BITI2513: INTRODUCTION TO DATA SCIENCE

ASSIGNMENT 1

PREDICTION OF RECIDIVISM

GROUP MEMBER	MATRIC NO
CHEING JIN XIAN	B032010431
LIEW SZE WEN	B032010178
LOO HEN SHEN	B032010149
MICHELLE TANG	B032010115

1.0 Aims of AI

The goal of recidivism prediction using AI is to assist in reducing the number of inmates who reoffend after being released from custody by identifying those who are more likely to do so and offering them specialized rehabilitation programmes and support services. It is intended that by doing this, the rates of recidivism can be decreased, enhancing public safety, lightening the load on the criminal justice system, and facilitating the effective reintegration of more persons who have served time in prison.

2.0 How to Deploy Our Prediction

1. **Read the dataset.**

```
[ ]  df= pd.read_csv("3-Year_Recidivism_for_Offenders_Released_from_Prison_in_Iowa_elaborated.csv")
     #print(df)
```

Output:

	Fiscal Year Released	Recidivism Reporting Year	Race - Ethnicity	Age At Release	Convicting Offense Classification	Convicting Offense Type	Convicting Offense Subtype	Main Supervising District	Release Type	Release Type: Paroled to Detainer united	Part of Target Population	Recidivism - Return to Prison numeric
0	2010	2013	White - Non-Hispanic	Under 25	D Felony	Violent	Assault	4JD	Parole	Parole	Yes	1
1	2010	2013	White - Non-Hispanic	55 and Older	D Felony	Public Order	OWI	7JD	Parole	Parole	Yes	1
2	2010	2013	White - Non-Hispanic	25-34	D Felony	Property	Burglary	6JD	Parole	Parole	Yes	1
3	2010	2013	White - Non-Hispanic	55 and Older	C Felony	Drug	Trafficking	8JD	Parole	Parole	Yes	1
4	2010	2013	Black - Non-Hispanic	25-34	D Felony	Drug	Trafficking	5JD	Parole	Parole	Yes	1
...	...	...	...	...	...	...	...	...	...	...	...	...
95	2010	2013	White - Non-Hispanic	55 and Older	D Felony	Public Order	OWI	4JD	Parole	Parole	Yes	1
96	2010	2013	White - Non-Hispanic	45-54	Felony - Enhanced	Property	Theft	6JD	Parole	Parole	Yes	1
97	2010	2013	White - Non-Hispanic	Under 25	C Felony	Violent	Sex	8JD	Special Sentence	Special Sentence	Yes	1
98	2010	2013	White - Non-Hispanic	25-34	D Felony	Property	Burglary	1JD	Parole	Parole	Yes	1
99	2010	2013	White - Non-Hispanic	25-34	Aggravated Misdemeanor	Property	Theft	5JD	Parole	Parole	Yes	1

2. **Data Cleaning**

   i. Drop the unnecessary column data.

```
data = df.drop(['Fiscal Year Released','Recidivism Reporting Year'],axis='columns')
data.head(100)
```

Output:

	Race - Ethnicity	Age At Release	Convicting Offense Classification	Convicting Offense Type	Convicting Offense Subtype	Main Supervising District	Release Type	Release type: Paroled to Detainer united	Part of Target Population	Recidivism - Return to Prison numeric
0	White - Non-Hispanic	Under 25	D Felony	Violent	Assault	4JD	Parole	Parole	Yes	1
1	White - Non-Hispanic	55 and Older	D Felony	Public Order	OWI	7JD	Parole	Parole	Yes	1
2	White - Non-Hispanic	25-34	D Felony	Property	Burglary	5JD	Parole	Parole	Yes	1
3	White - Non-Hispanic	55 and Older	C Felony	Drug	Trafficking	8JD	Parole	Parole	Yes	1
4	Black - Non-Hispanic	25-34	D Felony	Drug	Trafficking	5JD	Parole	Parole	Yes	1
...	...	...	...	...	...	...	...	...	...	...
95	White - Non-Hispanic	55 and Older	D Felony	Public Order	OWI	4JD	Parole	Parole	Yes	1
96	White - Non-Hispanic	45-54	Felony - Enhanced	Property	Theft	5JD	Parole	Parole	Yes	1
97	White - Non-Hispanic	Under 25	C Felony	Violent	Sex	8JD	Special Sentence	Special Sentence	Yes	1
98	White - Non-Hispanic	25-34	D Felony	Property	Burglary	1JD	Parole	Parole	Yes	1
99	White - Non-Hispanic	25-34	Aggravated Misdemeanor	Property	Theft	5JD	Parole	Parole	Yes	1

ii. Split the column.

```
data [['Race','a','b']] = data["Race - Ethnicity"].apply(lambda x:pd.Series(str(x).split("-"))) #split one into  columns
data['Ethnicity']=data["a"]+data["b"] #combine 2 columns into 1 column
data=data.drop(["Race - Ethnicity","a","b"],axis='columns') #axis = columns or 1
data.head(200)
```

iii. Check the null data from the dataset.

```
[ ]  data['Race'].unique()
```

Output:

```
array(['White ', 'Black ', 'American Indian or Alaska Native ',
       'Asian or Pacific Islander ', 'N/A ', 'nan'], dtype=object)
```

iv. Set data's null default to the same default value of null.

```
[ ]  data=data.replace(['nan','N/A '],np.NaN)
```

v. Replace the null data into the data.

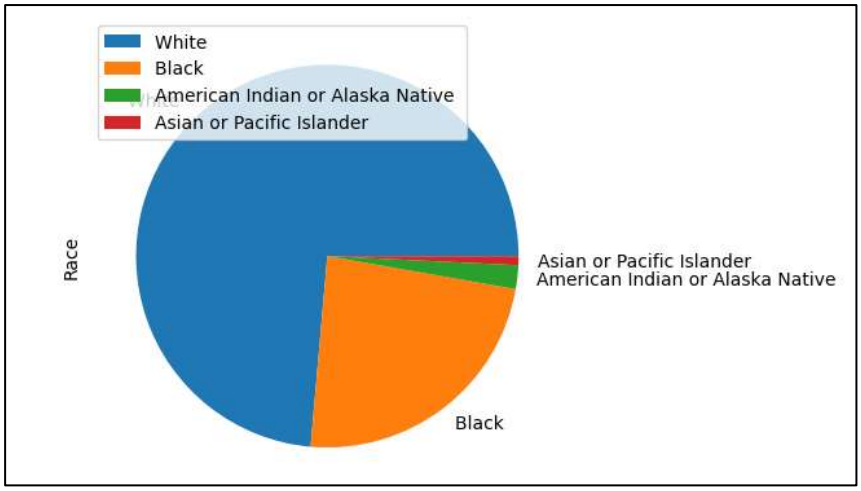
```
[ ]  data.ffill(inplace=True)
```

3. Data Analysis

i. Create the pie bar and bar for analysis the column data.

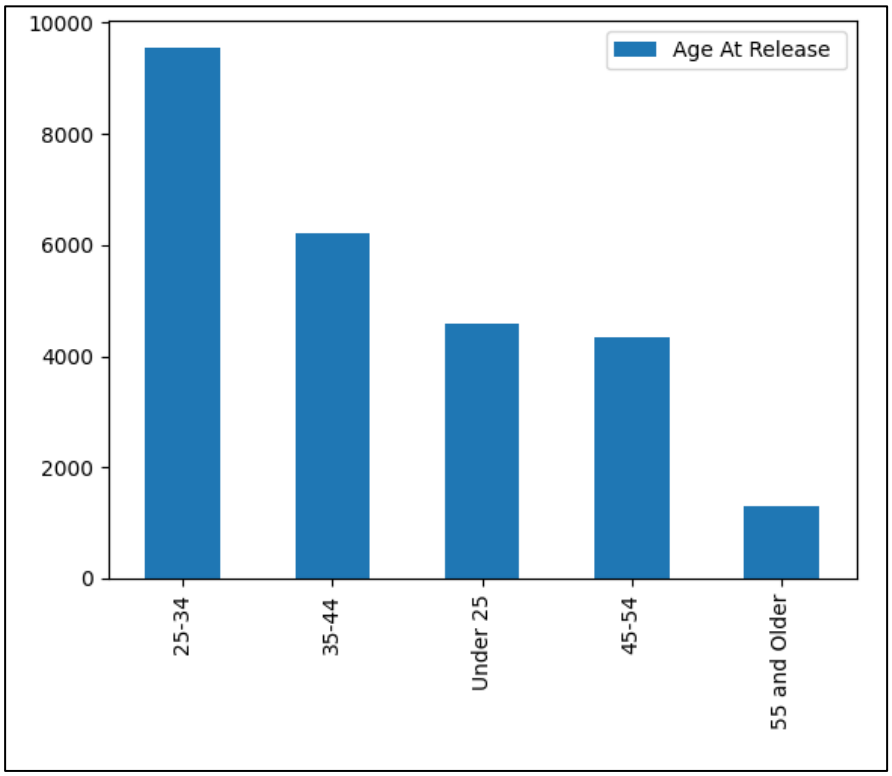
```
data.Race.value_counts().plot(kind='pie').legend()
plt.legend(loc='best')
```

Output:




```
[13] data["Age At Release "].value_counts().plot(kind='bar').legend()
     plt.legend(loc='best')
```

Output:

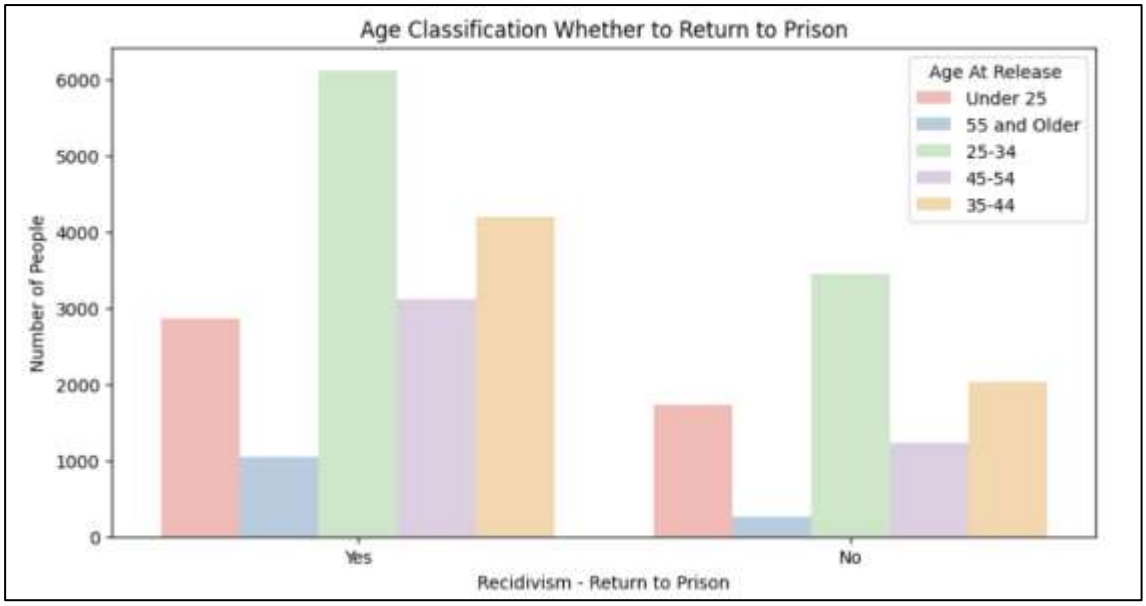


ii. Create histogram graphs for data analysis using the seaborn.

```
plt.figure(figsize=(10,5))
sns.countplot(x='Recidivism - Return to Prison numeric', hue='Age At Release ',data=data, palette="Pastel1", )
plt.xticks([0, 1], ['Yes', 'No'])
plt.title("Age Classification Whether to Return to Prison")
plt.xlabel('Recidivism - Return to Prison')
plt.ylabel('Number of People')

plt.show
```

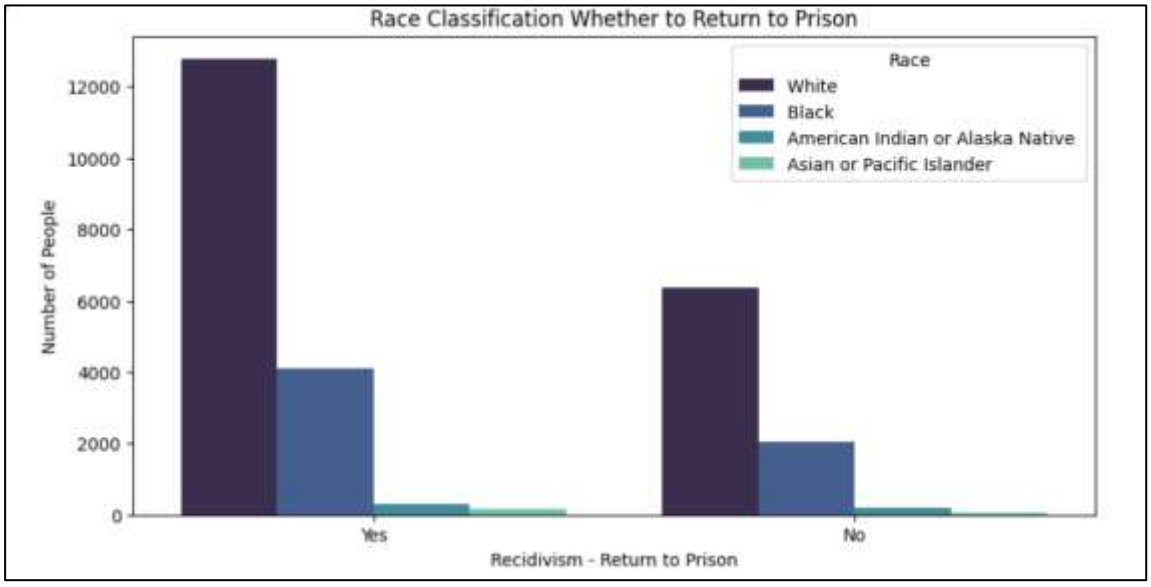
Output:




```
plt.figure(figsize=(10,5))
sns.countplot(x='Recidivism - Return to Prison numeric', hue='Race',data=data, palette="mako")
plt.xticks([0, 1], ['Yes', 'No'])
plt.title("Race Classification Whether to Return to Prison")
plt.xlabel('Recidivism - Return to Prison')
plt.ylabel('Number of People')

plt.show
```

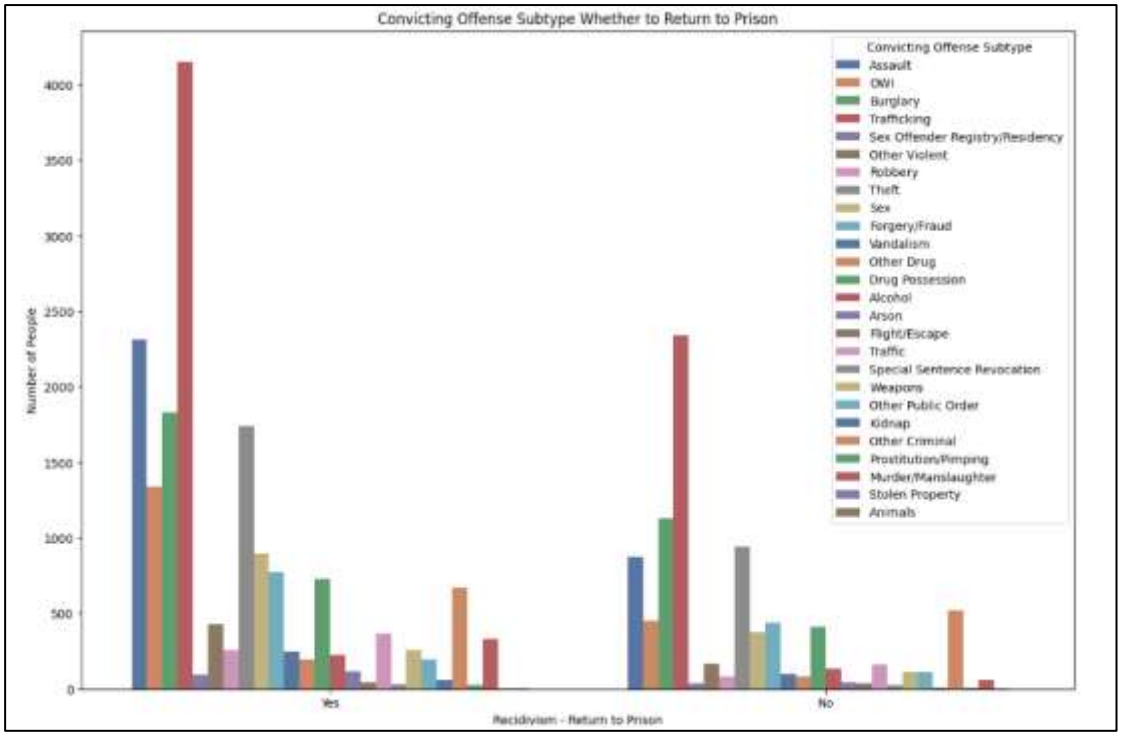
Output:




```
plt.figure(figsize=(15,10))
sns.countplot(x='Recidivism - Return to Prison numeric', hue='Convicting Offense Subtype',data=data, palette="deep")
plt.xticks([0, 1], ['Yes', 'No'])
plt.title("Convicting Offense Subtype Whether to Return to Prison")
plt.xlabel('Recidivism - Return to Prison')
plt.ylabel('Number of People')

plt.show
```

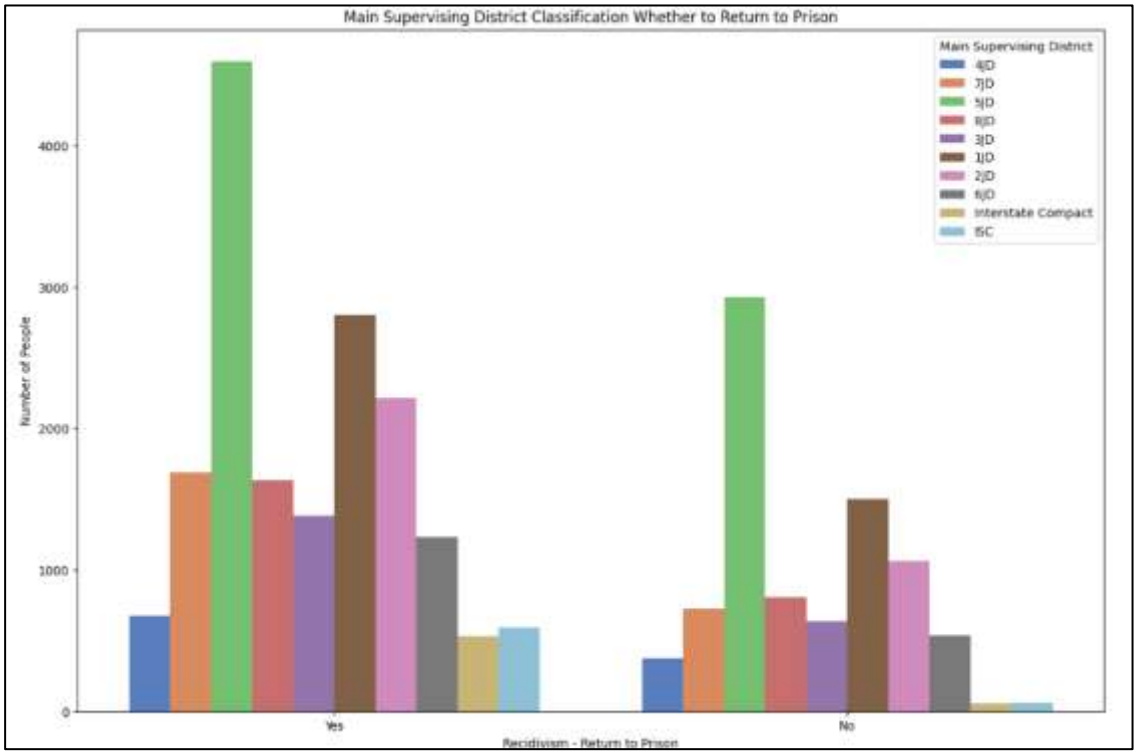
Output:




```
plt.figure(figsize=(15,10))
sns.countplot(x='Recidivism - Return to Prison numeric', hue='Main Supervising District',data=data, palette="muted")
plt.xticks([0, 1], ['Yes', 'No'])
plt.title("Main Supervising District Classification Whether to Return to Prison")
plt.xlabel('Recidivism - Return to Prison')
plt.ylabel('Number of People')

plt.show
```

Output:



4. **Checking the data whether suitable training in machine learning.**

i. Check the type of each column from the dataset.

```
data.dtypes
```

Output:

```
Age At Release                                  object
Convicting Offense Classification               object
Convicting Offense Type                         object
Convicting Offense Subtype                      object
Main Supervising District                       object
Release type: Paroled to Detainder united       object
Part of Target Population                       object
Recidivism - Return to Prison numeric            int64
Race                                            object
dtype: object
```

ii. Convert it into numeric data because the machine learning model will be easy to recognize and trained.

```
race_mapping={'White ':1, 'Black ':2, 'American Indian or Alaska Native ':3,'Asian or Pacific Islander ':4}
age_mapping ={'Under 25':5, '55 and Older':6, '25-34':7, '45-54':8, '35-44':9}
classification_mapping={'D Felony':10, 'C Felony':11, 'Aggravated Misdemeanor':12,
       'Felony - Enhanced':13, 'B Felony':14, 'Serious Misdemeanor':15,
       'Simple Misdemeanor':16, 'Other Felony (Old Code)':17,
       'Sexual Predator Community Supervision':18,
       'Felony - Enhancement to Original Penalty':19,
       'Special Sentence 2005':20, 'Felony - Mandatory Minimum':21, 'A Felony':22,
       'Other Felony':23, 'Other Misdemeanor':24}
convtype_mapping={'Violent':25, 'Public Order':26, 'Property':27, 'Drug':28, 'Other':29}
convsubtype_mapping={'Assault':30, 'OWI':31, 'Burglary':32, 'Trafficking':33,
       'Sex Offender Registry/Residency':34, 'Other Violent':35, 'Robbery':36,
       'Theft':37, 'Sex':38, 'Forgery/Fraud':39, 'Vandalism':40, 'Other Drug':41,
       'Drug Possession':42, 'Alcohol':43, 'Arson':44, 'Flight/Escape':45, 'Traffic':46,
       'Special Sentence Revocation':47, 'Weapons':48, 'Other Public Order':49,
       'Kidnap':50, 'Other Criminal':51, 'Prostitution/Pimping':52,
       'Murder/Manslaughter':53, 'Stolen Property':54, 'Animals':55}
detained_mapping={'Parole':56, 'Discharged End of Sentence':57, 'Special Sentence':58,
       'Paroled to Detainer':59}
main_mapping={'4JD':60, '7JD':61, '5JD':62, '8JD':63, '3JD':64, '1JD':65, '2JD':66, '6JD':67,
       'Interstate Compact':68, 'ISC':69}
target_mapping={'Yes':70, 'No':71}

data['Race']=data['Race'].map(race_mapping)
data['Age At Release ']=data['Age At Release '].map(age_mapping)
data['Convicting Offense Classification']=data['Convicting Offense Classification'].map(classification_mapping)
data['Convicting Offense Type']=data['Convicting Offense Type'].map(convtype_mapping)
data['Convicting Offense Subtype']=data['Convicting Offense Subtype'].map(convsubtype_mapping)
data['Release type: Paroled to Detainder united']=data['Release type: Paroled to Detainder united'].map(detained_mapping)
data['Main Supervising District']=data['Main Supervising District'].map(main_mapping)
data['Part of Target Population']=data['Part of Target Population'].map(target_mapping)
data
```

Output:

	Age At Release	Convicting Offense Classification	Convicting Offense Type	Convicting Offense Subtype	Main Supervising District	Release type: Paroled to Detainder united	Part of Target Population	Recidivism - Return to Prison numeric	Race
0	5	10	25	30	60	56	70	1	1
1	6	10	26	31	61	56	70	1	1
2	7	10	27	32	62	56	70	1	1
3	6	11	28	33	63	56	70	1	1
4	7	10	28	33	64	56	70	1	2
...	...	...	...	...	...	...	...	...	...
26015	7	11	25	53	61	57	70	0	1
26016	7	10	27	40	62	59	71	0	1
26017	5	12	25	35	65	57	70	0	2
26018	5	10	28	41	62	56	71	0	1
26019	9	11	25	35	62	56	71	0	1

iii. Check the type of data.

```
Age At Release                                 int64
Convicting Offense Classification              int64
Convicting Offense Type                        int64
Convicting Offense Subtype                     int64
Main Supervising District                      int64
Release type: Paroled to Detainder united      int64
Part of Target Population                      int64
Recidivism - Return to Prison numeric          int64
Race                                           int64
dtype: object
```

5. Recidivism Prediction using Support Vector Classifier (SVC).

i. Create the training and test.

```
X = data.drop("Recidivism - Return to Prison numeric", axis='columns')
y = data["Recidivism - Return to Prison numeric"] #to predict y using X
# Create training and test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1, stratify=y)
```

ii. Instantiate the Support Vector Classifier.

```
sc = StandardScaler()
sc.fit(X_train)
X_train_std = sc.transform(X_train)
X_test_std = sc.transform(X_test)
# Instantiate the Support Vector Classifier (SVC)
svc = SVC(C=1.0, random_state=1, kernel='linear')
# Fit the model
svc.fit(X_train_std, y_train)
```

iii. Do the prediction using SVC.

```
y_preds = svc.predict(X_test)
```

iv. Check the train data score.

```
svc.score(X_train, y_train)
```

Output:

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:432: UserWarning: X has feature names, but SVC was fitted without feature names
  warnings.warn(
0.6663555506753047
```

v. Check the test data score.

```
[37] svc.score(X_test, y_test)
```

Output:

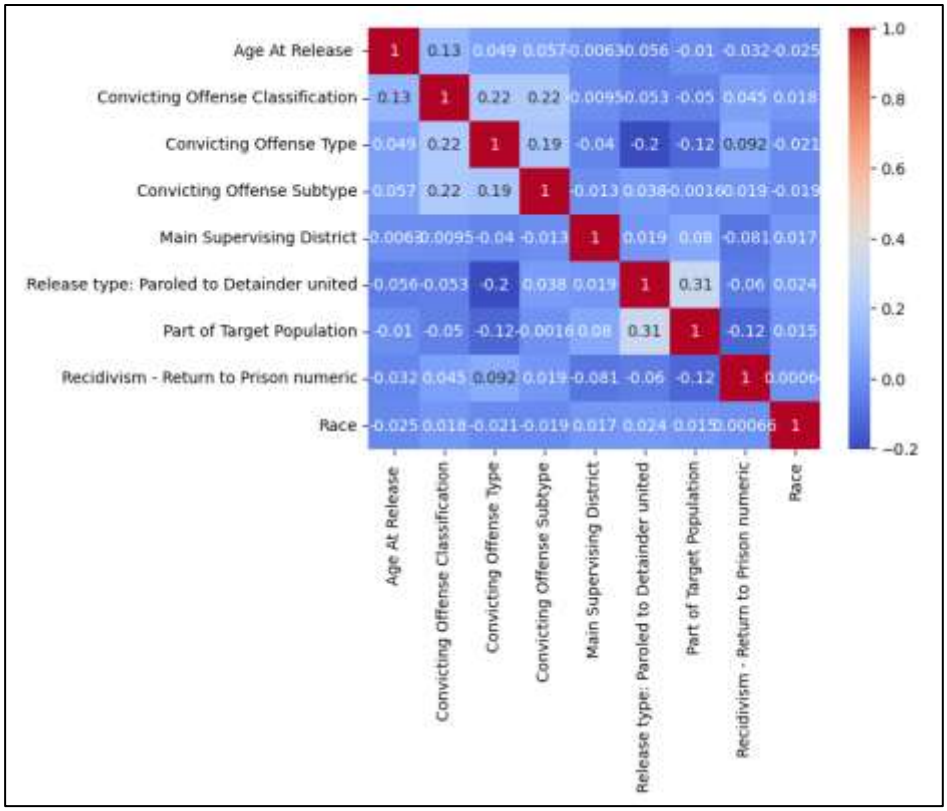
```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:432: UserWarning: X has feature names, but SVC was fitted without feature names
  warnings.warn(
0.6664104534973098
```

vi. Do the confusion matrix.

```
corr_matrix=data.corr()
```

```
[39] sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
     plt.show()
```

Output:



6. Deploy the prediction report. (Accuracy: 0.67)

```
print(classification_report(y_test, y_preds))
```

Output:

```
              precision    recall  f1-score   support

           0       0.67      1.00      0.80      5202
           1       0.00      0.00      0.00      2604

    accuracy                           0.67      7806
   macro avg       0.33      0.50      0.40      7806
weighted avg       0.44      0.67      0.53      7806
```

3.0 Dataset

Kaggle: 3 Year Recidivism for Offenders Released from Prison in Iowa

Link:

https://www.kaggle.com/slonnadube/recidivism-for-offenders-released-from-prison/activity?select=3-Year_Recidivism_for_Offenders_Released_from_Prison_in_Iowa_elaborated.csv

This data set is on re-offending in a 3-year period after an initial release from prison, among offenders serving a prison term in the State of Iowa, US between 2010 and 2015, with recidivism follow-up between 2013 and 2018.

The variables in the dataset include:

- Fiscal Year Released
- Recidivism Reporting Year
- Race – Ethnicity
- Age at Release

- Convicting Offense Classification
- Convicting Offense Type
- Convicting Offense Subtype
- Main Supervising District
- Release Type
- Part of Target Population

4.0 Tools and Programming Language

Google Colaboratory and Python