In [1]:
```
import pandas as pd
import numpy as np
import h2o
from h2o.automl import H2OAutoML
```

In [2]:
```
NIJ_Training_df= pd.read_csv("NIJ_s_Recidivism_Challenge_Training_Dataset-sara.csv")
```

In [3]:
```
NIJ_Training_df.columns
```

Out[3]:
```
Index(['ID', 'Gender', 'Race', 'Age_at_Release', 'Residence_PUMA',
       'Gang_Affiliated', 'Supervision_Risk_Score_First',
       'Supervision_Level_First', 'Education_Level', 'Dependents',
       'Prison_Offense', 'Prison_Years', 'Prior_Arrest_Episodes_Felony',
       'Prior_Arrest_Episodes_Misd', 'Prior_Arrest_Episodes_Violent',
       'Prior_Arrest_Episodes_Property', 'Prior_Arrest_Episodes_Drug',
       'Prior_Arrest_Episodes_PPViolationCharges',
       'Prior_Arrest_Episodes_DVCharges', 'Prior_Arrest_Episodes_GunCharges',
       'Prior_Conviction_Episodes_Felony', 'Prior_Conviction_Episodes_Misd',
       'Prior_Conviction_Episodes_Viol', 'Prior_Conviction_Episodes_Prop',
       'Prior_Conviction_Episodes_Drug',
       'Prior_Conviction_Episodes_PPViolationCharges',
       'Prior_Conviction_Episodes_DomesticViolenceCharges',
       'Prior_Conviction_Episodes_GunCharges', 'Prior_Revocations_Parole',
       'Prior_Revocations_Probation', 'Condition_MH_SA', 'Condition_Cog_Ed',
       'Condition_Other', 'Violations_ElectronicMonitoring',
       'Violations_Instruction', 'Violations_FailToReport',
       'Violations_MoveWithoutPermission', 'Delinquency_Reports',
       'Program_Attendances', 'Program_UnexcusedAbsences', 'Residence_Changes',
       'Avg_Days_per_DrugTest', 'DrugTests_THC_Positive',
       'DrugTests_Cocaine_Positive', 'DrugTests_Meth_Positive',
       'DrugTests_Other_Positive', 'Percent_Days_Employed', 'Jobs_Per_Year',
       'Employment_Exempt', 'Recidivism_Within_3years',
       'Recidivism_Arrest_Year1', 'Recidivism_Arrest_Year2',
       'Recidivism_Arrest_Year3'],
      dtype='object')
```

In [4]:
```
x_train =NIJ_Training_df[NIJ_Training_df.columns[0:-4]]
x_train.columns
```

Out[4]:
```
Index(['ID', 'Gender', 'Race', 'Age_at_Release', 'Residence_PUMA',
       'Gang_Affiliated', 'Supervision_Risk_Score_First',
       'Supervision_Level_First', 'Education_Level', 'Dependents',
       'Prison_Offense', 'Prison_Years', 'Prior_Arrest_Episodes_Felony',
       'Prior_Arrest_Episodes_Misd', 'Prior_Arrest_Episodes_Violent',
       'Prior_Arrest_Episodes_Property', 'Prior_Arrest_Episodes_Drug',
       'Prior_Arrest_Episodes_PPViolationCharges',
       'Prior_Arrest_Episodes_DVCharges', 'Prior_Arrest_Episodes_GunCharges',
       'Prior_Conviction_Episodes_Felony', 'Prior_Conviction_Episodes_Misd',
       'Prior_Conviction_Episodes_Viol', 'Prior_Conviction_Episodes_Prop',
       'Prior_Conviction_Episodes_Drug',
       'Prior_Conviction_Episodes_PPViolationCharges',
       'Prior_Conviction_Episodes_DomesticViolenceCharges',
       'Prior_Conviction_Episodes_GunCharges', 'Prior_Revocations_Parole',
       'Prior_Revocations_Probation', 'Condition_MH_SA', 'Condition_Cog_Ed',
       'Condition_Other', 'Violations_ElectronicMonitoring',
       'Violations_Instruction', 'Violations_FailToReport',
       'Violations_MoveWithoutPermission', 'Delinquency_Reports',
       'Program_Attendances', 'Program_UnexcusedAbsences', 'Residence_Changes',
```

```
       'Avg_Days_per_DrugTest', 'DrugTests_THC_Positive',
       'DrugTests_Cocaine_Positive', 'DrugTests_Meth_Positive',
       'DrugTests_Other_Positive', 'Percent_Days_Employed', 'Jobs_Per_Year',
       'Employment_Exempt'],
      dtype='object')
```

In [5]:

```
# Number of records having null Values in each column
x_train.isnull().sum()
```

Out[5]:

```
ID                                                      0
Gender                                                  0
Race                                                    0
Age_at_Release                                          0
Residence_PUMA                                          0
Gang_Affiliated                                      2217
Supervision_Risk_Score_First                          330
Supervision_Level_First                              1212
Education_Level                                         0
Dependents                                              0
Prison_Offense                                       2321
Prison_Years                                            0
Prior_Arrest_Episodes_Felony                            0
Prior_Arrest_Episodes_Misd                              0
Prior_Arrest_Episodes_Violent                           0
Prior_Arrest_Episodes_Property                          0
Prior_Arrest_Episodes_Drug                              0
Prior_Arrest_Episodes_PPViolationCharges                0
Prior_Arrest_Episodes_DVCharges                         0
Prior_Arrest_Episodes_GunCharges                        0
Prior_Conviction_Episodes_Felony                        0
Prior_Conviction_Episodes_Misd                          0
Prior_Conviction_Episodes_Viol                          0
Prior_Conviction_Episodes_Prop                          0
Prior_Conviction_Episodes_Drug                          0
Prior_Conviction_Episodes_PPViolationCharges            0
Prior_Conviction_Episodes_DomesticViolenceCharges       0
Prior_Conviction_Episodes_GunCharges                    0
Prior_Revocations_Parole                                0
Prior_Revocations_Probation                             0
Condition_MH_SA                                         0
Condition_Cog_Ed                                        0
Condition_Other                                         0
Violations_ElectronicMonitoring                         0
Violations_Instruction                                  0
Violations_FailToReport                                 0
Violations_MoveWithoutPermission                        0
Delinquency_Reports                                     0
Program_Attendances                                     0
Program_UnexcusedAbsences                               0
Residence_Changes                                       0
Avg_Days_per_DrugTest                                4260
DrugTests_THC_Positive                               3632
DrugTests_Cocaine_Positive                           3632
DrugTests_Meth_Positive                              3632
DrugTests_Other_Positive                             3632
Percent_Days_Employed                                 307
Jobs_Per_Year                                         534
Employment_Exempt                                       0
dtype: int64
```

In [6]:

```
# Total Number of Records in a Dataset with null values in any column
x_train.isna().any(axis=1).sum()
```

Out[6]: 8190

In [7]:
```
# Find the Categoricorical columns for feature engineering(get Dummies)
cat_columns=[col for col in x_train.columns if x_train[col].dtypes=='O']
x_train[cat_columns].isnull().sum()
```

Out[7]:
```
Gender                                 0
Race                                   0
Age_at_Release                         0
Gang_Affiliated                     2217
Supervision_Level_First             1212
Education_Level                        0
Dependents                             0
Prison_Offense                      2321
Prison_Years                           0
Prior_Arrest_Episodes_Felony           0
Prior_Arrest_Episodes_Misd             0
Prior_Arrest_Episodes_Violent          0
Prior_Arrest_Episodes_Property         0
Prior_Arrest_Episodes_Drug             0
Prior_Arrest_Episodes_PPViolationCharges       0
Prior_Conviction_Episodes_Felony       0
Prior_Conviction_Episodes_Misd         0
Prior_Conviction_Episodes_Prop         0
Prior_Conviction_Episodes_Drug         0
Delinquency_Reports                    0
Program_Attendances                    0
Program_UnexcusedAbsences              0
Residence_Changes                      0
dtype: int64
```

In [8]:
```
x_train[cat_columns].isna().any(axis=1).sum()
```

Out[8]: 5211

In [9]:
```
cat_variables = x_train[cat_columns]
# print(cat_variables)
cat_dummies = pd.get_dummies(cat_variables,dummy_na=True)
print(cat_dummies.head())
```

```
   Gender_F  Gender_M  Gender_nan  Race_BLACK  Race_WHITE  Race_nan  \
0         0         1           0           1           0         0
1         0         1           0           1           0         0
2         0         1           0           1           0         0
3         0         1           0           0           1         0
4         0         1           0           0           1         0

   Age_at_Release_18-22  Age_at_Release_23-27  Age_at_Release_28-32  \
0                     0                     0                     0
1                     0                     0                     0
2                     0                     0                     0
3                     0                     0                     0
4                     0                     0                     0

   Age_at_Release_33-37  ...  Program_UnexcusedAbsences_0  \
0                     0  ...                            1
1                     1  ...                            1
2                     0  ...                            1
3                     0  ...                            1
4                     1  ...                            1
```

```
   Program_UnexcusedAbsences_1  Program_UnexcusedAbsences_2  \
0                            0                            0
1                            0                            0
2                            0                            0
3                            0                            0
4                            0                            0

   Program_UnexcusedAbsences_3 or more  Program_UnexcusedAbsences_nan  \
0                                    0                              0
1                                    0                              0
2                                    0                              0
3                                    0                              0
4                                    0                              0

   Residence_Changes_0  Residence_Changes_1  Residence_Changes_2  \
0                    0                    0                    1
1                    0                    0                    1
2                    1                    0                    0
3                    0                    0                    0
4                    1                    0                    0

   Residence_Changes_3 or more  Residence_Changes_nan
0                            0                      0
1                            0                      0
2                            0                      0
3                            1                      0
4                            0                      0

[5 rows x 135 columns]
```

In [10]:

```
x_train.columns
```

Out[10]:

```
Index(['ID', 'Gender', 'Race', 'Age_at_Release', 'Residence_PUMA',
       'Gang_Affiliated', 'Supervision_Risk_Score_First',
       'Supervision_Level_First', 'Education_Level', 'Dependents',
       'Prison_Offense', 'Prison_Years', 'Prior_Arrest_Episodes_Felony',
       'Prior_Arrest_Episodes_Misd', 'Prior_Arrest_Episodes_Violent',
       'Prior_Arrest_Episodes_Property', 'Prior_Arrest_Episodes_Drug',
       'Prior_Arrest_Episodes_PPViolationCharges',
       'Prior_Arrest_Episodes_DVCharges', 'Prior_Arrest_Episodes_GunCharges',
       'Prior_Conviction_Episodes_Felony', 'Prior_Conviction_Episodes_Misd',
       'Prior_Conviction_Episodes_Viol', 'Prior_Conviction_Episodes_Prop',
       'Prior_Conviction_Episodes_Drug',
       'Prior_Conviction_Episodes_PPViolationCharges',
       'Prior_Conviction_Episodes_DomesticViolenceCharges',
       'Prior_Conviction_Episodes_GunCharges', 'Prior_Revocations_Parole',
       'Prior_Revocations_Probation', 'Condition_MH_SA', 'Condition_Cog_Ed',
       'Condition_Other', 'Violations_ElectronicMonitoring',
       'Violations_Instruction', 'Violations_FailToReport',
       'Violations_MoveWithoutPermission', 'Delinquency_Reports',
       'Program_Attendances', 'Program_UnexcusedAbsences', 'Residence_Changes',
       'Avg_Days_per_DrugTest', 'DrugTests_THC_Positive',
       'DrugTests_Cocaine_Positive', 'DrugTests_Meth_Positive',
       'DrugTests_Other_Positive', 'Percent_Days_Employed', 'Jobs_Per_Year',
       'Employment_Exempt'],
      dtype='object')
```

In [11]:

```
x_train = x_train.drop(cat_columns, axis=1)
```

In [12]:

```
x_train.columns
```

Out[12]:
```
Index(['ID', 'Residence_PUMA', 'Supervision_Risk_Score_First',
       'Prior_Arrest_Episodes_DVCharges', 'Prior_Arrest_Episodes_GunCharges',
       'Prior_Conviction_Episodes_Viol',
       'Prior_Conviction_Episodes_PPViolationCharges',
       'Prior_Conviction_Episodes_DomesticViolenceCharges',
       'Prior_Conviction_Episodes_GunCharges', 'Prior_Revocations_Parole',
       'Prior_Revocations_Probation', 'Condition_MH_SA', 'Condition_Cog_Ed',
       'Condition_Other', 'Violations_ElectronicMonitoring',
       'Violations_Instruction', 'Violations_FailToReport',
       'Violations_MoveWithoutPermission', 'Avg_Days_per_DrugTest',
       'DrugTests_THC_Positive', 'DrugTests_Cocaine_Positive',
       'DrugTests_Meth_Positive', 'DrugTests_Other_Positive',
       'Percent_Days_Employed', 'Jobs_Per_Year', 'Employment_Exempt'],
      dtype='object')
```

8 - null get dummies, dummy_na - gang_affiliated _false=np.null, gangaffliated_true= np.nullgang_affiliated_nan=1 drop ganga_affliated_nan

In [13]:
```
x_train = pd.concat([x_train, cat_dummies], axis=1)
print(x_train.columns)
```

```
Index(['ID', 'Residence_PUMA', 'Supervision_Risk_Score_First',
       'Prior_Arrest_Episodes_DVCharges', 'Prior_Arrest_Episodes_GunCharges',
       'Prior_Conviction_Episodes_Viol',
       'Prior_Conviction_Episodes_PPViolationCharges',
       'Prior_Conviction_Episodes_DomesticViolenceCharges',
       'Prior_Conviction_Episodes_GunCharges', 'Prior_Revocations_Parole',
       ...
       'Program_UnexcusedAbsences_0', 'Program_UnexcusedAbsences_1',
       'Program_UnexcusedAbsences_2', 'Program_UnexcusedAbsences_3 or more',
       'Program_UnexcusedAbsences_nan', 'Residence_Changes_0',
       'Residence_Changes_1', 'Residence_Changes_2',
       'Residence_Changes_3 or more', 'Residence_Changes_nan'],
      dtype='object', length=161)
```

In [14]:
```
x_train.loc[x_train.Gang_Affiliated_nan == 1, ["Gang_Affiliated_False", "Gang_Affiliate
# x_train.drop('Gang_Affiliated_nan',axis=1)
```

In [15]:
```
x_train.loc[x_train.Supervision_Level_First_nan == 1, ["Supervision_Level_First_High",
# x_train.drop('Supervision_Level_First_nan',axis=1)
```

In [16]:
```
x_train.loc[x_train.Prison_Offense_nan == 1, ["Prison_Offense_Drug", "Prison_Offense_Ot
# x_train.drop('Prison_Offense_nan',axis=1)
```

In [17]:
```
x_train.drop(['Gender_nan','Race_nan','Age_at_Release_nan','Gang_Affiliated_nan','Super
```

In [18]:
```
x_train_nan_values = x_train[x_train.isna().any(axis=1)]
len(x_train_nan_values)
```

Out[18]: 8190

In [19]:
```
from sklearn.impute import KNNImputer
imputer = KNNImputer(n_neighbors=10)
```

```
x_train = pd.DataFrame(imputer.fit_transform(x_train),columns = x_train.columns)
```

In [20]:

```
x_train.isna().any()
```

Out[20]:

```
ID                                   False
Residence_PUMA                       False
Supervision_Risk_Score_First         False
Prior_Arrest_Episodes_DVCharges      False
Prior_Arrest_Episodes_GunCharges     False
                                     ...
Program_UnexcusedAbsences_3 or more  False
Residence_Changes_0                  False
Residence_Changes_1                  False
Residence_Changes_2                  False
Residence_Changes_3 or more          False
Length: 138, dtype: bool
```

In [21]:

```
x_train.isna().sum().sum()
```

Out[21]:

```
0
```

In [2]:

```
h2o.init()
```

```
Checking whether there is an H2O instance running at http://localhost:54321 ..... not fo
und.
Attempting to start a local H2O server...
; Java HotSpot(TM) 64-Bit Server VM (build 25.291-b10, mixed mode)
  Starting server from C:\Users\Vimalathithan\anaconda3\Lib\site-packages\h2o\backend\bi
n\h2o.jar
  Ice root: C:\Users\VIMALA~1\AppData\Local\Temp\tmpikqfpg04
  JVM stdout: C:\Users\VIMALA~1\AppData\Local\Temp\tmpikqfpg04\h2o_Vimalathithan_started
_from_python.out
  JVM stderr: C:\Users\VIMALA~1\AppData\Local\Temp\tmpikqfpg04\h2o_Vimalathithan_started
_from_python.err
  Server is running at http://127.0.0.1:54321
Connecting to H2O server at http://127.0.0.1:54321 ... successful.
```

H2O_cluster_uptime:	06 secs
H2O_cluster_timezone:	America/Chicago
H2O_data_parsing_timezone:	UTC
H2O_cluster_version:	3.32.1.3
H2O_cluster_version_age:	8 days
H2O_cluster_name:	H2O_from_python_Vimalathithan_wz1old
H2O_cluster_total_nodes:	1
H2O_cluster_free_memory:	3.542 Gb
H2O_cluster_total_cores:	8
H2O_cluster_allowed_cores:	8
H2O_cluster_status:	accepting new members, healthy
H2O_connection_url:	http://127.0.0.1:54321

H2O_connection_proxy:	{"http": null, "https": null}
H2O_internal_security:	False
H2O_API_Extensions:	Amazon S3, Algos, AutoML, Core V3, TargetEncoder, Core V4
Python_version:	3.8.8 final

In [3]:

```
NIJ_Train_Mod = h2o.import_file("NIJ_s_Recidivism_Challenge_Training_Dataset-sara.csv")
```

```
Parse progress: |████████████████████████████████████████████████████████| 100%
```

In [4]:

```
NIJ_Train_Mod["Recidivism_Arrest_Year1"]=NIJ_Train_Mod["Recidivism_Arrest_Year1"].asfac
```

In [5]:

```
NIJ_Train_Mod.describe()
```

```
Rows:18028
Cols:53
```

	ID	Gender	Race	Age_at_Release	Residence_PUMA	Gang_Affiliated	Supe
type	int	enum	enum	enum	int	enum	
mins	1.0				1.0		
mean	13386.065342800088				12.307577102285334		
maxs	26761.0				25.0		
sigma	7721.4519921005185				7.143255483699997		
zeros	0				0		
missing	0	0	0	0	0	2217	
0	1.0	M	BLACK	43-47	16.0	false	
1	2.0	M	BLACK	33-37	16.0	false	
2	3.0	M	BLACK	48 or older	24.0	false	
3	4.0	M	WHITE	38-42	16.0	false	
4	5.0	M	WHITE	33-37	16.0	false	
5	7.0	M	BLACK	48 or older	18.0	false	
6	9.0	F	BLACK	43-47	5.0		

	ID	Gender	Race	Age_at_Release	Residence_PUMA	Gang_Affiliated	Supe
7	10.0	M	BLACK	43-47	16.0	false	
8	11.0	M	WHITE	43-47	5.0	false	
9	13.0	M	WHITE	48 or older	18.0	false	

In [6]:

```
y= "Recidivism_Arrest_Year1"
NIJ_Train_Mod=NIJ_Train_Mod.drop(["ID","Recidivism_Within_3years","Recidivism_Arrest_Ye
x= NIJ_Train_Mod.columns
print(x)
```

```
['Gender', 'Race', 'Age_at_Release', 'Residence_PUMA', 'Gang_Affiliated', 'Supervision_R
isk_Score_First', 'Supervision_Level_First', 'Education_Level', 'Dependents', 'Prison_Of
fense', 'Prison_Years', 'Prior_Arrest_Episodes_Felony', 'Prior_Arrest_Episodes_Misd', 'P
rior_Arrest_Episodes_Violent', 'Prior_Arrest_Episodes_Property', 'Prior_Arrest_Episodes_
Drug', 'Prior_Arrest_Episodes_PPViolationCharges', 'Prior_Arrest_Episodes_DVCharges', 'P
rior_Arrest_Episodes_GunCharges', 'Prior_Conviction_Episodes_Felony', 'Prior_Conviction_
Episodes_Misd', 'Prior_Conviction_Episodes_Viol', 'Prior_Conviction_Episodes_Prop', 'Pri
or_Conviction_Episodes_Drug', 'Prior_Conviction_Episodes_PPViolationCharges', 'Prior_Con
viction_Episodes_DomesticViolenceCharges', 'Prior_Conviction_Episodes_GunCharges', 'Prio
r_Revocations_Parole', 'Prior_Revocations_Probation', 'Condition_MH_SA', 'Condition_Cog_
Ed', 'Condition_Other', 'Violations_ElectronicMonitoring', 'Violations_Instruction', 'Vi
olations_FailToReport', 'Violations_MoveWithoutPermission', 'Delinquency_Reports', 'Prog
ram_Attendances', 'Program_UnexcusedAbsences', 'Residence_Changes', 'Avg_Days_per_DrugTe
st', 'DrugTests_THC_Positive', 'DrugTests_Cocaine_Positive', 'DrugTests_Meth_Positive',
'DrugTests_Other_Positive', 'Percent_Days_Employed', 'Jobs_Per_Year', 'Employment_Exemp
t', 'Recidivism_Arrest_Year1']
```

In [7]:

```
aml = H2OAutoML(max_models = 10, seed = 1)
aml.train(x = x, y = y, training_frame = NIJ_Train_Mod)
```

```
AutoML progress: |
06:17:10.678: AutoML: XGBoost is not available; skipping it.

████████████████████████████████████████████████████████| 100%
```

In [8]:

```
lb = aml.leaderboard
```

In [9]:

```
lb.head()
```

model_id	auc	logloss	aucpr	mean_per_class_error
StackedEnsemble_AllModels_AutoML_20210528_061710	0.867471	0.40513	0.721791	0.219541
GBM_1_AutoML_20210528_061710	0.864168	0.410363	0.713609	0.21977
StackedEnsemble_BestOfFamily_AutoML_20210528_061710	0.864058	0.409899	0.713256	0.219616
GBM_2_AutoML_20210528_061710	0.863699	0.410663	0.712505	0.219131
GBM_grid__1_AutoML_20210528_061710_model_1	0.862715	0.413537	0.712251	0.221504
GBM_5_AutoML_20210528_061710	0.862392	0.411906	0.711023	0.221938

model_id	auc	logloss	aucpr	mean_per_class_error
GBM_3_AutoML_20210528_061710	0.859444	0.41611	0.702074	0.224182
GBM_4_AutoML_20210528_061710	0.858616	0.418236	0.704776	0.228805
XRT_1_AutoML_20210528_061710	0.818852	0.473279	0.635521	0.257104
DRF_1_AutoML_20210528_061710	0.818494	0.473402	0.63413	0.255788

Out[9]:

In [10]:

```
lb.head(rows=lb.nrows)
```

model_id	auc	logloss	aucpr	mean_per_class_error
StackedEnsemble_AllModels_AutoML_20210528_061710	0.867471	0.40513	0.721791	0.219541
GBM_1_AutoML_20210528_061710	0.864168	0.410363	0.713609	0.21977
StackedEnsemble_BestOfFamily_AutoML_20210528_061710	0.864058	0.409899	0.713256	0.219616
GBM_2_AutoML_20210528_061710	0.863699	0.410663	0.712505	0.219131
GBM_grid__1_AutoML_20210528_061710_model_1	0.862715	0.413537	0.712251	0.221504
GBM_5_AutoML_20210528_061710	0.862392	0.411906	0.711023	0.221938
GBM_3_AutoML_20210528_061710	0.859444	0.41611	0.702074	0.224182
GBM_4_AutoML_20210528_061710	0.858616	0.418236	0.704776	0.228805
XRT_1_AutoML_20210528_061710	0.818852	0.473279	0.635521	0.257104
DRF_1_AutoML_20210528_061710	0.818494	0.473402	0.63413	0.255788
GLM_1_AutoML_20210528_061710	0.769877	0.509474	0.56241	0.295848
DeepLearning_1_AutoML_20210528_061710	0.764187	0.514955	0.559117	0.303828

Out[10]:

In [11]:

```
# Get model ids for all models in the AutoML Leaderboard
model_ids = list(aml.leaderboard['model_id'].as_data_frame().iloc[:,0])
# Get the "All Models" Stacked Ensemble model
se = h2o.get_model([mid for mid in model_ids if "StackedEnsemble_AllModels" in mid][0])
# Get the Stacked Ensemble metalearner model
metalearner = se.metalearner()
```

In [12]:

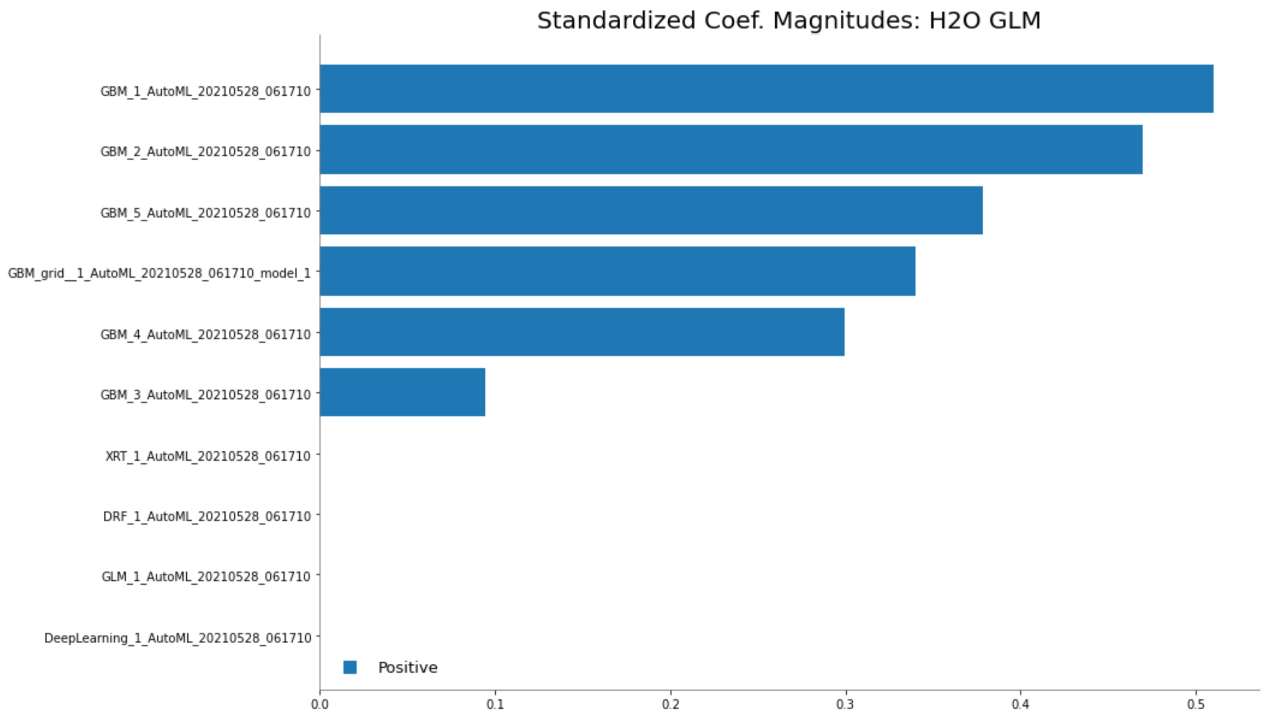
```
metalearner.coef_norm()
```

Out[12]:

```
{'Intercept': -1.5513084495297385,
 'GBM_1_AutoML_20210528_061710': 0.5105255495423108,
 'GBM_2_AutoML_20210528_061710': 0.46958082297640646,
 'GBM_grid__1_AutoML_20210528_061710_model_1': 0.3403355892744305,
 'GBM_5_AutoML_20210528_061710': 0.3784814600766142,
 'GBM_3_AutoML_20210528_061710': 0.09444636609576594,
 'GBM_4_AutoML_20210528_061710': 0.29942092112786983,
```

```
 'XRT_1_AutoML_20210528_061710': 0.0,
 'DRF_1_AutoML_20210528_061710': 0.0,
 'GLM_1_AutoML_20210528_061710': 0.0,
 'DeepLearning_1_AutoML_20210528_061710': 0.0}
```

In [13]:
```
%matplotlib inline
metalearner.std_coef_plot()
```



In [18]:
```
gbm = h2o.get_model([mid for mid in model_ids if "GBM" in mid][0])
print(gbm)
```

```
Model Details
=============
H2OGradientBoostingEstimator :  Gradient Boosting Machine
Model Key:  GBM_1_AutoML_20210528_061710


Model Summary:
```

	number_of_trees	number_of_internal_trees	model_size_in_bytes	min_depth	max_depth	mean_deptl
0	81.0	81.0	57885.0	6.0	6.0	6.

```
ModelMetricsBinomial: gbm
** Reported on train data. **

MSE: 0.10043555525361032
RMSE: 0.31691569108141415
LogLoss: 0.3235122611131131
Mean Per-Class Error: 0.144020838279167
AUC: 0.9339457207041229
AUCPR: 0.860621216295739
Gini: 0.8678914414082457

Confusion Matrix (Act/Pred) for max f1 @ threshold = 0.3957987447464135:
```

		false	true	Error	Rate
0	false	11012.0	1639.0	0.1296	(1639.0/12651.0)
1	true	889.0	4488.0	0.1653	(889.0/5377.0)
2	Total	11901.0	6127.0	0.1402	(2528.0/18028.0)

```
Maximum Metrics: Maximum metrics at their respective thresholds
```

	metric	threshold	value	idx
0	max f1	0.395799	0.780250	201.0
1	max f2	0.263771	0.845464	256.0
2	max f0point5	0.553802	0.798050	134.0
3	max accuracy	0.452559	0.865099	177.0
4	max precision	0.969915	1.000000	0.0
5	max recall	0.030815	1.000000	374.0
6	max specificity	0.969915	1.000000	0.0
7	max absolute_mcc	0.415246	0.681859	193.0
8	max min_per_class_accuracy	0.374371	0.854083	210.0
9	max mean_per_class_accuracy	0.363427	0.855979	214.0
10	max tns	0.969915	12651.000000	0.0
11	max fns	0.969915	5373.000000	0.0
12	max fps	0.004492	12651.000000	399.0
13	max tps	0.030815	5377.000000	374.0
14	max tnr	0.969915	1.000000	0.0
15	max fnr	0.969915	0.999256	0.0
16	max fpr	0.004492	1.000000	399.0
17	max tpr	0.030815	1.000000	374.0

```
Gains/Lift Table: Avg response rate: 29.83 %, avg score: 29.82 %
```

	group	cumulative_data_fraction	lower_threshold	lift	cumulative_lift	response_rate	score
0	1	0.010040	0.901151	3.352799	3.352799	1.000000	0.925328
1	2	0.020024	0.871201	3.315546	3.334224	0.988889	0.885018
2	3	0.030009	0.848312	3.241039	3.303220	0.966667	0.859313
3	4	0.040049	0.825817	3.204609	3.278499	0.955801	0.836330
4	5	0.050033	0.806206	3.296919	3.282175	0.983333	0.815322
5	6	0.100011	0.725528	3.077430	3.179859	0.917869	0.765024
6	7	0.150044	0.654739	2.877014	3.078873	0.858093	0.690183
7	8	0.200022	0.581294	2.482039	2.929748	0.740289	0.618011

	group	cumulative_data_fraction	lower_threshold	lift	cumulative_lift	response_rate	score
8	9	0.300033	0.449254	1.906056	2.588517	0.568497	0.514883
9	10	0.399989	0.322884	1.175898	2.235509	0.350721	0.383975
10	11	0.500000	0.216594	0.619236	1.912219	0.184692	0.268179
11	12	0.600011	0.133813	0.316126	1.646179	0.094287	0.172895
12	13	0.699967	0.067562	0.104194	1.425982	0.031077	0.098390
13	14	0.799978	0.029232	0.018596	1.250035	0.005546	0.045841
14	15	0.899989	0.013605	0.000000	1.111125	0.000000	0.020177
15	16	1.000000	0.003708	0.000000	1.000000	0.000000	0.009175

```
ModelMetricsBinomial: gbm
** Reported on cross-validation data. **

MSE: 0.13547005586359534
RMSE: 0.36806257058222497
LogLoss: 0.41036321129186654
Mean Per-Class Error: 0.21941240313571475
AUC: 0.8641679774825592
AUCPR: 0.7136088615519571
Gini: 0.7283359549651185

Confusion Matrix (Act/Pred) for max f1 @ threshold = 0.3238127858709052:
```

		false	true	Error	Rate
0	false	9662.0	2989.0	0.2363	(2989.0/12651.0)
1	true	1093.0	4284.0	0.2033	(1093.0/5377.0)
2	Total	10755.0	7273.0	0.2264	(4082.0/18028.0)

```
Maximum Metrics: Maximum metrics at their respective thresholds
```

	metric	threshold	value	idx
0	max f1	0.323813	0.677312	231.0
1	max f2	0.156366	0.789721	307.0
2	max f0point5	0.485704	0.669430	163.0
3	max accuracy	0.485704	0.800089	163.0
4	max precision	0.978521	1.000000	0.0
5	max recall	0.015320	1.000000	388.0
6	max specificity	0.978521	1.000000	0.0
7	max absolute_mcc	0.354775	0.524296	218.0
8	max min_per_class_accuracy	0.338593	0.776539	225.0
9	max mean_per_class_accuracy	0.293433	0.780588	245.0
10	max tns	0.978521	12651.000000	0.0

	metric	threshold	value	idx
11	max fns	0.978521	5376.000000	0.0
12	max fps	0.004891	12651.000000	399.0
13	max tps	0.015320	5377.000000	388.0
14	max tnr	0.978521	1.000000	0.0
15	max fnr	0.978521	0.999814	0.0
16	max fpr	0.004891	1.000000	399.0
17	max tpr	0.015320	1.000000	388.0

```
Gains/Lift Table: Avg response rate: 29.83 %, avg score: 29.53 %
```

	group	cumulative_data_fraction	lower_threshold	lift	cumulative_lift	response_rate	score
0	1	0.010040	0.897360	3.204609	3.204609	0.955801	0.922177
1	2	0.020024	0.862566	2.905759	3.055598	0.866667	0.877426
2	3	0.030009	0.835535	2.943012	3.018139	0.877778	0.848007
3	4	0.040049	0.812713	2.815610	2.967366	0.839779	0.824251
4	5	0.050033	0.794527	2.887132	2.951355	0.861111	0.803615
5	6	0.100011	0.702162	2.467154	2.709389	0.735849	0.745913
6	7	0.150044	0.631082	2.271131	2.563249	0.677384	0.665807
7	8	0.200022	0.563882	2.072707	2.440681	0.618202	0.598294
8	9	0.300033	0.437947	1.742414	2.207926	0.519689	0.499244
9	10	0.399989	0.326062	1.302419	1.981643	0.388457	0.380038
10	11	0.500000	0.224432	0.881435	1.761577	0.262895	0.273726
11	12	0.600011	0.139814	0.669444	1.579538	0.199667	0.180544
12	13	0.699967	0.072736	0.377702	1.407915	0.112653	0.104769
13	14	0.799978	0.031845	0.133889	1.248640	0.039933	0.049954
14	15	0.899989	0.015102	0.011157	1.111125	0.003328	0.022306
15	16	1.000000	0.003255	0.000000	1.000000	0.000000	0.010081

```
Cross-Validation Metrics Summary:
```

		mean	sd	cv_1_valid	cv_2_valid	cv_3_valid	cv_4_valid	
0	accuracy	0.78067493	0.010983764	0.7698281	0.77066	0.7936772	0.79029125	0
1	auc	0.8642062	0.0054015317	0.86432254	0.8550114	0.867677	0.8684897	
2	err	0.21932505	0.010983764	0.23017193	0.22933999	0.20632279	0.20970874	0
3	err_count	790.8	39.644672	830.0	827.0	744.0	756.0	
4	f0point5	0.6314402	0.016336285	0.61463416	0.6222425	0.65323085	0.6439571	

		mean	sd	cv_1_valid	cv_2_valid	cv_3_valid	cv_4_valid	
5	f1	0.67986244	0.010331308	0.6800308	0.6642306	0.6912863	0.6865672	
6	f2	0.73681825	0.017454185	0.7610009	0.71229535	0.73405004	0.7352158	0
7	lift_top_group	3.2269464	0.15848134	3.1055498	3.2100096	3.045608	3.359739	
8	logloss	0.41036272	0.0072479476	0.40887854	0.42308584	0.40788755	0.40510845	0
9	max_per_class_error	0.23894164	0.013036208	0.25403702	0.2516011	0.234375	0.22833177	0
10	mcc	0.5280259	0.016884942	0.5289355	0.5006719	0.54429764	0.53922886	
11	mean_per_class_accuracy	0.7807899	0.009282224	0.7862898	0.76437056	0.78571165	0.78492576	
12	mean_per_class_error	0.2192101	0.009282224	0.21371017	0.23562944	0.21428837	0.21507426	
13	mse	0.13546981	0.0029535464	0.13520809	0.14049165	0.13505216	0.13325684	0
14	pr_auc	0.71393496	0.005750112	0.7086458	0.707401	0.7154973	0.7208689	
15	precision	0.602962	0.021094475	0.57760316	0.5970803	0.6301059	0.6183719	
16	r2	0.35273805	0.01082751	0.35102645	0.3348964	0.35898474	0.36256394	0
17	recall	0.7807951	0.0299197	0.8266167	0.7483989	0.765625	0.7716682	
18	rmse	0.36804494	0.0039903303	0.36770654	0.37482217	0.36749443	0.3650436	0
19	specificity	0.78078467	0.023410624	0.745963	0.7803422	0.80579823	0.79818326	

```
Scoring History:
```

	timestamp	duration	number_of_trees	training_rmse	training_logloss	training_auc	training_pr_au
0	2021-05-28 06:18:05	11.065 sec	0.0	0.457493	0.609381	0.500000	0.2982
1	2021-05-28 06:18:05	11.270 sec	5.0	0.416537	0.523970	0.856017	0.6987
2	2021-05-28 06:18:05	11.411 sec	10.0	0.393472	0.477137	0.871462	0.7242
3	2021-05-28 06:18:05	11.625 sec	15.0	0.377093	0.442998	0.882104	0.7495
4	2021-05-28 06:18:05	11.753 sec	20.0	0.367003	0.421426	0.888843	0.7630
5	2021-05-28 06:18:06	11.888 sec	25.0	0.359184	0.404474	0.894401	0.7754
6	2021-05-28 06:18:06	12.018 sec	30.0	0.353004	0.391459	0.899891	0.7875
7	2021-05-28 06:18:06	12.147 sec	35.0	0.348575	0.382768	0.903974	0.7961
8	2021-05-28 06:18:06	12.273 sec	40.0	0.344660	0.374886	0.907749	0.8041
9	2021-05-28 06:18:06	12.434 sec	45.0	0.340521	0.366383	0.911266	0.8122

	timestamp	duration	number_of_trees	training_rmse	training_logloss	training_auc	training_pr_au
10	2021-05-28 06:18:06	12.568 sec	50.0	0.336431	0.359107	0.915530	0.8217
11	2021-05-28 06:18:06	12.696 sec	55.0	0.333063	0.352487	0.918756	0.8282
12	2021-05-28 06:18:07	12.825 sec	60.0	0.329617	0.346242	0.922037	0.8351
13	2021-05-28 06:18:07	12.961 sec	65.0	0.326021	0.340007	0.925507	0.8427
14	2021-05-28 06:18:07	13.086 sec	70.0	0.323053	0.334475	0.928334	0.8487
15	2021-05-28 06:18:07	13.208 sec	75.0	0.320467	0.329565	0.930519	0.8534
16	2021-05-28 06:18:07	13.330 sec	80.0	0.317398	0.324319	0.933521	0.8597
17	2021-05-28 06:18:07	13.381 sec	81.0	0.316916	0.323512	0.933946	0.8606

```
Variable Importances:
```

	variable	relative_importance	scaled_importance	percentage
0	Percent_Days_Employed	2207.446533	1.000000	0.240241
1	Jobs_Per_Year	1596.135376	0.723069	0.173710
2	Delinquency_Reports	565.253235	0.256067	0.061518
3	Avg_Days_per_DrugTest	447.108673	0.202546	0.048660
4	Age_at_Release	438.510162	0.198650	0.047724
5	DrugTests_THC_Positive	379.995056	0.172142	0.041356
6	Residence_Changes	345.457336	0.156496	0.037597
7	Program_Attendances	269.460999	0.122069	0.029326
8	Prior_Arrest_Episodes_Felony	259.170441	0.117407	0.028206
9	Gang_Affiliated	241.904129	0.109585	0.026327
10	Supervision_Risk_Score_First	231.313385	0.104788	0.025174
11	Prior_Arrest_Episodes_PPViolationCharges	211.192078	0.095673	0.022984
12	DrugTests_Meth_Positive	201.253433	0.091170	0.021903
13	Prior_Arrest_Episodes_Property	195.097519	0.088382	0.021233
14	Residence_PUMA	136.996063	0.062061	0.014910
15	Prison_Offense	124.228493	0.056277	0.013520
16	Prior_Arrest_Episodes_Misd	120.916656	0.054777	0.013160
17	Supervision_Level_First	103.752914	0.047001	0.011292

	variable	relative_importance	scaled_importance	percentage
18	Prior_Conviction_Episodes_Misd	97.146889	0.044009	0.010573
19	Prison_Years	80.273933	0.036365	0.008736

```
See the whole table with table.as_data_frame()
```