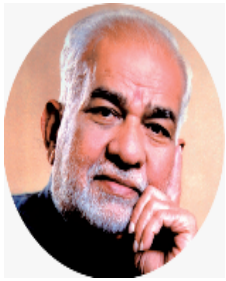
BVRIT HYDERABAD College of Engineering for Women

Department of Information Technology



CRIME ANALYSIS AND PREDICTION AGAINST WOMEN

Under the guidance of:

Guide Name: Ms. R. Sravani

Designation: Assistant Professor

Team- 17

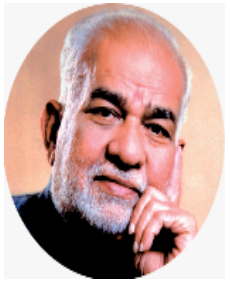
A. Spoorthi (19WH1A1250)

Riya Fathima (19WH1A1251)

Ch. Supraja (19WH1A1254)

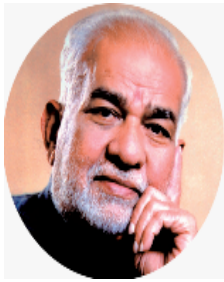
K. Sowmya (19WH1A1256)

Contents



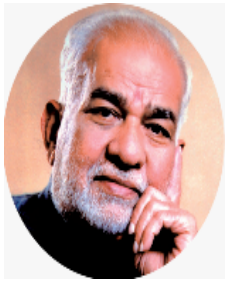
- Abstract
- Introduction
- Problem Definition
- Literature Survey
- Proposed System
- Project Modules
- Implementation and Results
- Performance Measures
- Conclusion
- References

Abstract



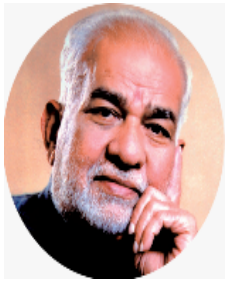
- Crime against women is one of the dangerous aspects of our society which is growing continuously in intensity and complexity.
- The primary objective of this project is to distinguish various crimes using clustering techniques based on the occurrences and regularity. In this project, the crime data is classified using the K-means clustering algorithm, Linear Regression, ARIMA model.
- This proposed system can indicate the crime ahead which has a high probability of crime and thus effectively help in significantly reducing the crime rate in various parts of the country.

Introduction



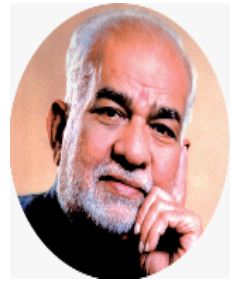
- Crimes in India are increasing at a very tremendous rate.
- India is one of the countries which has tried to balance between the advancement and their culture.
- There has been a huge increase in the number of crimes against women.
- This project aimed at extracting knowledge from crime records to better understand criminal behavior and ultimately prevent future crimes.

Problem Definition



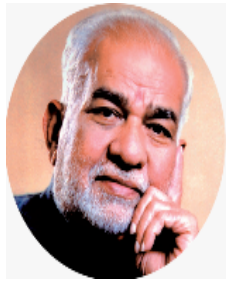
- The aim of this project is to make crime predictions using the features present in the dataset. The dataset is extracted from the official sites. With the help of machine learning algorithms, using python as core we can predict the type of crime that will occur in a particular area.

Literature Survey



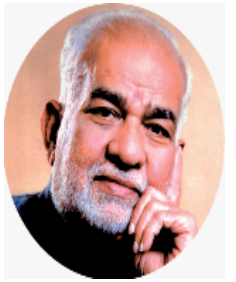
Author	Journal (Title of Paper)	Description	Techniques
Aman Kumar, Nikhil Tiwari , Prakhar Gupta, Dr.S.N.Rajan	Women Crime Prediction(2020)	To predict major types of crimes occurring on women using various data mining techniques	Data Mining techniques, Naïve Bayes
K. Zakhir Hussain	Analyzing violent criminal behavior on women by simulation model (2019)	Data mining approaches is a very useful tool that can help and support in identifying violent criminal behavior on women.	data mining techniques

Literature Survey



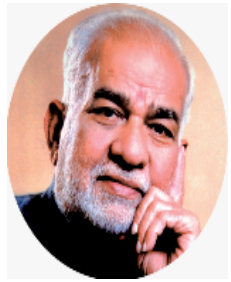
Author	Journal (Title of Paper)	Description	Techniques
P.Tamilarasi, R.Rani	Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning.(2020)	K-fold cross-validation algorithm is tested for accuracy.	KNN and decision trees, Naïve Bayes
Varshitha D N ,Vidyashree K P, Aishwarya P Janya T S, K R Dhananjay Gupta Sahana R	Paper on Different Approaches for Crime Prediction system(2018)	Here grouping is done according to various types of crimes against women taking place in different states and cities of India.	data mining techniques

Proposed System



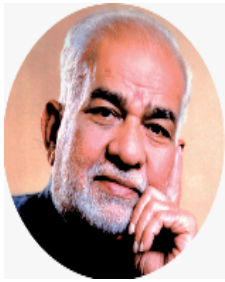
- The aim of this project is to make crime prediction using the features present in the dataset.
- Building the model will be done using a better algorithm depending upon the accuracy.
- The K-means clustering and a few classification algorithms will be used for crime prediction.

Project Modules



- Data Pre-processing
- Data Analysis
- Training Model
- Comparison Results

Implementation and Results



Data Pre-processing

```
df.isna().sum()
```

```
Unnamed: 0                                          0
STATE/UT                                            0
DISTRICT                                            0
Year                                                0
Rape                                                0
Kidnapping and Abduction                            0
Dowry Deaths                                        0
Assault on women with intent to outrage her modesty 0
Insult to modesty of Women                          0
Cruelty by Husband or his Relatives                 0
Importation of Girls                                0
dtype: int64
```

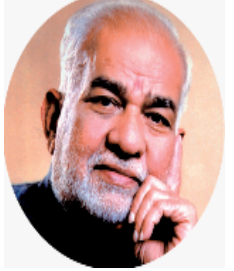
```
df.drop('Unnamed: 0',axis=1,inplace=True)
```

```
df.head()
```

	Unnamed: 0	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
0	0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0
1	1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0
2	2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0
3	3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0
4	4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0

	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0
1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0
2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0
3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0
4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0

VISHNU
UNIVERSAL LEARNING



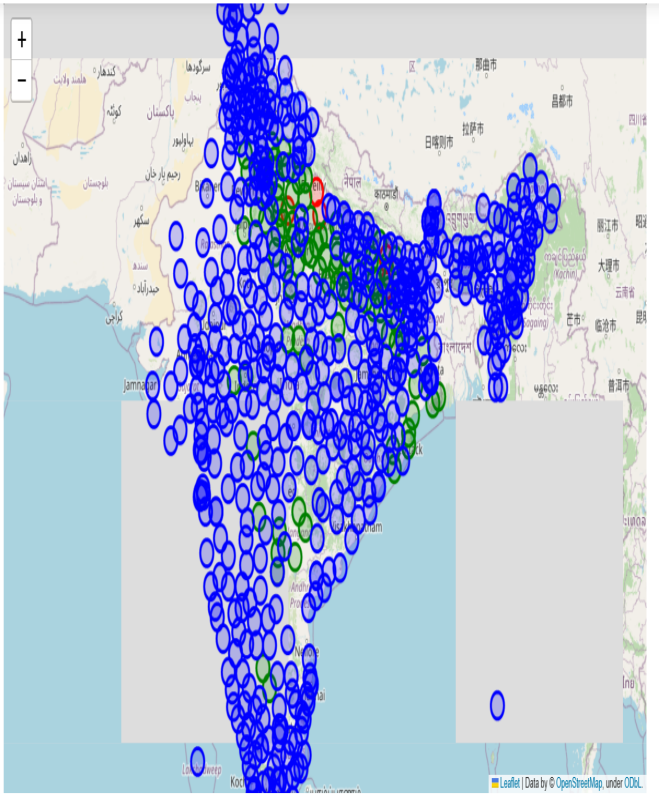
Implementation and Results

K-Means clustering

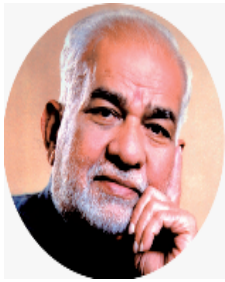
```
In [174]: import pandas as pd
          import numpy as np
          df = pd.read_csv('C:/Users/supraja reddy/Documents/Projects/kmeans.csv')
          # df

          #three clusters _(centroidsspecific)
          c1 = df['Dowry_Deaths'].min()
          c2 = df['Dowry_Deaths'].max()
          c3 = df['Dowry_Deaths'].max()/2
```

```
In [176]: lclust_values = dict( zip(keyss,valuess))
          mclust_values = dict( zip(keyss,valuess))
          hclust_values = dict( zip(keyss,valuess))
```

```
In [177]: i=0
          lclust = 0
          hclust = 0
          mclust = 0
          # print(length)
          for i in range(length):
              a = abs(df.Dowry_Deaths[i]-c1)
              b = abs(df.Dowry_Deaths[i]-c2)
              c = abs(df.Dowry_Deaths[i]-c3)
              #the one with the minimum distance will be considered
              val = min(a,b,c)
              if val == a:
                  lclust = df.Dowry_Deaths[i]
                  lclust_values[i] = lclust
              elif val == b:
                  hclust = df.Dowry_Deaths[i]
                  hclust_values[i] = hclust
              else:
                  mclust = df.Dowry_Deaths[i]
                  mclust_values[i] = mclust
```

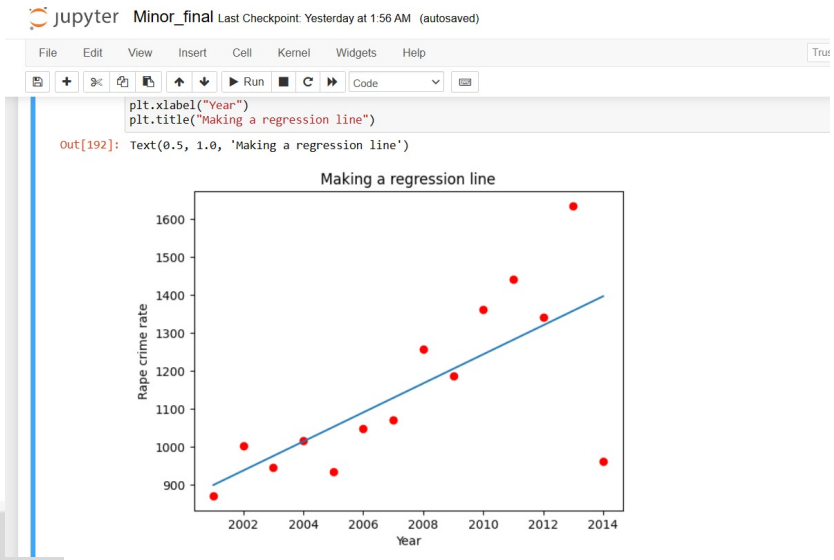


Implementation and Results



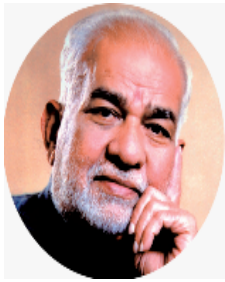
Linear Regression before prediction

```
In [192]: m,b=slope_intercept(xs,ys)
          reg_line=[(m*x)+b for x in xs]
          plt.scatter(xs,ys,color="red")
          plt.plot(xs,reg_line)
          plt.ylabel("Rape crime rate")
          plt.xlabel("Year")
          plt.title("Making a regression line")

Out[192]: Text(0.5, 1.0, 'Making a regression line')
```




```
X1 = df.iloc[:, :-7].values
y1 = df.iloc[:, 1].values
from sklearn.model_selection import train_test_split
x1_train, x1_test, y1_train, y11_test = train_test_split(X1, y1, test_size=0.3, random_state=0)
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(x1_train,y1_train)
```

Implementation and Results



Linear Regression after prediction

```
]: trmergedlist=np.concatenate((xs, years), axis=0)
print("Years: "+str(trmergedlist))
#print("Rape rate :"+str(temergedlist))
#print(ypredd)

#print(temergedlist)
#m,b=slope_intercept(trmergedlist,temergedlist)
reg_line=[(m*x)+b for x in trmergedlist]
temergedlist=np.concatenate((ys,ypredd), axis=0)
print("Rape: "+str(temergedlist))
plt.scatter(trmergedlist,temergedlist,color="red")
plt.plot(trmergedlist,reg_line)
plt.xlim((2001,2018))
#plt.ylim((800,1900))
plt.ylabel("Rape crime rate")
plt.xlabel("Year")
plt.title("Making a regression line")
```

```
Years: [2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
 2015 2016 2017 2018]
Rape: [ 871.   1002.    946.   1016.    935.   1049.   1070.   1257.   1188.
 1362.   1442.   1341.   1635.    961.   1435.16 1473.42 1511.68 1549.94]
```

```
]: Text(0.5, 1.0, 'Making a regression line')
```

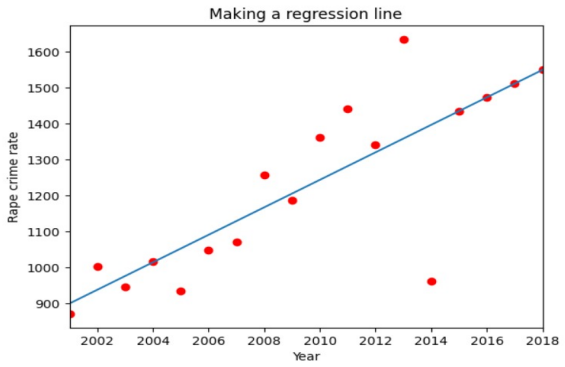
jupyter Minor_final Last Checkpoint: Yesterday at 1:56 AM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

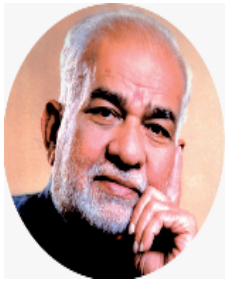
Run Code

```
Years: [2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
 2015 2016 2017 2018]
Rape: [ 871.   1002.    946.   1016.    935.   1049.   1070.   1257.   1188.
 1362.   1442.   1341.   1635.    961.   1435.16 1473.42 1511.68 1549.94]
```

Out[197]: Text(0.5, 1.0, 'Making a regression line')



Implementation and Results



Arima model

In [255]:
```
#using the ARIMA model to forecast the future 5 years values
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.stattools import adfuller
```

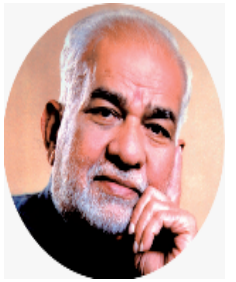
In [256]:
```
def check_stationarity(series):
    # Copied from https://machinelearningmastery.com/time-series-data-sta

    result = adfuller(series.values)

    print('ADF Statistic: %f' % result[0])
    print('p-value: %f' % result[1])
    print('Critical Values:')
    for key, value in result[4].items():
        print('\t%s: %.3f' % (key, value))

    if (result[1] <= 0.05) & (result[4]['5%'] > result[0]):
        print("\u001b[32mStationary\u001b[0m")
    else:
        print("\x1b[31mNon-stationary\x1b[0m")
check_stationarity(df_AP['Dowry_Deaths'])
```

```
ADF Statistic: -4.253928
p-value: 0.000533
Critical Values:
        1%: -4.012
        5%: -3.104
        10%: -2.691
```

In [260]:
```
fc = model_fit.forecast(4, alpha=0.05)
for i in fc:
    print(round(i,0))
```

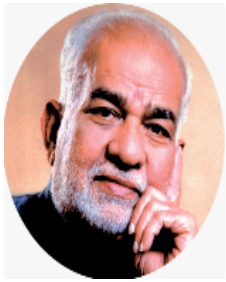
```
149.0
97.0
57.0
26.0
```

Performance Measure



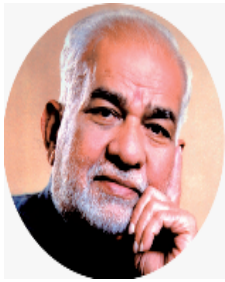
- We used linear regression and Arima model to predict future patterns of crime
- By using linear regression we got accuracy of 70%

Conclusion



- Crime prediction is one the current trends in the society. Crime prediction intends to reduce crime occurrences. It does this by predicting which type of crime may occur in future. Here, analysis of crime and prediction are performed with the help of various approaches. In this system, we get to classify and cluster to improve the accuracy of location and pattern- based crimes. This software predicts frequently occurring crimes, especially for particular state, and occurrences.

References



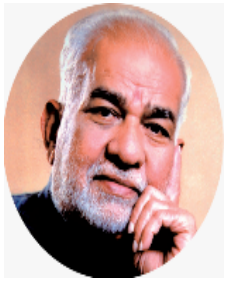
- Varshitha D N, Vidyashree K P, Aishwarya P Janya T S, K R Dhananjay Gupta Sahana R," Paper on Different Approaches for Crime Prediction system", International Journal of Engineering Research Technology (IJERT), ISSN: 2278-0181, 2018.

- K.Zakhir Hussain, "Analyzing violent criminal behaviour on women by simulation model," 2019.

- Aman Kumar, Nikhil Tiwari, Prakhar Gupta, Dr.S.N.Rajan," Women Crime Prediction",2020.

- P.Tamilarasi, R.Rani "Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning",2020.

VISHNU
UNIVERSAL LEARNING



THANK YOU