Crime Prediction Using Machine Learning

Crime prediction is an emerging field of study that uses data analysis and machine learning to predict criminal behavior and prevent crime.

Predictive models use advanced algorithms to identify patterns in criminal behavior and take into account external factors to make more accurate predictions.

Predictive models can help law enforcement agencies allocate resources, prevent crime, and improve public safety, as well as help investigators identify potential suspects and gather evidence more efficiently.

Researchers and practitioners must be aware of ethical and legal concerns surrounding crime prediction, such as potential biases and privacy and civil liberties, and address them through transparent and ethical practices.

Training Dataset Information

For more information click on the following link :

Dataset Info

Attribute Information:

Attribute Information: (122 predictive, 5 non-predictive, 1 goal)

-- state: US state (by number) - not counted as predictive above, but if considered, should be consided nominal (nominal)

-- county: numeric code for county - not predictive, and many missing values (numeric)

-- community: numeric code for community - not predictive and many missing values (numeric)

-- communityname: community name - not predictive - for information only (string)

-- fold: fold number for non-random 10 fold cross validation, potentially useful for debugging, paired tests - not predictive (numeric)

-- population: population for community: (numeric - decimal)

-- householdsize: mean people per household (numeric - decimal)

-- racepctblack: percentage of population that is african american (numeric - decimal)

-- racePctWhite: percentage of population that is caucasian (numeric - decimal)

-- racePctAsian: percentage of population that is of asian heritage (numeric - decimal)

-- racePctHisp: percentage of population that is of hispanic heritage (numeric - decimal)

-- agePct12t21: percentage of population that is 12-21 in age (numeric - decimal)

-- agePct12t29: percentage of population that is 12-29 in age (numeric - decimal)

-- agePct16t24: percentage of population that is 16-24 in age (numeric - decimal)

-- agePct65up: percentage of population that is 65 and over in age (numeric - decimal)

-- numbUrban: number of people living in areas classified as urban (numeric - decimal)

```
-- pctUrban: percentage of people living in areas classified as urban (numeric - decimal)

-- medIncome: median household income (numeric - decimal)

-- pctWWage: percentage of households with wage or salary income in 1989 (numeric - decimal)

-- pctWFarmSelf: percentage of households with farm or self employment income in 1989 (numeric - decimal)

-- pctWInvInc: percentage of households with investment / rent income in 1989 (numeric - decimal)

-- pctWSocSec: percentage of households with social security income in 1989 (numeric - decimal)

-- pctWPubAsst: percentage of households with public assistance income in 1989 (numeric - decimal)

-- pctWRetire: percentage of households with retirement income in 1989 (numeric - decimal)

-- medFamInc: median family income (differs from household income for non-family households) (numeric - decimal)

-- perCapInc: per capita income (numeric - decimal)

-- whitePerCap: per capita income for caucasians (numeric - decimal)

-- blackPerCap: per capita income for african americans (numeric - decimal)

-- indianPerCap: per capita income for native americans (numeric - decimal)

-- AsianPerCap: per capita income for people with asian heritage (numeric - decimal)

-- OtherPerCap: per capita income for people with 'other' heritage (numeric - decimal)

-- HispPerCap: per capita income for people with hispanic heritage (numeric - decimal)
```

-- NumUnderPov: number of people under the poverty level (numeric - decimal)

-- PctPopUnderPov: percentage of people under the poverty level (numeric - decimal)

-- PctLess9thGrade: percentage of people 25 and over with less than a 9th grade education (numeric - decimal)

-- PctNotHSGrad: percentage of people 25 and over that are not high school graduates (numeric - decimal)

-- PctBSorMore: percentage of people 25 and over with a bachelors degree or higher education (numeric - decimal)

-- PctUnemployed: percentage of people 16 and over, in the labor force, and unemployed (numeric - decimal)

-- PctEmploy: percentage of people 16 and over who are employed (numeric - decimal)

-- PctEmplManu: percentage of people 16 and over who are employed in manufacturing (numeric - decimal)

-- PctEmplProfServ: percentage of people 16 and over who are employed in professional services (numeric - decimal)

-- PctOccupManu: percentage of people 16 and over who are employed in manufacturing (numeric - decimal) ########

-- PctOccupMgmtProf: percentage of people 16 and over who are employed in management or professional occupations (numeric - decimal)

-- MalePctDivorce: percentage of males who are divorced (numeric - decimal)

-- MalePctNevMarr: percentage of males who have never married (numeric - decimal)

-- FemalePctDiv: percentage of females who are divorced (numeric - decimal)

-- TotalPctDiv: percentage of population who are divorced (numeric - decimal)

-- PersPerFam: mean number of people per family (numeric - decimal)

-- PctFam2Par: percentage of families (with kids) that are headed by two parents (numeric - decimal)

-- PctKids2Par: percentage of kids in family housing with two parents (numeric - decimal)

-- PctYoungKids2Par: percent of kids 4 and under in two parent households (numeric - decimal)

-- PctTeen2Par: percent of kids age 12-17 in two parent households (numeric - decimal)

-- PctWorkMomYoungKids: percentage of moms of kids 6 and under in labor force (numeric - decimal)

-- PctWorkMom: percentage of moms of kids under 18 in labor force (numeric - decimal)

-- NumIlleg: number of kids born to never married (numeric - decimal)

-- PctIlleg: percentage of kids born to never married (numeric - decimal)

-- NumImmig: total number of people known to be foreign born (numeric - decimal)

-- PctImmigRecent: percentage of *immigrants* who immigated within last 3 years (numeric - decimal)

-- PctImmigRec5: percentage of *immigrants* who immigated within last 5 years (numeric - decimal)

-- PctImmigRec8: percentage of *immigrants* who immigated within last 8 years (numeric - decimal)

-- PctImmigRec10: percentage of *immigrants* who immigated within last 10 years (numeric - decimal)

-- PctRecentImmig: percent of *population* who have immigrated within the last 3 years (numeric - decimal)

-- PctRecImmig5: percent of *population* who have immigrated within the last 5 years (numeric - decimal)

-- PctRecImmig8: percent of *population* who have immigrated within the last 8 years (numeric - decimal)

-- PctRecImmig10: percent of *population* who have immigrated within the last 10 years (numeric - decimal)

-- PctSpeakEnglOnly: percent of people who speak only English (numeric - decimal)

-- PctNotSpeakEnglWell: percent of people who do not speak English well (numeric - decimal)

-- PctLargHouseFam: percent of family households that are large (6 or more) (numeric - decimal)

-- PctLargHouseOccup: percent of all occupied households that are large (6 or more people) (numeric - decimal)

-- PersPerOccupHous: mean persons per household (numeric - decimal)

-- PersPerOwnOccHous: mean persons per owner occupied household (numeric - decimal)

-- PersPerRentOccHous: mean persons per rental household (numeric - decimal)

-- PctPersOwnOccup: percent of people in owner occupied households (numeric - decimal)

-- PctPersDenseHous: percent of persons in dense housing (more than 1 person per room) (numeric - decimal)

-- PctHousLess3BR: percent of housing units with less than 3 bedrooms (numeric - decimal)

-- MedNumBR: median number of bedrooms (numeric - decimal)

-- HousVacant: number of vacant households (numeric - decimal)

-- PctHousOccup: percent of housing occupied (numeric - decimal)

-- PctHousOwnOcc: percent of households owner occupied (numeric - decimal)

-- PctVacantBoarded: percent of vacant housing that is boarded up (numeric - decimal)

-- PctVacMore6Mos: percent of vacant housing that has been vacant more than 6 months (numeric - decimal)

-- MedYrHousBuilt: median year housing units built (numeric - decimal)

-- PctHousNoPhone: percent of occupied housing units without phone (in 1990, this was rare!) (numeric - decimal)

-- PctWOFullPlumb: percent of housing without complete plumbing facilities (numeric - decimal)

-- OwnOccLowQuart: owner occupied housing - lower quartile value (numeric - decimal)

-- OwnOccMedVal: owner occupied housing - median value (numeric - decimal)

-- OwnOccHiQuart: owner occupied housing - upper quartile value (numeric - decimal)

-- RentLowQ: rental housing - lower quartile rent (numeric - decimal)

-- RentMedian: rental housing - median rent (Census variable H32B from file STF1A) (numeric - decimal)

-- RentHighQ: rental housing - upper quartile rent (numeric - decimal)

-- MedRent: median gross rent (Census variable H43A from file STF3A - includes utilities) (numeric - decimal)

-- MedRentPctHousInc: median gross rent as a percentage of household income (numeric - decimal)

-- MedOwnCostPctInc: median owners cost as a percentage of household income - for owners with a mortgage (numeric - decimal)

-- MedOwnCostPctIncNoMtg: median owners cost as a percentage of household income - for owners without a mortgage (numeric - decimal)

-- NumInShelters: number of people in homeless shelters (numeric - decimal)

-- NumStreet: number of homeless people counted in the street (numeric - decimal)

-- PctForeignBorn: percent of people foreign born (numeric - decimal)

-- PctBornSameState: percent of people born in the same state as currently living (numeric - decimal)

-- PctSameHouse85: percent of people living in the same house as in 1985 (5 years before) (numeric - decimal)

-- PctSameCity85: percent of people living in the same city as in 1985 (5 years before) (numeric - decimal)

-- PctSameState85: percent of people living in the same state as in 1985 (5 years before) (numeric - decimal)

-- LemasSwornFT: number of sworn full time police officers (numeric - decimal)

-- LemasSwFTPerPop: sworn full time police officers per 100K population (numeric - decimal)

-- LemasSwFTFieldOps: number of sworn full time police officers in field operations (on the street as opposed to administrative etc) (numeric - decimal)

-- LemasSwFTFieldPerPop: sworn full time police officers in field operations (on the street as opposed to administrative etc) per 100K population (numeric - decimal)

-- LemasTotalReq: total requests for police (numeric - decimal)

-- LemasTotReqPerPop: total requests for police per 100K popuation (numeric - decimal)

-- PolicReqPerOffic: total requests for police per police officer (numeric - decimal)

-- PolicPerPop: police officers per 100K population (numeric - decimal)

-- RacialMatchCommPol: a measure of the racial match between the community and the police force. High values indicate proportions in community and police force are similar (numeric - decimal)

-- PctPolicWhite: percent of police that are caucasian (numeric - decimal)

-- PctPolicBlack: percent of police that are african american (numeric - decimal)

-- PctPolicHisp: percent of police that are hispanic (numeric - decimal)

-- PctPolicAsian: percent of police that are asian (numeric - decimal)

-- PctPolicMinor: percent of police that are minority of any kind (numeric - decimal)

-- OfficAssgnDrugUnits: number of officers assigned to special drug units (numeric - decimal)

-- NumKindsDrugsSeiz: number of different kinds of drugs seized (numeric - decimal)

-- PolicAveOTWorked: police average overtime worked (numeric - decimal)

-- LandArea: land area in square miles (numeric - decimal)

-- PopDens: population density in persons per square mile (numeric - decimal)

-- PctUsePubTrans: percent of people using public transit for commuting (numeric - decimal)

-- PolicCars: number of police cars (numeric - decimal)

-- PolicOperBudg: police operating budget (numeric - decimal)

-- LemasPctPolicOnPatr: percent of sworn full time police officers on patrol (numeric - decimal)

-- LemasGangUnitDeploy: gang unit deployed (numeric - decimal - but really ordinal - 0 means NO, 1 means YES, 0.5 means Part Time)

-- LemasPctOfficDrugUn: percent of officers assigned to drug units (numeric - decimal)

-- PolicBudgPerPop: police operating budget per population (numeric - decimal)

-- ViolentCrimesPerPop: total number of violent crimes per 100K popuation (numeric - decimal) GOAL attribute (to be predicted)

Summary Statistics:

Min Max Mean SD Correl Median Mode Missing

population 0 1 0.06 0.13 0.37 0.02 0.01 0

householdsize 0 1 0.46 0.16 -0.03 0.44 0.41 0

racepctblack 0 1 0.18 0.25 0.63 0.06 0.01 0

racePctWhite 0 1 0.75 0.24 -0.68 0.85 0.98 0

racePctAsian 0 1 0.15 0.21 0.04 0.07 0.02 0

racePctHisp 0 1 0.14 0.23 0.29 0.04 0.01 0

agePct12t21 0 1 0.42 0.16 0.06 0.4 0.38 0

agePct12t29 0 1 0.49 0.14 0.15 0.48 0.49 0

agePct16t24 0 1 0.34 0.17 0.10 0.29 0.29 0

agePct65up 0 1 0.42 0.18 0.07 0.42 0.47 0

numbUrban 0 1 0.06 0.13 0.36 0.03 0 0

pctUrban 0 1 0.70 0.44 0.08 1 1 0

medIncome 0 1 0.36 0.21 -0.42 0.32 0.23 0

pctWWage 0 1 0.56 0.18 -0.31 0.56 0.58 0

pctWFarmSelf 0 1 0.29 0.20 -0.15 0.23 0.16 0

pctWInvInc 0 1 0.50 0.18 -0.58 0.48 0.41 0

pctWSocSec 0 1 0.47 0.17 0.12 0.475 0.56 0

pctWPubAsst 0 1 0.32 0.22 0.57 0.26 0.1 0

pctWRetire 0 1 0.48 0.17 -0.10 0.47 0.44 0

medFamInc 0 1 0.38 0.20 -0.44 0.33 0.25 0

perCapInc 0 1 0.35 0.19 -0.35 0.3 0.23 0

whitePerCap 0 1 0.37 0.19 -0.21 0.32 0.3 0

blackPerCap 0 1 0.29 0.17 -0.28 0.25 0.18 0

indianPerCap 0 1 0.20 0.16 -0.09 0.17 0 0

AsianPerCap 0 1 0.32 0.20 -0.16 0.28 0.18 0

OtherPerCap 0 1 0.28 0.19 -0.13 0.25 0 1

HispPerCap 0 1 0.39 0.18 -0.24 0.345 0.3 0

NumUnderPov 0 1 0.06 0.13 0.45 0.02 0.01 0

PctPopUnderPov 0 1 0.30 0.23 0.52 0.25 0.08 0

PctLess9thGrade 0 1 0.32 0.21 0.41 0.27 0.19 0

PctNotHSGrad 0 1 0.38 0.20 0.48 0.36 0.39 0

PctBSorMore 0 1 0.36 0.21 -0.31 0.31 0.18 0

PctUnemployed 0 1 0.36 0.20 0.50 0.32 0.24 0

PctEmploy 0 1 0.50 0.17 -0.33 0.51 0.56 0

PctEmplManu 0 1 0.40 0.20 -0.04 0.37 0.26 0

PctEmplProfServ 0 1 0.44 0.18 -0.07 0.41 0.36 0

PctOccupManu 0 1 0.39 0.20 0.30 0.37 0.32 0

PctOccupMgmtProf 0 1 0.44 0.19 -0.34 0.4 0.36 0

MalePctDivorce 0 1 0.46 0.18 0.53 0.47 0.56 0

MalePctNevMarr 0 1 0.43 0.18 0.30 0.4 0.38 0

FemalePctDiv 0 1 0.49 0.18 0.56 0.5 0.54 0

TotalPctDiv 0 1 0.49 0.18 0.55 0.5 0.57 0

PersPerFam 0 1 0.49 0.15 0.14 0.47 0.44 0

PctFam2Par 0 1 0.61 0.20 -0.71 0.63 0.7 0

PctKids2Par 0 1 0.62 0.21 -0.74 0.64 0.72 0

PctYoungKids2Par 0 1 0.66 0.22 -0.67 0.7 0.91 0

PctTeen2Par 0 1 0.58 0.19 -0.66 0.61 0.6 0

PctWorkMomYoungKids 0 1 0.50 0.17 -0.02 0.51 0.51 0

PctWorkMom 0 1 0.53 0.18 -0.15 0.54 0.57 0

NumIlleg 0 1 0.04 0.11 0.47 0.01 0 0

PctIlleg 0 1 0.25 0.23 0.74 0.17 0.09 0

NumImmig 0 1 0.03 0.09 0.29 0.01 0 0

PctImmigRecent 0 1 0.32 0.22 0.17 0.29 0 0

PctImmigRec5 0 1 0.36 0.21 0.22 0.34 0 0

PctImmigRec8 0 1 0.40 0.20 0.25 0.39 0.26 0

PctImmigRec10 0 1 0.43 0.19 0.29 0.43 0.43 0

PctRecentImmig 0 1 0.18 0.24 0.23 0.09 0.01 0

PctRecImmig5 0 1 0.18 0.24 0.25 0.08 0.02 0

PctRecImmig8 0 1 0.18 0.24 0.25 0.09 0.02 0

PctRecImmig10 0 1 0.18 0.23 0.26 0.09 0.02 0

PctSpeakEnglOnly 0 1 0.79 0.23 -0.24 0.87 0.96 0

PctNotSpeakEnglWell 0 1 0.15 0.22 0.30 0.06 0.03 0

PctLargHouseFam 0 1 0.27 0.20 0.38 0.2 0.17 0

PctLargHouseOccup 0 1 0.25 0.19 0.29 0.19 0.19 0

PersPerOccupHous 0 1 0.46 0.17 -0.04 0.44 0.37 0

PersPerOwnOccHous 0 1 0.49 0.16 -0.12 0.48 0.45 0

PersPerRentOccHous 0 1 0.40 0.19 0.25 0.36 0.32 0

PctPersOwnOccup 0 1 0.56 0.20 -0.53 0.56 0.54 0

PctPersDenseHous 0 1 0.19 0.21 0.45 0.11 0.06 0

PctHousLess3BR 0 1 0.50 0.17 0.47 0.51 0.53 0

MedNumBR 0 1 0.31 0.26 -0.36 0.5 0.5 0

HousVacant 0 1 0.08 0.15 0.42 0.03 0.01 0

PctHousOccup 0 1 0.72 0.19 -0.32 0.77 0.88 0

PctHousOwnOcc 0 1 0.55 0.19 -0.47 0.54 0.52 0

PctVacantBoarded 0 1 0.20 0.22 0.48 0.13 0 0

PctVacMore6Mos 0 1 0.43 0.19 0.02 0.42 0.44 0

MedYrHousBuilt 0 1 0.49 0.23 -0.11 0.52 0 0

PctHousNoPhone 0 1 0.26 0.24 0.49 0.185 0.01 0

PctWOFullPlumb 0 1 0.24 0.21 0.36 0.19 0 0

OwnOccLowQuart 0 1 0.26 0.22 -0.21 0.18 0.09 0

OwnOccMedVal 0 1 0.26 0.23 -0.19 0.17 0.08 0

OwnOccHiQuart 0 1 0.27 0.24 -0.17 0.18 0.08 0

RentLowQ 0 1 0.35 0.22 -0.25 0.31 0.13 0

RentMedian 0 1 0.37 0.21 -0.24 0.33 0.19 0

RentHighQ 0 1 0.42 0.25 -0.23 0.37 1 0

MedRent 0 1 0.38 0.21 -0.24 0.34 0.17 0

MedRentPctHousInc 0 1 0.49 0.17 0.33 0.48 0.4 0

MedOwnCostPctInc 0 1 0.45 0.19 0.06 0.45 0.41 0

MedOwnCostPctIncNoMtg 0 1 0.40 0.19 0.05 0.37 0.24 0

NumInShelters 0 1 0.03 0.10 0.38 0 0 0

NumStreet 0 1 0.02 0.10 0.34 0 0 0

PctForeignBorn 0 1 0.22 0.23 0.19 0.13 0.03 0

PctBornSameState 0 1 0.61 0.20 -0.08 0.63 0.78 0

PctSameHouse85 0 1 0.54 0.18 -0.16 0.54 0.59 0

PctSameCity85 0 1 0.63 0.20 0.08 0.67 0.74 0

PctSameState85 0 1 0.65 0.20 -0.02 0.7 0.79 0

LemasSwornFT 0 1 0.07 0.14 0.34 0.02 0.02 1675

LemasSwFTPerPop 0 1 0.22 0.16 0.15 0.18 0.2 1675

LemasSwFTFieldOps 0 1 0.92 0.13 -0.33 0.97 0.98 1675

LemasSwFTFieldPerPop 0 1 0.25 0.16 0.16 0.21 0.19 1675

LemasTotalReq 0 1 0.10 0.16 0.35 0.04 0.02 1675

LemasTotReqPerPop 0 1 0.22 0.16 0.27 0.17 0.14 1675

PolicReqPerOffic 0 1 0.34 0.20 0.17 0.29 0.23 1675

PolicPerPop 0 1 0.22 0.16 0.15 0.18 0.2 1675

RacialMatchCommPol 0 1 0.69 0.23 -0.46 0.74 0.78 1675

PctPolicWhite 0 1 0.73 0.22 -0.44 0.78 0.72 1675

PctPolicBlack 0 1 0.22 0.24 0.54 0.12 0 1675

PctPolicHisp 0 1 0.13 0.20 0.12 0.06 0 1675

PctPolicAsian 0 1 0.11 0.23 0.10 0 0 1675

PctPolicMinor 0 1 0.26 0.23 0.49 0.2 0.07 1675

OfficAssgnDrugUnits 0 1 0.08 0.12 0.34 0.04 0.03 1675

NumKindsDrugsSeiz 0 1 0.56 0.20 0.13 0.57 0.57 1675

PolicAveOTWorked 0 1 0.31 0.23 0.03 0.26 0.19 1675

LandArea 0 1 0.07 0.11 0.20 0.04 0.01 0

PopDens 0 1 0.23 0.20 0.28 0.17 0.09 0

PctUsePubTrans 0 1 0.16 0.23 0.15 0.07 0.01 0

PolicCars 0 1 0.16 0.21 0.38 0.08 0.02 1675

PolicOperBudg 0 1 0.08 0.14 0.34 0.03 0.02 1675

LemasPctPolicOnPatr 0 1 0.70 0.21 -0.08 0.75 0.74 1675

LemasGangUnitDeploy 0 1 0.44 0.41 0.12 0.5 0 1675

LemasPctOfficDrugUn 0 1 0.09 0.24 0.35 0 0 0

ML model used : Randam Forest

PolicBudgPerPop 0 1 0.20 0.16 0.10 0.15 0.12 1675

Algorithm

ViolentCrimesPerPop 0 1 0.24 0.23 1.00 0.15 0.03 0

Random Forest is a popular machine learning algorithm that is used for both classification and regression tasks. It is an ensemble method that combines multiple decision trees to make predictions.The idea behind the algorithm is to generate a large number of decision trees, where each tree is trained on a randomly sampled subset of the training data and a random subset of the input features.

During prediction, each decision tree in the forest independently makes a prediction, and the final prediction is then determined by taking the majority vote of all the trees. This helps to reduce overfitting and improve the accuracy of the predictions.

Random Forest is a versatile algorithm that can be used for a wide range of applications, including image classification, fraud detection, and customer churn prediction. It is also relatively easy to use, as it requires minimal data preprocessing and hyperparameter tuning. Overall, Random Forest is a powerful and popular algorithm in the field of machine learning.

Enter the 100 features according to given info provided in the dataset section in the form (feature values seprated by ,)

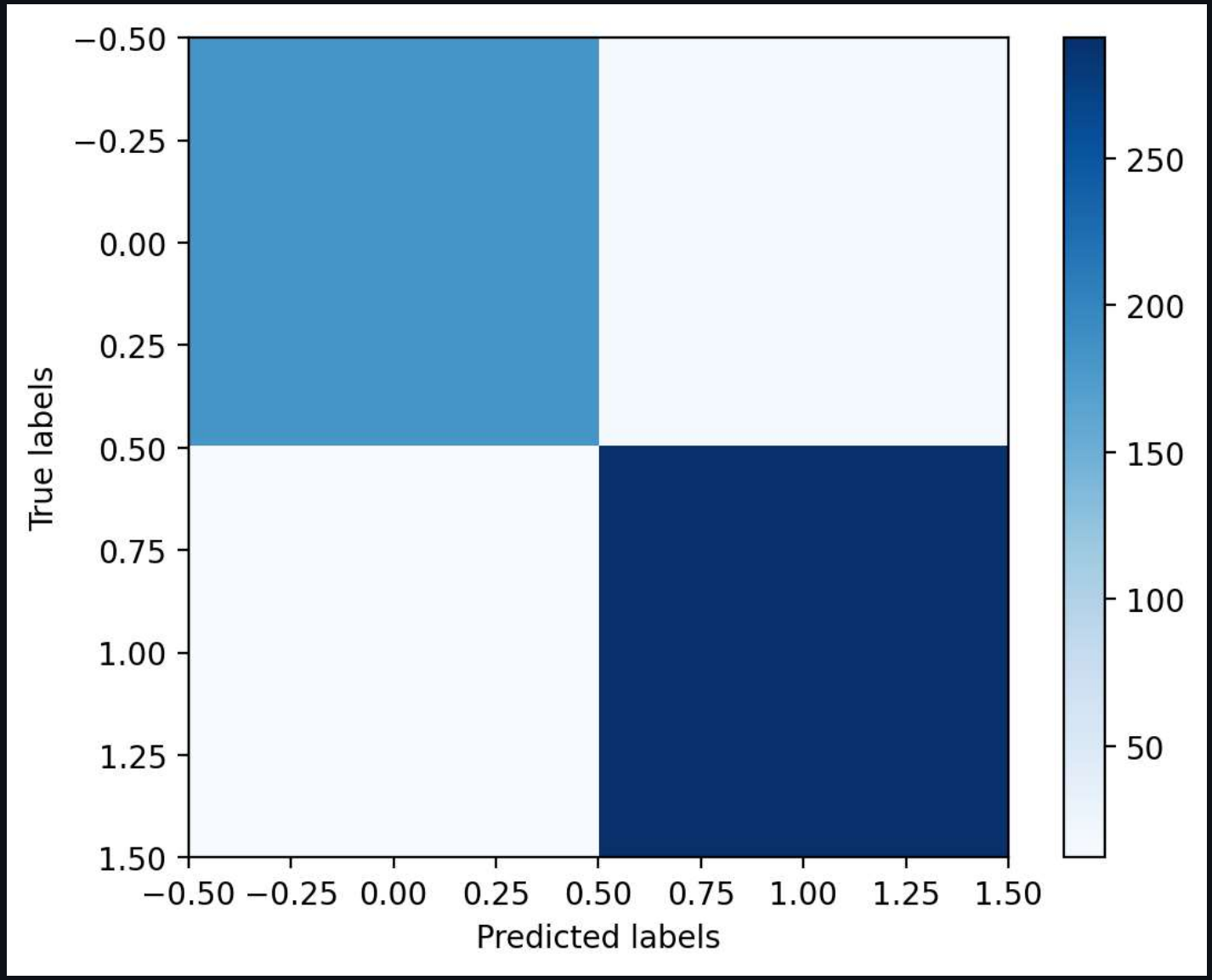
Enter Input Array

0.05,0.67,0.04,0.64,0.5,0.46,0.52,0.61,0.45,0.2,0.07,1,0.65,0.82,0.22,0.56,0.2,0.2,0.35,0.62,0.47,0.53,0.37,0

Submit

High Chance of Crime

Confusin Matrix and accuracy score



Accuracy of model is : 0.944 or 94.389%

This confusion matrix represents the performance of a binary classification model on a test dataset. The rows of the matrix correspond to the actual class labels, while the columns correspond to the predicted class labels.

In this particular case, the model predicted 196 instances to be positive (belonging to the second class) and 303 instances to be negative (belonging to the first class). Out of the 196 positive instances, the model correctly predicted 180, but misclassified 16 as negative. Out of the 303 negative instances, the model correctly predicted 291, but misclassified 12 as positive.

Therefore, the confusion matrix shows that the model has a high true positive rate (TPR) of 0.938 and a high true negative rate (TNR) of 0.960, but a relatively low precision (0.948) and F1-score (0.943) due to the false positive and false negative classifications.