Estimating Future Crime Rates in Los Angeles Based on Crime Data from 2023 to Present

jiaxi li

Introduction

This project aims to forecast future crime rates in Los Angeles by analyzing crime data from 2023 to the present. The goal is to explore how factors such as socio-economic changes, mobility patterns, and land-use influence crime dynamics and to use machine learning models to predict future trends. The study will help identify areas of concern, allowing law enforcement to allocate resources more effectively.

Motivations:

- **Post-Pandemic Crime Evolution:** The COVID-19 pandemic changed crime trends in cities. Lockdowns, reduced mobility, and economic stress led to fewer property crimes but an increase in domestic violence and cybercrime. This shift provides a chance to study how crises affect criminal behavior and how law enforcement can prepare for future changes.

  Localized Crime Hotspots: In a large city like Los Angeles, crime isn't spread evenly. Factors like population density, land use, and socio-economic conditions create crime hotspots. Knowing these patterns is key for law enforcement to use resources wisely and stop crimes before they grow. This study aims to offer targeted insights to improve urban safety.

  Mobility and Criminal Opportunities: Los Angeles, with its high mobility and frequent visitors, sees crime rates shift based on movement patterns. Areas with heavy foot traffic often have more crime opportunities, partly due to visitor anonymity. This project explores the link between urban mobility and crime to better understand how movement shapes criminal chances.

Goal of the Project:

The main goal of this project is to predict crime rates in Los Angeles based on data from 2023-2024 and to forecast crime trends for the year 2025. By analyzing historical data, this study aims to provide insights that help law enforcement agencies prepare for and address emerging crime patterns. Several key questions will guide this research:

Impact of Post-Pandemic Socio-Economic Changes: How have the socio-economic factors in 2023-2024, such as unemployment rates, population density, and mobility patterns, influenced different types of crime in Los Angeles? Are there observable shifts in criminal behavior that need targeted intervention?

Spatial and Temporal Crime Dynamics: Which neighborhoods and times of day are most prone to criminal activities? How do urban characteristics, such as mixed land use and foot traffic, affect crime rates in these areas? Understanding these dynamics will allow for better allocation of law enforcement resources.

Predictive Modeling: Can machine learning models accurately forecast crime trends for 2025 based on historical and real-time data? Which methods and models provide the best accuracy for predicting crime rates and identifying emerging hotspots?

Illustration / Figure

Omissions and Context

This project aims to fill these gaps by integrating diverse and dynamic data sources, like real-time mobility patterns and socio-economic indicators, to improve the accuracy of crime predictions. By advancing traditional methods and adopting machine learning techniques, the project recognizes the need for continuous model updates and the inclusion of new data to better predict crime trends in an ever-evolving urban landscape.

Related Work

1. Socio-economic, built environment, and mobility conditions associated with crime: A study of multiple cities

   13 Apr 2020 De Nadai Marco, Xu Yanyan, Letouzé Emmanuel, González Marta C., Lepri Bruno

   https://cs.paperswithcode.com/paper/socio-economic-built-environment-and-mobility

2. Crime Prediction Based On Crime Types And Using Spatial And Temporal Criminal Hotspots

   9 Aug 2015 **Tahani Almanie**, **Rsha Mirza**, **Elizabeth Lor**

   https://paperswithcode.com/paper/crime-prediction-based-on-crime-types-and

3. Changes in Crime Rates During the COVID-19 Pandemic

   19 May 2021 - Mikaela Meyer, Ahmed Hassafy, Gina Lewis, Prasun Shrestha, Amelia M. Haviland, Daniel S. Nagin ·

   https://stat.paperswithcode.com/paper/changes-in-crime-rates-during-the-covid-19

Data Processing

```
packages <- c(
"tibble",
"dplyr",
"readr",
"tidyr",
"purrr",
"broom",
"magrittr",
"corrplot",
"caret",
```

```r
"rpart",
"rpart.plot",
"e1071",
"torch",
"luz"
)
# renv::install(packages)
sapply(packages, require, character.only=T)
```

```
Loading required package: tibble

Loading required package: dplyr


Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

Loading required package: readr

Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
logical.return = TRUE, : there is no package called 'readr'

Loading required package: tidyr

Loading required package: purrr

Loading required package: broom

Loading required package: magrittr


Attaching package: 'magrittr'
```

```
The following object is masked from 'package:purrr':

    set_names

The following object is masked from 'package:tidyr':

    extract

Loading required package: corrplot

Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
logical.return = TRUE, : there is no package called 'corrplot'

Loading required package: caret

Loading required package: ggplot2

Loading required package: lattice


Attaching package: 'caret'

The following object is masked from 'package:purrr':

    lift

Loading required package: rpart

Loading required package: rpart.plot

Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
logical.return = TRUE, : there is no package called 'rpart.plot'

Loading required package: e1071

Loading required package: torch

Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
logical.return = TRUE, : there is no package called 'torch'
```

```
Loading required package: luz

Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
logical.return = TRUE, : there is no package called 'luz'

    tibble      dplyr      readr      tidyr      purrr      broom   magrittr
      TRUE       TRUE      FALSE       TRUE       TRUE       TRUE       TRUE
  corrplot      caret      rpart rpart.plot      e1071      torch        luz
     FALSE       TRUE       TRUE      FALSE       TRUE      FALSE      FALSE
```

```r
library(caret)
```

```r
library(lubridate)
```

```
Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```r
library(dplyr)

# Read the CSV files
crime_data <- read.csv("Crime_Data_from_2023_to_Present.csv", header = TRUE)
```

```r
head(crime_data)
```

```
      DR_NO     Date.Rptd      DATE.OCC TIME.OCC AREA   AREA.NAME Rpt.Dist.No
1 231000510 1/5/2023 0:00 1/5/2023 0:00     2050   10 West Valley        1067
2 231404137 1/5/2023 0:00 1/4/2023 0:00     1400   14     Pacific        1441
3 232104453 1/5/2023 0:00 1/3/2023 0:00      249   21     Topanga        2126
4 231604110 1/5/2023 0:00 1/4/2023 0:00     1200   16    Foothill        1672
5 230704222 1/5/2023 0:00 1/5/2023 0:00     2200    7    Wilshire         736
6 230900519 1/5/2023 0:00 1/4/2023 0:00     1005    9    Van Nuys         994
  Part.1.2 Crm.Cd                  Crm.Cd.Desc            Mocodes Vict.Age
1        1    330         BURGLARY FROM VEHICLE 1822 0344 1300 1402       24
2        1    510              VEHICLE - STOLEN                            0
3        2    354             THEFT OF IDENTITY                930       37
```

```
4        1    510               VEHICLE - STOLEN                       0
5        2    901 VIOLATION OF RESTRAINING ORDER      2038 2004 1218       51
6        2    623         BATTERY POLICE (SIMPLE)           1212 0417        0
  Vict.Sex Vict.Descent Premis.Cd            Premis.Desc Weapon.Used.Cd
1        M            B       101                 STREET            500
2                             101                 STREET             NA
3        F            H       501 SINGLE FAMILY DWELLING             NA
4                             101                 STREET             NA
5        F            W       710          OTHER PREMISE             NA
6        X            X       101                 STREET            400
                                     Weapon.Desc Status  Status.Desc Crm.Cd.1
1                 UNKNOWN WEAPON/OTHER WEAPON     AA Adult Arrest      330
2                                                 IC  Invest Cont      510
3                                                 IC  Invest Cont      354
4                                                 IC  Invest Cont      510
5                                                 IC  Invest Cont      901
6 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)     AA Adult Arrest      623
  Crm.Cd.2 Crm.Cd.3 Crm.Cd.4                               LOCATION
1      998       NA       NA 17400    VENTURA                      BL
2       NA       NA       NA          WESTMINSTER                  AV
3       NA       NA       NA 20900    SATICOY                      ST
4       NA       NA       NA 11900    ART                          ST
5       NA       NA       NA  5700 W  3RD                          ST
6       NA       NA       NA  3600    BEVERLY GLEN                 BL
                      Cross.Street     LAT       LON
1                                  34.1660 -118.5095
2 E  MAIN                       ST 33.9843 -118.4643
3                                  34.2136 -118.5912
4                                  34.2337 -118.3915
5                                  34.0689 -118.3440
6                                  34.1360 -118.4527
```

```r
colnames(crime_data)
```

```
 [1] "DR_NO"          "Date.Rptd"      "DATE.OCC"       "TIME.OCC"
 [5] "AREA"           "AREA.NAME"      "Rpt.Dist.No"    "Part.1.2"
 [9] "Crm.Cd"         "Crm.Cd.Desc"    "Mocodes"        "Vict.Age"
[13] "Vict.Sex"       "Vict.Descent"   "Premis.Cd"      "Premis.Desc"
[17] "Weapon.Used.Cd" "Weapon.Desc"    "Status"         "Status.Desc"
[21] "Crm.Cd.1"       "Crm.Cd.2"       "Crm.Cd.3"       "Crm.Cd.4"
[25] "LOCATION"       "Cross.Street"   "LAT"            "LON"
```

```r
sapply(crime_data, function(x) sum(is.na(x)))
```

```
         DR_NO      Date.Rptd       DATE.OCC       TIME.OCC           AREA
             0              0              0              0              0
     AREA.NAME    Rpt.Dist.No       Part.1.2         Crm.Cd    Crm.Cd.Desc
             0              0              0              0              0
       Mocodes      Vict.Age       Vict.Sex   Vict.Descent      Premis.Cd
             0              0              0              0              6
   Premis.Desc Weapon.Used.Cd    Weapon.Desc         Status    Status.Desc
             0         233338              0              0              0
      Crm.Cd.1       Crm.Cd.2       Crm.Cd.3       Crm.Cd.4       LOCATION
             4         314787         334512         335164              0
  Cross.Street            LAT            LON
             0              0              0
```

```r
crime_data <- crime_data %>%
  select(-Crm.Cd.2, -Crm.Cd.3, -Crm.Cd.4,-Weapon.Used.Cd)
```

```r
crime_data$Vict.Age[is.na(crime_data$Vict.Age)] <- median(crime_data$Vict.Age, na.rm = TRUE)
```

```r
crime_data$DATE.OCC <- as.Date(crime_data$DATE.OCC, format = "%m/%d/%Y")
crime_data$Date.Rptd <- as.Date(crime_data$Date.Rptd, format = "%m/%d/%Y")
```

```r
crime_data$AREA.NAME <- as.factor(crime_data$AREA.NAME)
crime_data$Crm.Cd.Desc <- as.factor(crime_data$Crm.Cd.Desc)
crime_data$Vict.Sex <- as.factor(crime_data$Vict.Sex)

# Step 1.4: Extract day of the week, month, and time of day from date and time columns
crime_data$Day_of_Week <- weekdays(crime_data$DATE.OCC)
crime_data$Month <- month(crime_data$DATE.OCC, label = TRUE)

# Step 1.5: Create additional relevant features based on data insights (e.g., categorize cri
crime_data$Time_of_Day <- case_when(
  crime_data$TIME.OCC >= 0 & crime_data$TIME.OCC < 600 ~ "Night",
  crime_data$TIME.OCC >= 600 & crime_data$TIME.OCC < 1200 ~ "Morning",
  crime_data$TIME.OCC >= 1200 & crime_data$TIME.OCC < 1800 ~ "Afternoon",
  TRUE ~ "Evening"
)
```

```r
single_class_rows <- crime_data %>%
  group_by(Crm.Cd.Desc) %>%
  filter(n() == 1)

# Remove these from the main dataset and create a train-test split without them
main_data <- anti_join(crime_data, single_class_rows)
```

```
Joining with `by = join_by(DR_NO, Date.Rptd, DATE.OCC, TIME.OCC, AREA,
AREA.NAME, Rpt.Dist.No, Part.1.2, Crm.Cd, Crm.Cd.Desc, Mocodes, Vict.Age,
Vict.Sex, Vict.Descent, Premis.Cd, Premis.Desc, Weapon.Desc, Status,
Status.Desc, Crm.Cd.1, LOCATION, Cross.Street, LAT, LON, Day_of_Week, Month,
Time_of_Day)`
```

Decision Tree Model building

```r
set.seed(123)
trainIndex <- createDataPartition(main_data$Crm.Cd.Desc, p = 0.7, list = FALSE)
```

```
Warning in createDataPartition(main_data$Crm.Cd.Desc, p = 0.7, list = FALSE):
Some classes have no records ( BRIBERY, FIREARMS EMERGENCY PROTECTIVE ORDER
(FIREARMS EPO), MANSLAUGHTER, NEGLIGENT, PETTY THEFT - AUTO REPAIR, THEFT, COIN
MACHINE - ATTEMPT, TRAIN WRECKING ) and these will be ignored
```

```r
train_data <- main_data[trainIndex, ]
test_data <- main_data[-trainIndex, ]

train_data <- bind_rows(train_data, single_class_rows)
```

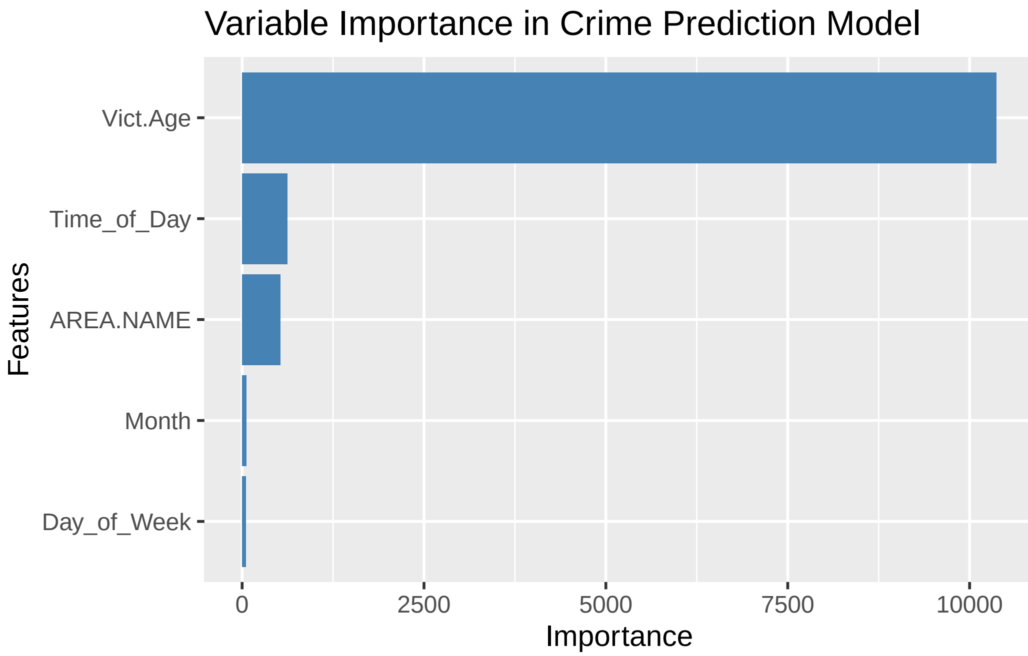
```r
# Train the model using relevant features
tree_model <- rpart(Crm.Cd.Desc ~ AREA.NAME + Vict.Age + Day_of_Week + Month + Time_of_Day, d

# View the model's summary
summary(tree_model)
```

```
Call:
rpart(formula = Crm.Cd.Desc ~ AREA.NAME + Vict.Age + Day_of_Week +
```

```
    Month + Time_of_Day, data = train_data, method = "class")
  n= 234685

          CP nsplit rel error    xerror         xstd
1 0.07998278      0 1.0000000 1.0000000 0.0007598157
2 0.01000000      1 0.9200172 0.9200172 0.0009190726

Variable importance
Vict.Age
     100

Node number 1: 234685 observations,    complexity param=0.07998278
  predicted class=VEHICLE - STOLEN          expected loss=0.8806784  P(node) =1
    class counts:   507   155 11298   973 16838    61   515   928     3     3  1313     2
   probabilities: 0.002 0.001 0.048 0.004 0.072 0.000 0.002 0.004 0.000 0.000 0.006 0.000 0.0
  left son=2 (163724 obs) right son=3 (70961 obs)
  Primary splits:
      Vict.Age    < 1 to the right, improve=10366.38000, (0 missing)
      Time_of_Day splits as  RLRL, improve=  623.09260, (0 missing)
      AREA.NAME   splits as  RLLRRRLRLRLLLLRRLLLLL, improve=  527.51970, (0 missing)
      Month       splits as  LLLRRRRRRRRR, improve=   57.36366, (0 missing)
      Day_of_Week splits as  RRLLRRR, improve=   48.30511, (0 missing)

Node number 2: 163724 observations
  predicted class=BATTERY - SIMPLE ASSAULT  expected loss=0.8984572  P(node) =0.697633
    class counts:   294    36 10876   834 16625    43   136   921     2     2  1282     2
   probabilities: 0.002 0.000 0.066 0.005 0.102 0.000 0.001 0.006 0.000 0.000 0.008 0.000 0.0

Node number 3: 70961 observations
  predicted class=VEHICLE - STOLEN          expected loss=0.6066995  P(node) =0.302367
    class counts:   213   119   422   139   213    18   379     7     1     1    31     0
   probabilities: 0.003 0.002 0.006 0.002 0.003 0.000 0.005 0.000 0.000 0.000 0.000 0.000 0.0
```

```r
predictions <- predict(tree_model, test_data, type = "class")
```

```r
importance <- varImp(tree_model, scale = FALSE)

# Plot the variable importance using ggplot2
ggplot2::ggplot(importance, aes(x = reorder(rownames(importance), Overall), y = Overall)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(title = "Variable Importance in Crime Prediction Model", x = "Features", y = "Importa
```



Describe a simple, baseline model that you will compare your neural network against. This can be a simple model that you build.

Quantitative Results

random forest

Linear regression

A description of the quantitative measures of your result. What measurements can you use to illustrate how your model performs?

Qualitative Results

Include some sample outputs of your model, to help your readers better understand what your model can do. The qualitative results should also put your quantitative results into context (e.g. Why did your model perform well? Is there a type of input that the model does not do well on?)

Discussion

Discuss your results. Do you think your model is performing well? Why or why not? What is unusual, surprising, or interesting about your results? What did you learn?

Ethical Considerations

Description of a use of the system that could give rise to ethical issues. Are there limitations of your model? Your training data?

(Note that the expectations are higher here than in the project proposal.)

Conclusion(Optional)

Summarize the whole report.