCT_DEM_HOUSE
SEX_Male
YEAR_2021
SEX_Unknown
CHARGE_DEGREE_First Degree
OFFENSE_Cocaine-Possess
RACE_Black

0.2.2 RF Classifier - Theft Charges

```
theft_sa_tree = get_tree(theft_sa_df, target = 'FINAL_ACTION_DESC', paramdict =␣
 ↪bag, model = RandomForestClassifier, seed = 42)
```

40 fits failed out of a total of 80.
The score on these train-test partitions for these parameters will be set to
nan.
If these failures are not expected, you can try to debug them by setting

16

error_score='raise'.

Below are more details about the failures:
----
40 fits failed with the following error:
Traceback (most recent call last):
  File "c:\ProgramData\Anaconda3\lib\site-
packages\sklearn\model_selection_validation.py", line 680, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "c:\ProgramData\Anaconda3\lib\site-packages\sklearn\ensemble_forest.py",
line 450, in fit
    trees = Parallel(
  File "c:\ProgramData\Anaconda3\lib\site-packages\joblib\parallel.py", line
1043, in __call__
    if self.dispatch_one_batch(iterator):
  File "c:\ProgramData\Anaconda3\lib\site-packages\joblib\parallel.py", line
861, in dispatch_one_batch
    self._dispatch(tasks)
  File "c:\ProgramData\Anaconda3\lib\site-packages\joblib\parallel.py", line
779, in _dispatch
    job = self._backend.apply_async(batch, callback=cb)
  File "c:\ProgramData\Anaconda3\lib\site-
packages\joblib_parallel_backends.py", line 208, in apply_async
    result = ImmediateResult(func)
  File "c:\ProgramData\Anaconda3\lib\site-
packages\joblib_parallel_backends.py", line 572, in __init__
    self.results = batch()
  File "c:\ProgramData\Anaconda3\lib\site-packages\joblib\parallel.py", line
262, in __call__
    return [func(*args, **kwargs)
  File "c:\ProgramData\Anaconda3\lib\site-packages\joblib\parallel.py", line
262, in <listcomp>
    return [func(*args, **kwargs)
  File "c:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\fixes.py", line
216, in __call__
    return self.function(*args, **kwargs)
  File "c:\ProgramData\Anaconda3\lib\site-packages\sklearn\ensemble_forest.py",
line 185, in _parallel_build_trees
    tree.fit(X, y, sample_weight=curr_sample_weight, check_input=False)
  File "c:\ProgramData\Anaconda3\lib\site-packages\sklearn\tree_classes.py",
line 937, in fit
    super().fit(
  File "c:\ProgramData\Anaconda3\lib\site-packages\sklearn\tree_classes.py",
line 308, in fit
    raise ValueError("max_features must be in (0, n_features]")
ValueError: max_features must be in (0, n_features]

One or more of the test scores are non-finite: [0.8845972 0.8845972 0.8845972

17

```
0.8845972        nan        nan        nan
        nan 0.8845972 0.8845972 0.8845972 0.8845972        nan        nan
        nan        nan]
```

Confusion matrix:
```
 [[  0 103]
 [  0 741]] Classification Report:
          precision   recall  f1-score   support

         0      0.00     0.00      0.00       103
         1      0.88     1.00      0.94       741

   accuracy                        0.88       844
  macro avg      0.44     0.50      0.47       844
weighted avg      0.77     0.88      0.82       844
 Accuracy: 0.8779620853080569
```
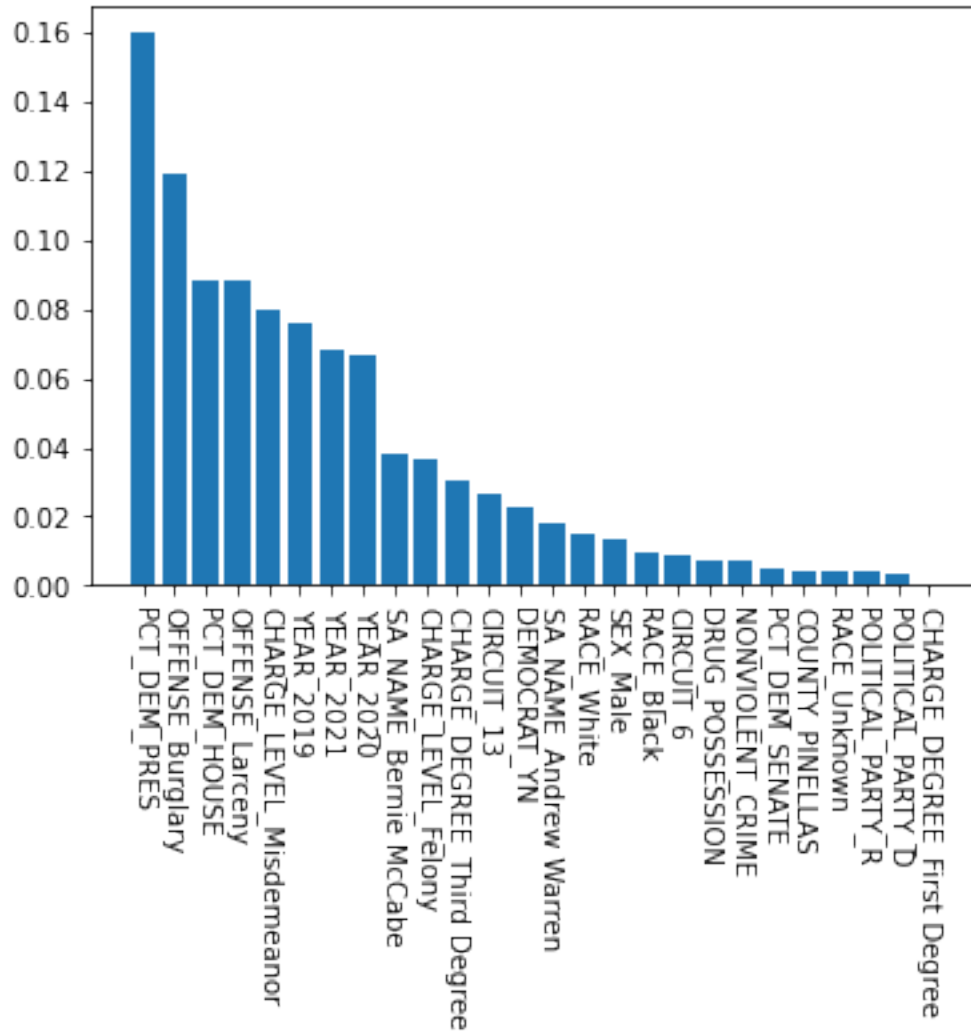Feature Importance Table

Precision and F-score are ill-defined and being set to 0.0 in labels with no
predicted samples. Use `zero_division` parameter to control this behavior.
Precision and F-score are ill-defined and being set to 0.0 in labels with no
predicted samples. Use `zero_division` parameter to control this behavior.
Precision and F-score are ill-defined and being set to 0.0 in labels with no
predicted samples. Use `zero_division` parameter to control this behavior.

	Importance	Features
32	0.159540	PCT_DEM_PRES
19	0.118935	OFFENSE_Burglary
30	0.088298	PCT_DEM_HOUSE
22	0.087798	OFFENSE_Larceny
8	0.079552	CHARGE_LEVEL_Misdemeanor
53	0.075724	YEAR_2019
55	0.068065	YEAR_2021
54	0.066693	YEAR_2020
63	0.037704	SA_NAME_Bernie McCabe
7	0.036427	CHARGE_LEVEL_Felony
5	0.030447	CHARGE_DEGREE_Third Degree
59	0.026219	CIRCUIT_13
33	0.023023	DEMOCRAT_YN
62	0.017839	SA_NAME_Andrew Warren
15	0.015255	RACE_White
17	0.013704	SEX_Male
12	0.009772	RACE_Black
57	0.009028	CIRCUIT_6
35	0.007316	DRUG_POSSESSION
36	0.007040	NONVIOLENT_CRIME
31	0.004664	PCT_DEM_SENATE
44	0.004534	COUNTY_PINELLAS
14	0.004490	RACE_Unknown

```
69    0.004409          POLITICAL_PARTY_R
68    0.003357          POLITICAL_PARTY_D
0     0.000168  CHARGE_DEGREE_First Degree
```

Histogram of Feature Importance

Feature Beeswarm Plot
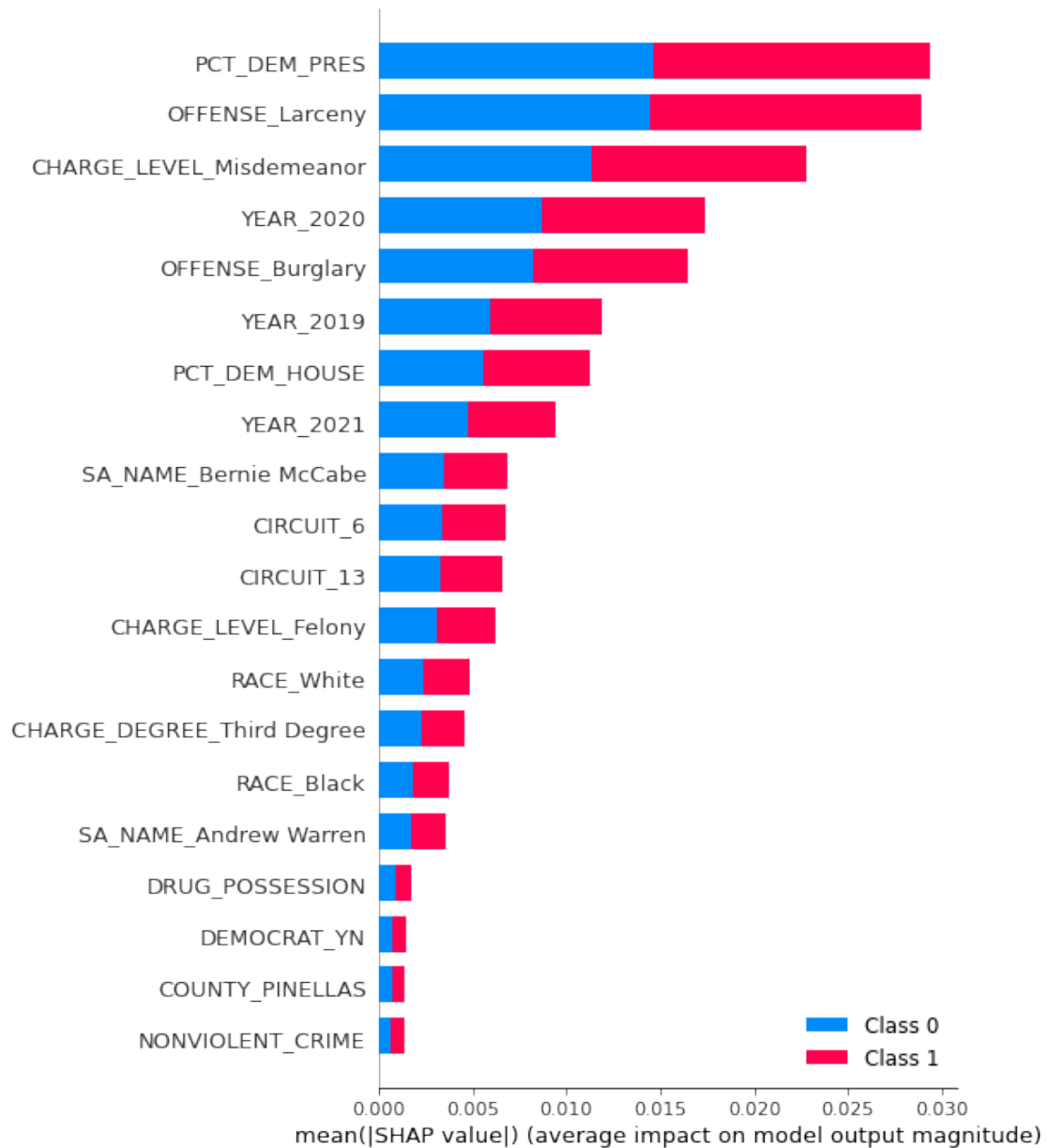
0.2.3 RF Regressor - Drug Charges

```
drug_off_tree = get_tree(drug_off_df, target = 'TERM_YEARS', paramdict = bag,
↪model = RandomForestRegressor, seed = 42)
```

```
train RMSE: 3.0750664907114205
test RMSE: 3.1701806859900636
Feature Importance Table
```
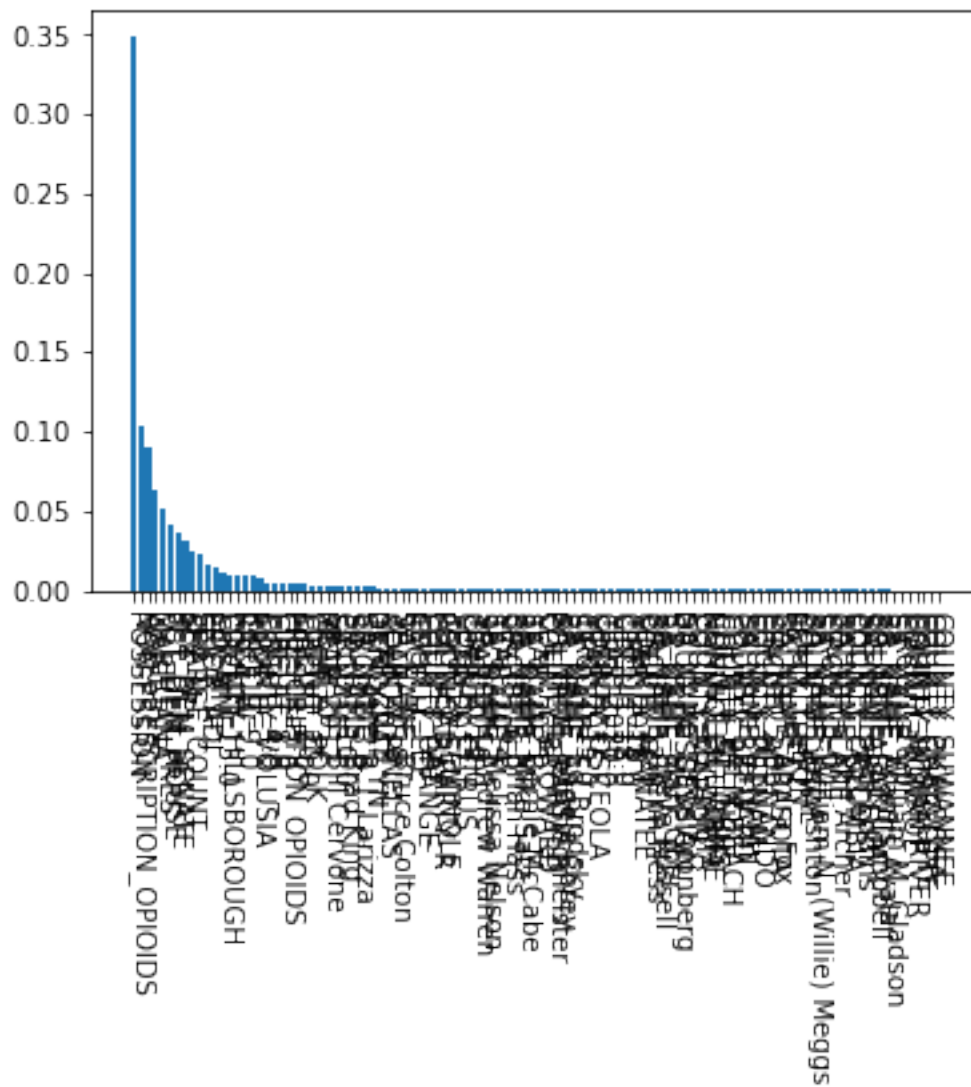
	Importance	Features
3	0.348333	POSSESSION

```
7       0.103840   NON-PRESCRIPTION_OPIOIDS
2       0.090793                        AGE
4       0.063462                  MARIJUANA
24      0.050611               PCT_DEM_PRES
..         …                           …
73      0.000046              COUNTY_MONROE
58      0.000034         COUNTY_INDIAN RIVER
36      0.000029           COUNTY_CHARLOTTE
75      0.000025           COUNTY_OKALOOSA
91      0.000007            COUNTY_SUWANNEE

[109 rows x 2 columns]
```
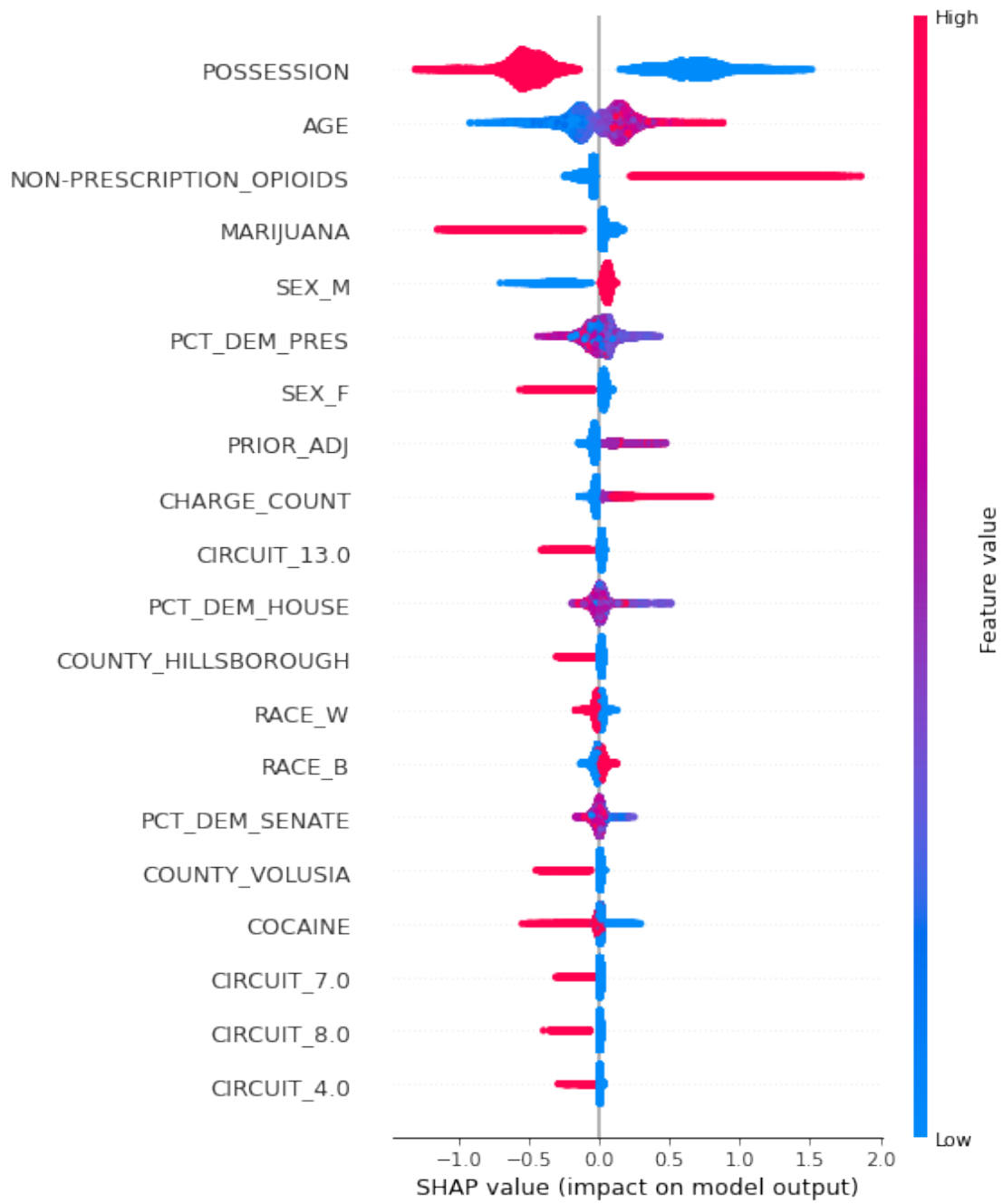
Histogram of Feature Importance

Feature Beeswarm Plot

0.2.4 RF Regressor - Theft Charges

```
theft_off_tree = get_tree(theft_off_df, 'TERM_YEARS', paramdict = bag, model =␣
↪RandomForestRegressor, seed = 42)
```

train RMSE: 3.8921053328608
test RMSE: 3.9502667128129887
Feature Importance Table

	Importance	Features
8	2.536840e-01	ROBBERY
16	1.510565e-01	NON-DEADLY_WEAPON
14	1.363828e-01	DEADLY_WEAPON
2	1.043347e-01	AGE
15	9.848255e-02	NO_WEAPON
..	…	…
121	1.062976e-05	CIRCUIT_12.0
91	7.731568e-06	COUNTY_ST. JOHNS
152	5.568020e-06	SA_NAME_Jeff Siegmeister
115	2.512912e-06	CIRCUIT_5.0
42	1.992044e-07	COUNTY_COLLIER

[102 rows x 2 columns]

Histogram of Feature Importance

Feature Beeswarm Plot
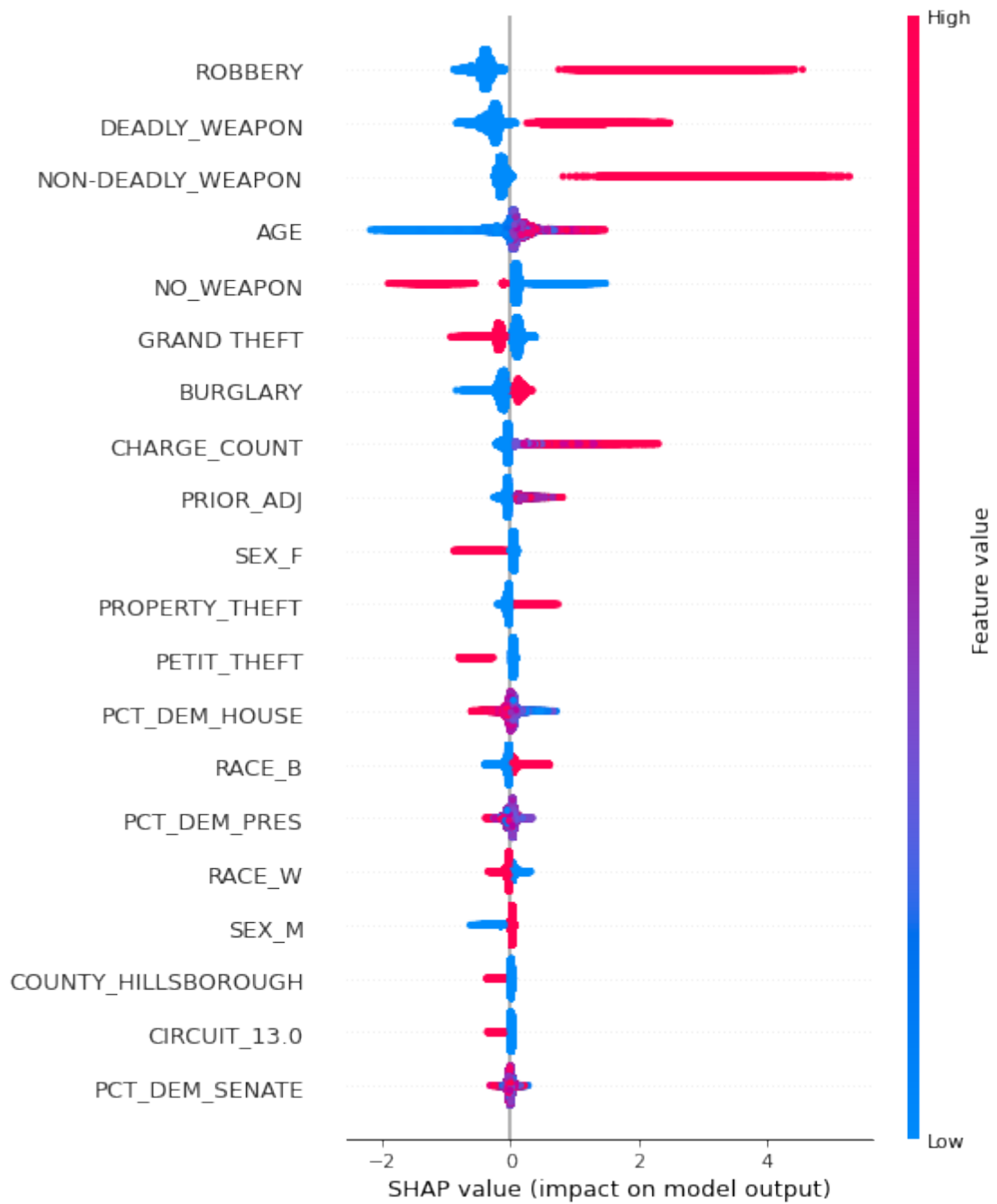
0.3 XGBoost Models

0.3.1 XGB Classifier - Drug Charges

```python
bag = {'max_depth':[4, 6],
       'min_child_weight':[10, 20],
       'alpha': [0.1, 0.2],
       'colsample_bytree': [0.6, 0.75]}

# XGB hyperparameter space
drug_sa_tree = get_tree(drug_sa_df, target = 'FINAL_ACTION_DESC', paramdict =
   bag, model = XGBClassifier, seed =10)
```

```
Confusion matrix:
 [[  55  201]
 [  34 1168]] Classification Report:
              precision    recall  f1-score   support

           0       0.62      0.21      0.32       256
           1       0.85      0.97      0.91      1202

    accuracy                           0.84      1458
   macro avg       0.74      0.59      0.61      1458
weighted avg       0.81      0.84      0.81      1458
 Accuracy: 0.8388203017832647
Feature Importance Table
```

	Importance	Features
64	0.210552	COUNTY_VOLUSIA
37	0.181287	OFFENSE_Hallucinogen-Possess
62	0.119848	COUNTY_PINELLAS
42	0.057983	OFFENSE_Marijuana-Possess
81	0.046652	SA_NAME_Bernie McCabe
0	0.039829	CHARGE_DEGREE_First Degree
8	0.037425	CHARGE_LEVEL_Misdemeanor
26	0.026139	OFFENSE_Drug Abuse Prevention and Control
4	0.023752	CHARGE_DEGREE_Second Degree
27	0.023718	OFFENSE_Drug Equip-Possess
7	0.023427	CHARGE_LEVEL_Felony
49	0.019630	PCT_DEM_PRES
48	0.018194	PCT_DEM_SENATE
75	0.017768	CIRCUIT_6
47	0.016123	PCT_DEM_HOUSE
72	0.012300	YEAR_2021
5	0.012288	CHARGE_DEGREE_Third Degree
70	0.011662	YEAR_2019
52	0.011069	DRUG_POSSESSION
22	0.009100	OFFENSE_Cocaine-Possess
71	0.008868	YEAR_2020

```
51     0.008230                            PETTY_THEFT
31     0.007583                    OFFENSE_Drugs-Possess
12     0.007220                             RACE_Black
16     0.006941                               SEX_Male
77     0.006496                             CIRCUIT_13
14     0.006088                             RACE_White
17     0.005681                            SEX_Unknown
13     0.005194                           RACE_Unknown
15     0.005003                             SEX_Female
76     0.004959                              CIRCUIT_7
53     0.003788                       NONVIOLENT_CRIME
80     0.002229                   SA_NAME_Andrew Warren
50     0.002077                             DEMOCRAT_YN
58     0.000898                    COUNTY_HILLSBOROUGH
```
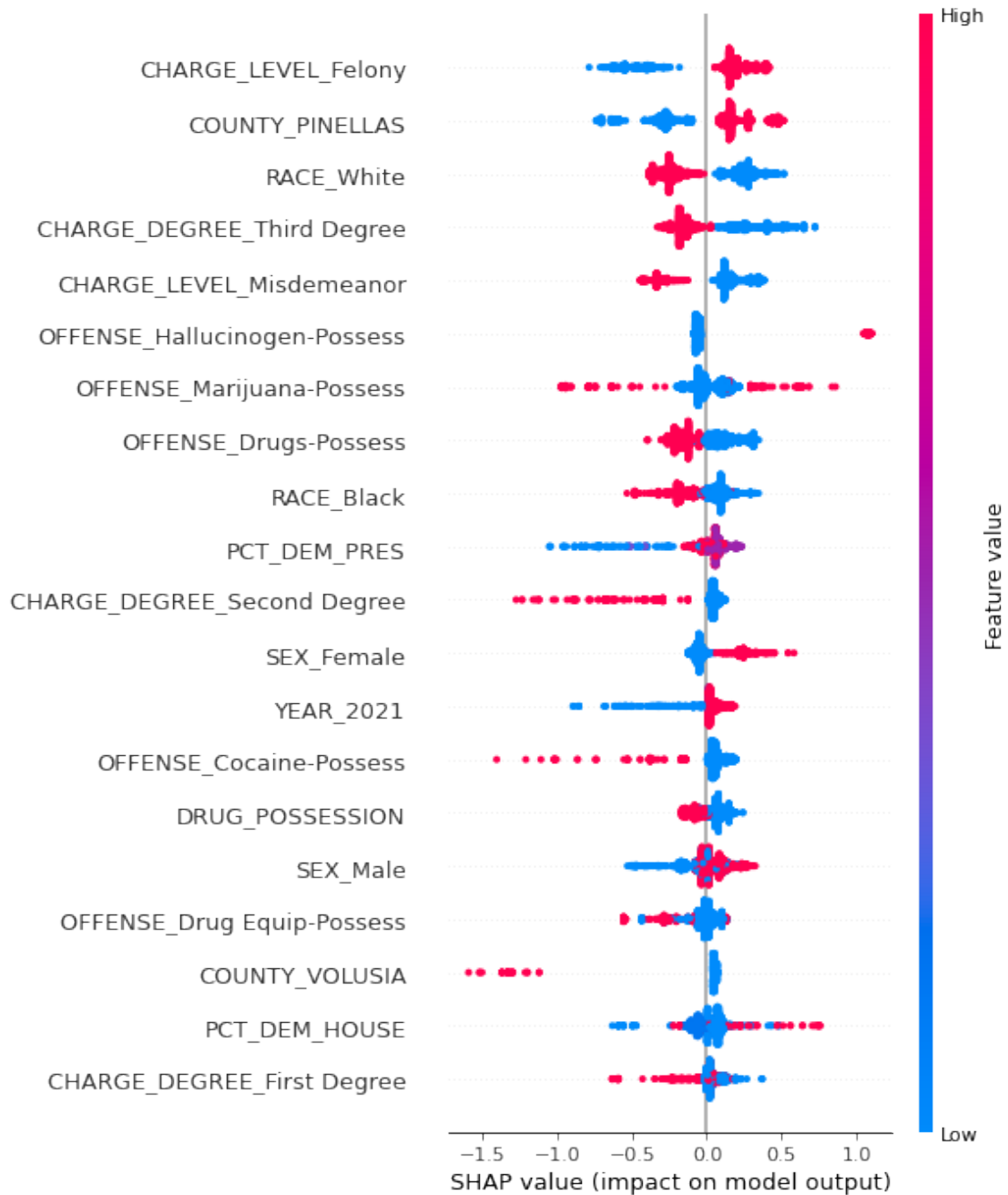
Histogram of Feature Importance
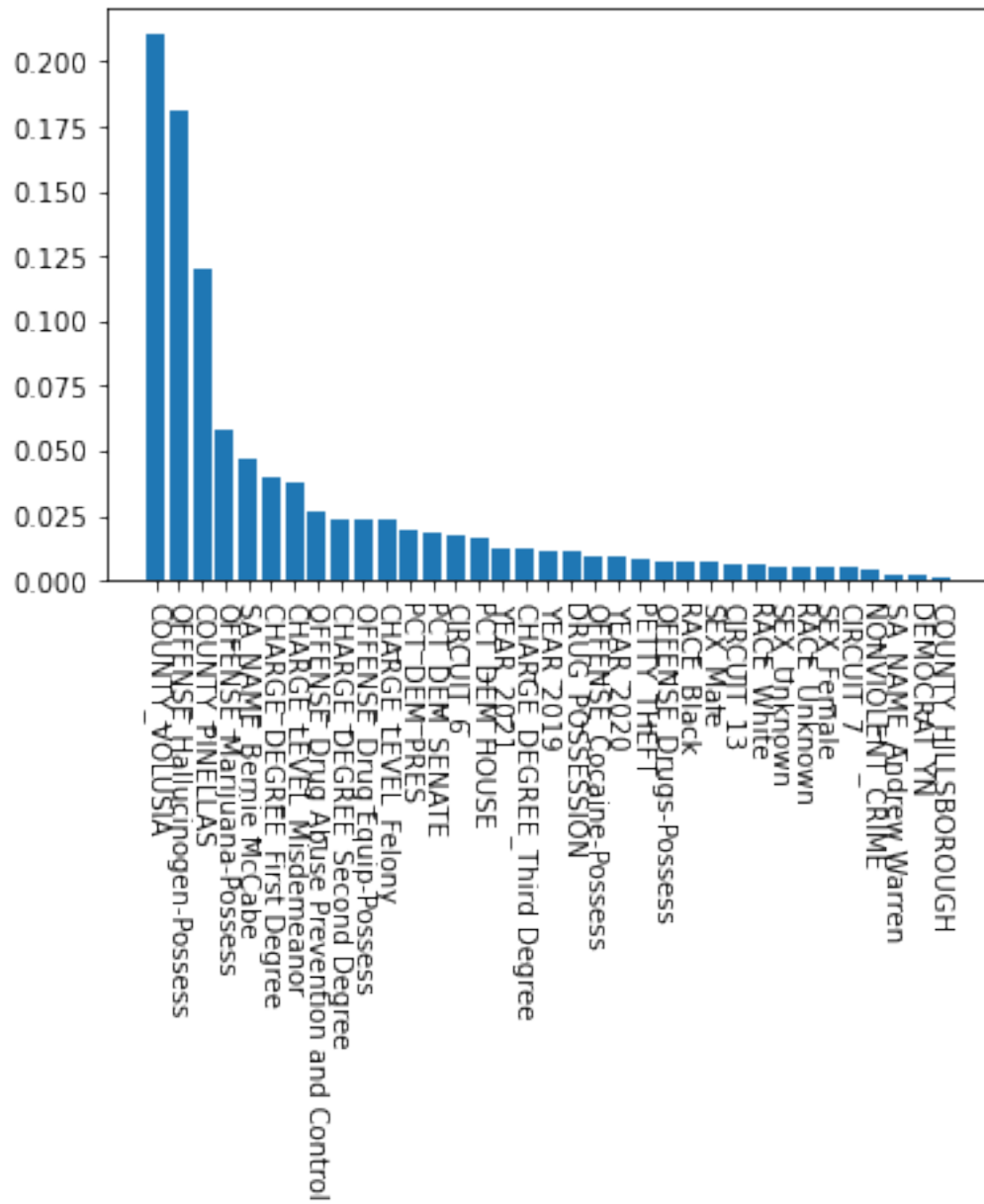
ntree_limit is deprecated, use `iteration_range` or model slicing instead.

Feature Beeswarm Plot

0.3.2 XGB Classifier - Theft Charges

```
theft_sa_tree = get_tree(theft_sa_df, target = 'FINAL_ACTION_DESC', paramdict =␣
 ↪bag, model = XGBClassifier,seed=10)
```

```
Confusion matrix:
 [[ 16  89]
 [  4 735]] Classification Report:
```
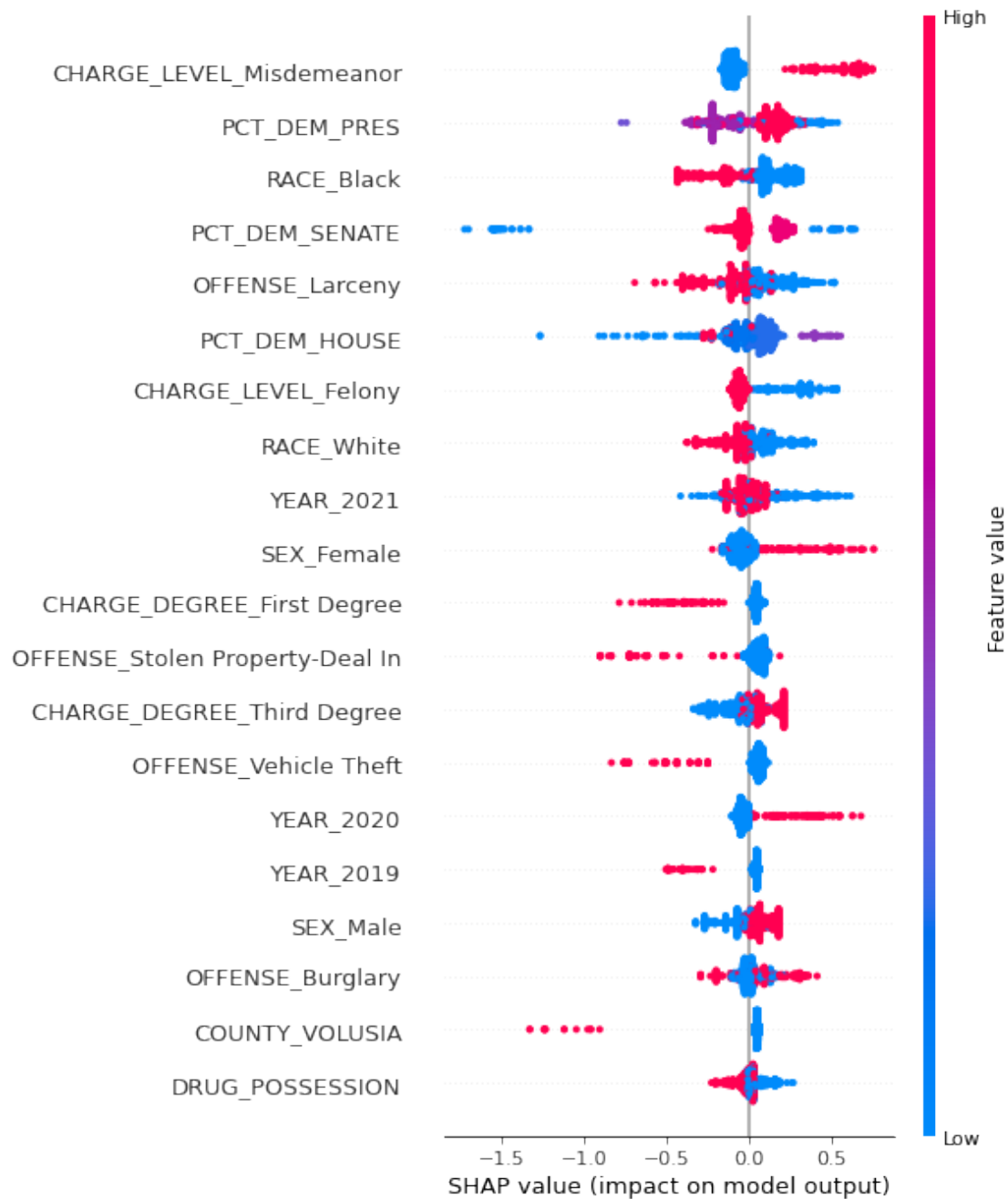
	precision	recall	f1-score	support
0	0.80	0.15	0.26	105
1	0.89	0.99	0.94	739
accuracy			0.89	844
macro avg	0.85	0.57	0.60	844
weighted avg	0.88	0.89	0.86	844

Accuracy: 0.8898104265402843

Feature Importance Table

	Importance	Features
65	0.385052	SA_NAME_R. J. Larizza
46	0.105756	COUNTY_VOLUSIA
31	0.048875	PCT_DEM_SENATE
44	0.039396	COUNTY_PINELLAS
23	0.029780	OFFENSE_Robbery
27	0.028457	OFFENSE_Stolen Property-Deal In
58	0.026575	CIRCUIT_7
8	0.026419	CHARGE_LEVEL_Misdemeanor
55	0.019185	YEAR_2021
18	0.018504	SEX_Unknown
54	0.017357	YEAR_2020
32	0.016348	PCT_DEM_PRES
29	0.016015	OFFENSE_Vehicle Theft
35	0.014867	DRUG_POSSESSION
12	0.014061	RACE_Black
30	0.013163	PCT_DEM_HOUSE
19	0.012785	OFFENSE_Burglary
41	0.012776	COUNTY_HILLSBOROUGH
7	0.012500	CHARGE_LEVEL_Felony
17	0.012280	SEX_Male
57	0.011885	CIRCUIT_6
4	0.011252	CHARGE_DEGREE_Second Degree
5	0.011209	CHARGE_DEGREE_Third Degree
34	0.010745	PETTY_THEFT
15	0.010690	RACE_White
16	0.010631	SEX_Female
22	0.010184	OFFENSE_Larceny
63	0.009745	SA_NAME_Bernie McCabe
53	0.008315	YEAR_2019
1	0.007952	CHARGE_DEGREE_First Degree punishable by life
14	0.007710	RACE_Unknown
33	0.006108	DEMOCRAT_YN
0	0.005611	CHARGE_DEGREE_First Degree
36	0.003954	NONVIOLENT_CRIME
59	0.003857	CIRCUIT_13

Histogram of Feature Importance

ntree_limit is deprecated, use `iteration_range` or model slicing instead.

Feature Beeswarm Plot

0.3.3 XGB Regressor - Drug Charges

```
boost = {'max_depth':[8, 12],
         'min_child_weight':[10, 20],
         'alpha': [0.1, 0.2],
         'colsample_bytree': [0.6, 0.75]}
```

```
# XGB Regressors
drug_off_tree = get_tree(drug_off_df, target = 'TERM_YEARS', paramdict = boost,⮠
  ↪model = XGBRegressor, seed=42)
```
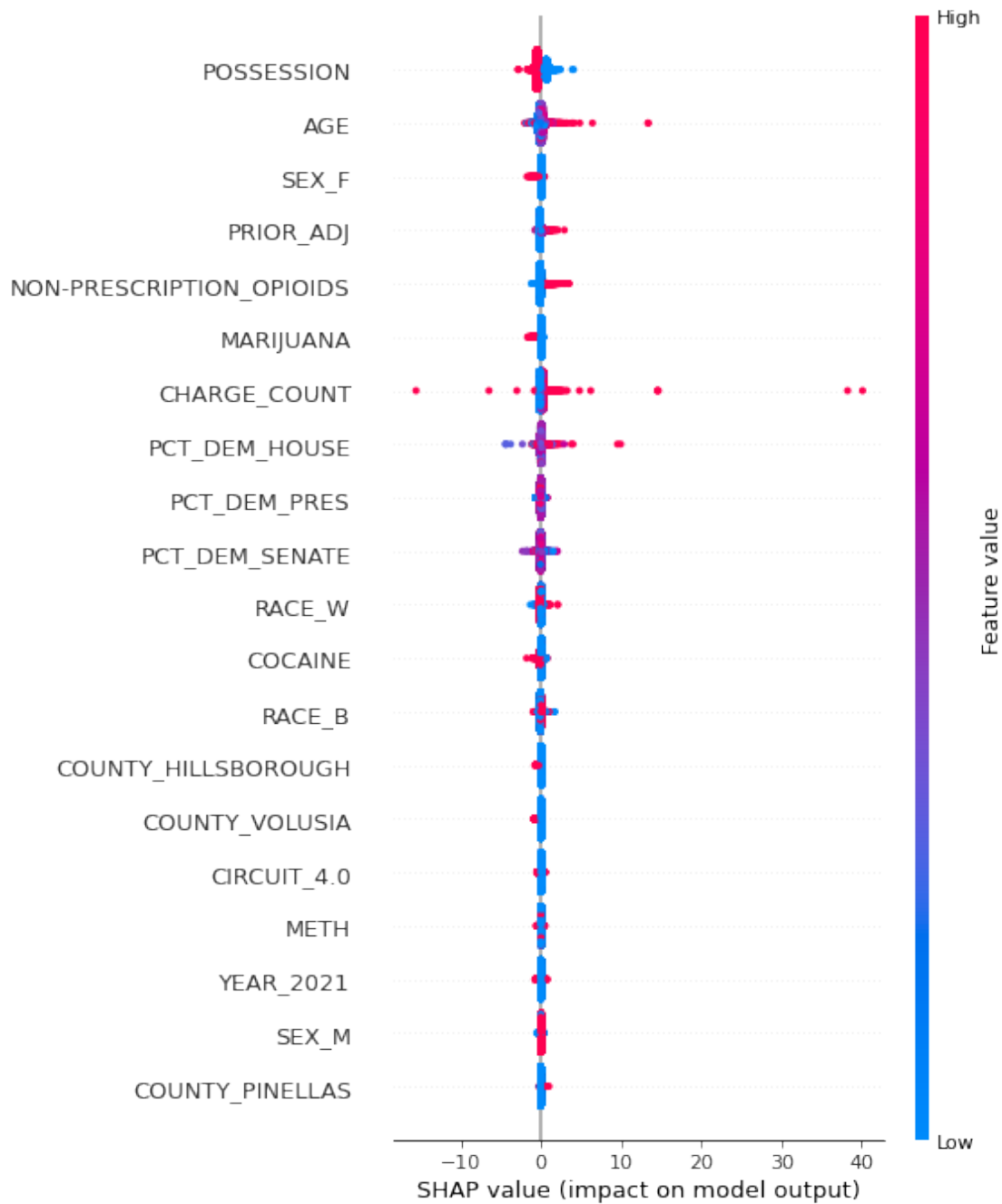
train RMSE: 2.792298127166822
test RMSE: 3.1339048348190426
Feature Importance Table

```
      Importance                     Features
121     0.046903                 CIRCUIT_15.0
80      0.035520            COUNTY_PALM BEACH
56      0.032114          COUNTY_HILLSBOROUGH
3       0.026963                   POSSESSION
133     0.024509            SA_NAME_Bill Eddins
..           ...                          ...
47      0.000812               COUNTY_GADSDEN
10      0.000812                    DATE_RAPE
14      0.000330                       RACE_A
67      0.000324               COUNTY_LIBERTY
167     0.000126  POLITICAL_PARTY_Appointee
```

[165 rows x 2 columns]

Histogram of Feature Importance

ntree_limit is deprecated, use `iteration_range` or model slicing instead.

Feature Beeswarm Plot

0.3.4 XGB Regressor - Theft Charges

```
theft_off_tree = get_tree(theft_off_df, 'TERM_YEARS', paramdict = boost, model↵
 ↪=XGBRegressor, seed = 42)
```

train RMSE: 3.6627520954662947
test RMSE: 3.911753631973393

```
Feature Importance Table

     Importance               Features
8      0.117143                ROBBERY
16     0.072479      NON-DEADLY_WEAPON
15     0.057819              NO_WEAPON
24     0.033785                  SEX_M
37     0.025726         COUNTY_BROWARD
..          ...                     ...
144    0.000659    SA_NAME_Dennis W. Ward
141    0.000455    SA_NAME_Catherine Vogel
71     0.000321         COUNTY_MADISON
13     0.000225        SUBSTANCE_THEFT
55     0.000146          COUNTY_HARDEE

[163 rows x 2 columns]

Histogram of Feature Importance
```

ntree_limit is deprecated, use `iteration_range` or model slicing instead.

Feature Beeswarm Plot