CS584 – machine learning

fall 2016

**SAFETY PREDICTION IN VARIOUS DISTRICTS OF CHICAGO AT A GIVEN TIME**

Group Members:

**Sami Ahmad Khan, A20352677**

**Zeeshan A. Khavas, A20341778**

SamSi

Table of Contents

[Task 2](#_Toc467172373)

[Dataset 2](#_Toc467172374)

[Data source 2](#_Toc467172375)

[Target variable 2](#_Toc467172376)

[Features 2](#_Toc467172377)

[Data size 2](#_Toc467172378)

[Preprocessing 2](#_Toc467172379)

[Visualization 2](#_Toc467172380)

[Target 2](#_Toc467172381)

[Features 2](#_Toc467172382)

[Evaluation 2](#_Toc467172383)

[Performance Measure 2](#_Toc467172384)

[Classifiers 3](#_Toc467172385)

[Evaluation Strategy 3](#_Toc467172386)

[Performance Results 3](#_Toc467172387)

[Top Features 3](#_Toc467172388)

[Discussion 3](#_Toc467172389)

[Interesting/Unexpected Results 3](#_Toc467172390)

[Contributions of Each Group Member 3](#_Toc467172391)

[Conclusion 3](#_Toc467172392)

[References 3](#_Toc467172393)

**SAFETY PREDICTION IN VARIOUS DISTRICTS OF CHICAGO AT A GIVEN TIME**

Group Members: Sami Ahmad Khan, Zeeshan Aamir Khavas

# Task

Given a dataset of reported incidents of crimes that occurred in the city of Chicago in the year 2016, we classify the districts as “Safe” and “Unsafe”.

Hence we can predict the safety of a particular district at a given time based on prediction model.

<Describe the task here. What is the problem? What is the classification/regression problem? Why is it interesting?>

# Dataset

## The raw dataset was collected from “City of Chicago” data portal.

## The original dataset contained millions of records for the year 2016. This was reduced to a size of 11,670 records by randomly choosing ~1000 entries from each month.

## From the dataset we selected few features for the prediction model. Then manually added weightage to each crime on the basis of crime severity.

## We created Target Labels on the basis of severity of crimes i.e. ‘Primary type’ and ‘Arrest’ and manually labelled them as ‘0’ for Safe and ‘1’ for Unsafe.

## So, our dataset has 11,670 rows, 5 columns and 1 Label column.

## Data source

<Where did you get the dataset from? Did you collect any additional data? Did you manually label any data?>

## Target variable

<What is the target variable?>

## Features

<What are the input features and how many are there?>

## Data size

<How many instances do you have in the data?>

# Preprocessing

<Describe any preprocessing you have done for this dataset.>

# Visualization

## Target

<Provide statistics about the target variable; if classification, provide counts for each class; if regression, print mean and variance.>

<Provide a histogram of the target variable.>

## Features

<Visualize the features. For categorical variables, provide bar plots, for numerical features, provide histograms and statistics. If there are too many features, visualize and provide statistics for only 10 of them that you choose. If your data is text, print the list of most frequent words. If your data is raw (like images), describe how your model is handling it.>

# Evaluation

## Performance Measure

<Which performance measure you chose and why?>

## Classifiers

<Which classifiers and parameter settings did you try and why?>

## Evaluation Strategy

<Did you do train-test split or cross-validation and why?>

## Performance Results

<Report your results, including baselines, using a table similar to the one on slide 7 of the project presentation template file.>

## Top Features

<Present the top features with respect to your model.>

## Discussion

<Briefly discuss your results. Did the best classifier perform as well as you expected? If things did not work out as well, why do you think they did not work? Did one classifier perform much better (or worse) than others? And so on.>

# Interesting/Unexpected Results

<Discuss a few interesting/unexpected cases. See slide 9 of the project presentation template file.>

# Contributions of Each Group Member

<If you are working in a group, please discuss in detail what each member did for this project.>

# Conclusion

<Provide concluding remarks.>

# References

<Provide references if you have any.>

Notes (erase these notes before you save and submit):

1. Once you are done editing, this file, update Table of Contents; you can simply click on Table of Contents and click on the button Update Table once it appears. Then, save this file as a pdf file.
2. Do not copy text from any website/paper. All the wording has to be your own.
3. Delete instructions (the text that appear inside <> above) and these notes before you save and submit.
4. Anywhere it says “discuss”, one or two paragraphs should be sufficient. Do not write anything unnecessary, but also do not omit any necessary details.