My grades for **Project proposal**

Proposal

6.6 / 10

Here you submit your proposal in pdf format

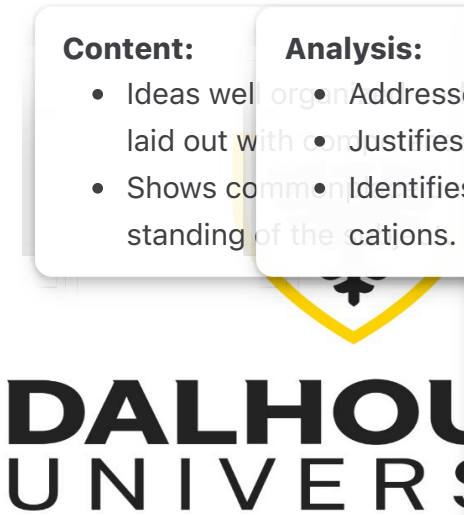
Content:
- Ideas well organized and laid out with coherence.
- Shows comprehension and understanding of the subject.

Analysis:
- Addresses most of the points.
- Justifies most design decisions.
- Identifies criteria used in justifications.

2.3

Delivery:
- Makes good word choices.
- Mostly uses standard conventions.
- The document could use some editing.

2



Crime Zone Classifier(CZC) Project Proposal

Submitted by:

Venkata Sreenivas Prasad Kasibhatla

B00972626 (vn769140@dal.ca)

Jahnavi Gajjala

B00969773 (jh707761@dal.ca)

Sichen Qiu

B00839638 (sc985823@dal.ca)

Submitted in fulfilment of the requirements for CSCI 4146&6409 to:

Prof. Evangelos Milios

Faculty of Computer Science

Dalhousie University

Halifax, Nova Scotia

February 1st. 2024

Introduction:

Urban crime is a complex issue that has an impact on people's safety and quality of life in cities all over the world. Effective management of crime is a difficulty for law enforcement authorities due to its dynamic character, which is influenced by a variety of socio-economic factors. The project aims to leverage geospatial analysis to identify and visualise crime hotspots within a city and predict potential areas where crime rates may increase. By employing advanced data science techniques, we aim to analyse patterns in crime data, offering actionable insights that can aid in preventing future incidents.

Police can focus their efforts on areas that need the greatest attention by identifying crime hotspots. They can also potentially prevent crime by strategically present in the community and engaging with the locals. The advantages also apply to community development and urban planning, enabling well-informed choices that can improve the quality of life in affected neighbourhoods.

Background:

In the past, police agencies used pin maps, which are actual maps on which crime scenes are marked with pins. Although the scale and scope of this manual method were constrained, it was nevertheless possible to perform a basic spatial analysis. Geographic Information Systems (GIS), which enabled the investigation of crime at a much more granular level and over wider areas, revolutionised this procedure with the advent of the digital age.

From this evolution, the idea of "crime hotspots" was born, designating parts of a city where crime is noticeably higher than in other places. Hotspots may remain constant or fluctuate over time in response to modifications in urban dynamics and strategies used by law enforcement.

Motivation:

This initiative's main goal is to improve public safety by using data to inform decision-making. Given the resource limitations that law enforcement agencies frequently have, it is critical to focus resources in the regions where crime is most prevalent. Predictive analytics can be used in addition to traditional reactive approaches to crime prevention, enabling a more planned use of police resources. Furthermore, by comprehending the spatial distribution of crime, policymakers in the fields of social services and urban development can better address the root causes of crime.

Benefits:

The application of geospatial analysis in crime data has the potential to provide numerous benefits:

- **Enhanced Public Safety:** The primary benefit is the potential reduction in crime rates through more effective policing and community engagement strategies.
- **Efficient Use of Resources:** Law enforcement agencies often face budgetary and manpower constraints. Predictive analytics helps optimise the use of resources, directing them to where they are most needed.

- **Community Empowerment:** Providing communities with information about crime in their areas empowers them to take part in safety initiatives and fosters a collaborative relationship with law enforcement.

- **Evidence-Based Policy Making:** When urban policies are informed by reliable data, they are more likely to be effective and have a lasting positive impact.

Description of the Dataset:

A table of features is expected here.

The dataset for this project is a co[illegible] Los Angeles, USA. It contains det[illegible] attributes:

- **DR_NO:** Unique report number for each crime incident.
- **Date Rptd:** The date when the crime was reported.
- **DATE OCC:** The date when the crime occurred.
- **TIME OCC:** The time when the crime occurred.
- **AREA:** Numeric code representing the area where the crime was reported.
- **AREA NAME:** The name of the area or neighbourhood where the crime occurred.
- **Rpt Dist No:** Reporting district number, a specific code that indicates a smaller segment within the larger area.
- **Part 1-2:** Indicator of whether the crime is a Part 1 (more serious) or Part 2 offence.
- **Crm Cd:** Numeric code for the crime category.
- **Crm Cd Desc:** Description of the crime category.
- **Mocodes**: Modus operandi codes providing details on the crime.
- **Vict Age:** Age of the victim.
- **Vict Sex:** Sex of the victim.
- **Vict Descent:** Ethnic/racial descent of the victim.
- **Premis Cd:** Numeric code indicating the type of location where the crime occurred.
- **Premis Desc:** Description of the location type.
- **Weapon Used Cd:** Code for the type of weapon used (if any).
- **Weapon Desc:** Description of the weapon used.
- **Status:** Code indicating the status of the crime report (e.g., investigated, closed).
- **Status Desc:** Description of the report status.
- **Crm Cd 1-4:** Up to four additional crime codes if multiple offences occurred in the same incident.
- **LOCATION:** Street address or intersection where the crime occurred.
- **Cross Street:** Nearest cross street (if applicable).
- **LAT:** Latitude coordinate of the crime location.
- **LON:** Longitude coordinate of the crime location.

The dataset is extensive in both size and dimensionality, offering a rich source of information for analysis.
Number of columns:28
Number of rows: 883987
The types of data range from categorical (e.g., AREA NAME, Vict Sex) to numerical (e.g., Vict Age, TIME OCC), as well as geospatial coordinates (LAT, LON).

What data science problem will your system solve?

Clustering:

- **Nature of Task:** Clustering is about grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). It's a form of unsupervised learning because the grouping is not based on any pre-known labels or outputs.

Application in this Case:

- Using clustering (K-means or DBSCAN) to group geographic locations based on crime data. The goal is to identify areas with similar crime characteristics, which are not predefined but discovered through the algorithm.

How are you going to choose K?

Hotspot Identification:

- The system will identify areas with high crime rates and classify these into hotspots using threshold based classification and density-based clustering techniques. This analysis will extend to categorise hotspots into high, medium, and low risk based on crime density, providing a nuanced view of the urban crime landscape.

Functionality:
Minimum Functionality:

1. **Data Cleaning and Preprocessing:**
   Standardised formats, handle missing values, and remove duplicates.
   Extract useful features from datetime fields, such as day of the week, time of day, etc.

2. **Basic Visualization:**
   Create static maps to visualise crime incidents using latitude and longitude data.
   Generate basic charts and graphs to illustrate crime trends and distributions.

3. **Identification of Crime Hotspots:**
   Use a simple threshold classification to identify areas with higher frequencies of crime.
   Classify these areas into broad categories like high, medium, and low incidence zones.

Expected Functionality:

1. **Model development (clustering techniques):**
   - Implement density-based clustering (e.g., DBSCAN) to identify and visualise crime hotspots.Analyse spatial patterns to understand the distribution of different types of crimes.
2. **Heatmaps and visualisation:**
   - Generate heat maps to visualise areas with high crime densities.

3. **Validation and testing:**
    - Use a portion of your data for training (setting thresholds) and another portion for validation.

Bonus Functionality:

1. **Temporal Analysis of Crime Data:**
    - Conduct time series analysis to identify trends and patterns over time and how the hotspots evolved over time. Explore variations in crime rates by time of day, week, and season.
2. **Correlation with External Datasets:**
    - Integrate external demographic and socioeconomic data to examine correlations of unemployment or budget allocation with crime rates.
3. **Location expansion other than LA**

High-Level Organization and Design:

1. **Data Acquisition and Preprocessing:**
    - **Data Collection:** Obtain crime reports from the metropolitan police department in Los Angeles, USA.
    - **Data Cleaning:** Standardize formats, handle missing values, remove duplicates, and extract useful features from datetime fields.
    - **Data Integration:** Integrate external datasets containing demographic and socioeconomic information.
2. **Visualization Module:**
    - **Basic Visualization:** Create static maps to visualize crime incidents using latitude and longitude data.
    - **Graphical Representation:** Generate basic charts and graphs to illustrate crime trends and distributions.
3. **Spatial Analysis Module:**
    - **Threshold-Based Classification:** We would define a threshold (e.g., number of crimes within a certain radius) to classify an area as a hotspot.
    - **Density-Based Clustering:** We plan to use clustering algorithms like DBSCAN or K-means to identify areas with high densities of crime incidents. These clusters can be labelled as hotspots. Further into high, medium, and low risk/incidence zones based on crime density.
    - **Advanced Spatial Analysis:** Analyze spatial patterns to understand the distribution of different types of crimes within hotspots.
4. **Model Development and Optimization:**
    - **Clustering Algorithms**: Employ clustering algorithms like K-Means or DBSCAN for spatial clustering

- **Hyperparameter Tuning:** Fine-tune machine learning models through hyperparameter optimization for improved prediction accuracy and robustness.

5. **Temporal Analysis Module:**
   - **Time Series Analysis:** Utilize ARIMA or SARIMA to
     time and identify temporal patterns.
   - **Pattern Recognition:** Identify variations in crime rat
     season.

You said you are not predicting anything.

6. **Model Evaluation:**
   - The model will be evaluate
     score, Stability measures,
   - The system will include mechanisms to assess the effectiveness of deployed law enforcement and community-based interventions over time, based on changes in hotspot dynamics.

Regression task evaluation is missing

Milestones and Timeline

Data Collection(Venkata Sreenivas Prasad Kasibhatla, Jahnavi Gajjala, Sichen Qiu):

- Find the dataset which has enough context to do analysis, The dataset should have enough categories and length to do basic model training.

Data Analysis(Venkata Sreenivas Prasad Kasibhatla, Jahnavi Gajjala, Sichen Qiu):

- Set the project problem based on the project categories lists.

Project implementation(Venkata Sreenivas Prasad Kasibhatla, Sichen Qiu):

- Build the model to classify the criminal distribution in LA, also find the frequency of different criminals and find the occurrence frequency of different types of crimes.

Testing and Evaluation(Venkata Sreenivas Prasad Kasibhatla, Jahnavi Gajjala):

- Test the model and check if it matches the assumption or not. For example, the model can classify the relationship between a criminal's characteristics and criminal type.

Model Enhancement(Venkata Sreenivas Prasad Kasibhatla, Jahnavi Gajjala, Sichen Qiu):

- Based on the model output, change the weight of input categories to make the model more precisely display the most relevant condition to each criminal rate. Based on this, make the prediction for future criminal distribution.

Finalization and Demonstration(Venkata Sreenivas Prasad Kasibhatla Jahnavi Gajjala, Sichen Qiu):

- Testing the final model again and displaying the outcome demonstration.

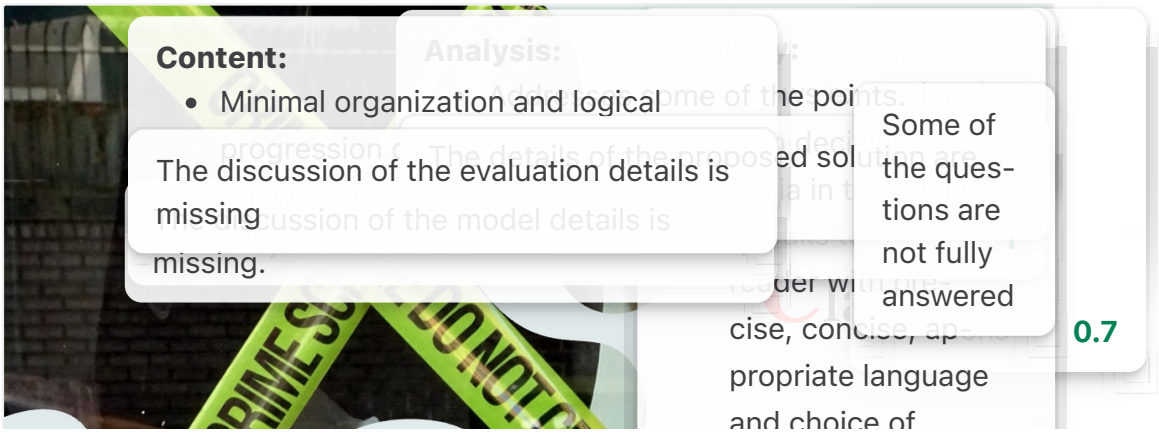
Writing Report(Venkata Sreenivas Prasad Kasibhatla, Jahnavi Gajjala, Sichen Qiu):

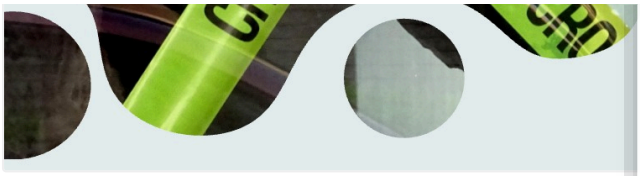
- Write the report to display the project processing.

Slides

4.7 / 10

Here you submit the slides of your project proposal presentation in pdf format





words.

- Always uses standard conventions.
- The document looks professional. 2

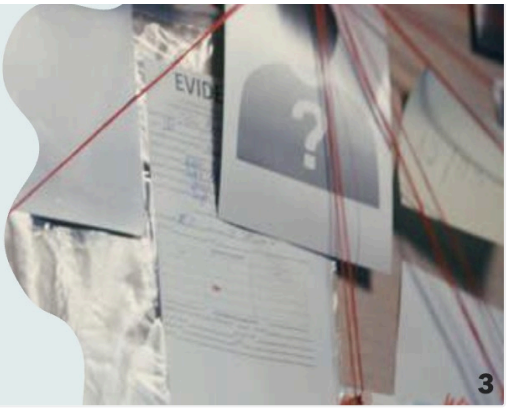
BUSINESS PROBLEM

- In metropolitan cities like Los Angeles, there might be numerous crime patterns, which can be both enduring and constantly evolving, are common in urban areas.
- Design and implement a system that uses past crime data to precisely identify and categorize crime hotspots in a metropolitan region.
- Categorizing different areas based on their likelihood of experiencing criminal activities.

2

WHY?

1. Public Safety Enhancement
2. Effective Resource Allocation
3. Informed Community Awareness
4. Aid in Urban Planning



3

Source Used:

https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/data_preview

What's in this dataset:

Rows: 884K

Columns: 28

Time Range: February 10,2020 to January 18,2024

Each row is a crime incident

4

INPUT DATA

Few of the Features:

DR_NO : Division of Records Number

Date Rptd : Crime Report Date

DATE OCC : Crime Occurance date

TIME OCC : Crime Occurance Time

AREA : The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department.

AREA NAME : The 21 Geographic Areas or Patrol Divisions

Rpt Dist No : A 4-digit code that represents a sub-area.

Crm Cd : Indicates the crime committed.

Crm Cd Desc : Defines the Crime Code provided.

Vict Age : Victim age

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 883987 entries, 0 to 883986
Data columns (total 28 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   DR_NO           883987 non-null  int64
 1   Date Rptd       883987 non-null  object
 2   DATE OCC        883987 non-null  object
 3   TIME OCC        883987 non-null  int64
 4   AREA            883987 non-null  int64
 5   AREA NAME       883987 non-null  object
 6   Rpt Dist No     883987 non-null  int64
 7   Part 1-2        883987 non-null  int64
 8   Crm Cd          883987 non-null  int64
 9   Crm Cd Desc     883987 non-null  object
 10  Mocodes         760974 non-null  object
 11  Vict Age        883987 non-null  int64
 12  Vict Sex        766927 non-null  object
 13  Vict Descent    766919 non-null  object
 14  Premis Cd       883977 non-null  float64
 15  Premis Desc     883447 non-null  object
 16  Weapon Used Cd  306604 non-null  float64
 17  Weapon Desc     306604 non-null  object
 18  Status          883987 non-null  object
 19  Status Desc     883987 non-null  object
 20  Crm Cd 1        883976 non-null  float64
 21  Crm Cd 2        64497 non-null   float64
 22  Crm Cd 3        2176 non-null    float64
 23  Crm Cd 4        62 non-null      float64
 24  LOCATION        883987 non-null  object
 25  Cross Street    139779 non-null  object
 26  LAT             883987 non-null  float64
 27  LON             883987 non-null  float64
dtypes: float64(8), int64(7), object(13)
memory usage: 188.8+ MB
```

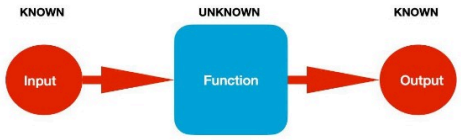
5

RESULT/OUTPUT

The model will classify areas into categories indicating their potential as crime hotspots.

These might consist of **Risk categories** include designations such as "High Risk," "Medium Risk," and "Low Risk.

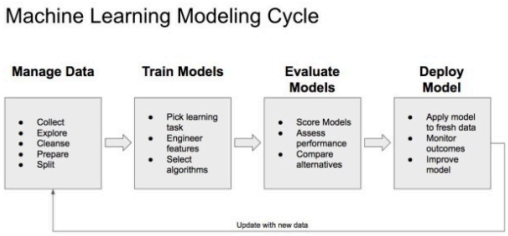
Visualization and Analysis: Maps with coloured zones that illustrate the different levels of crime danger around the city are examples of visual representations that display data.



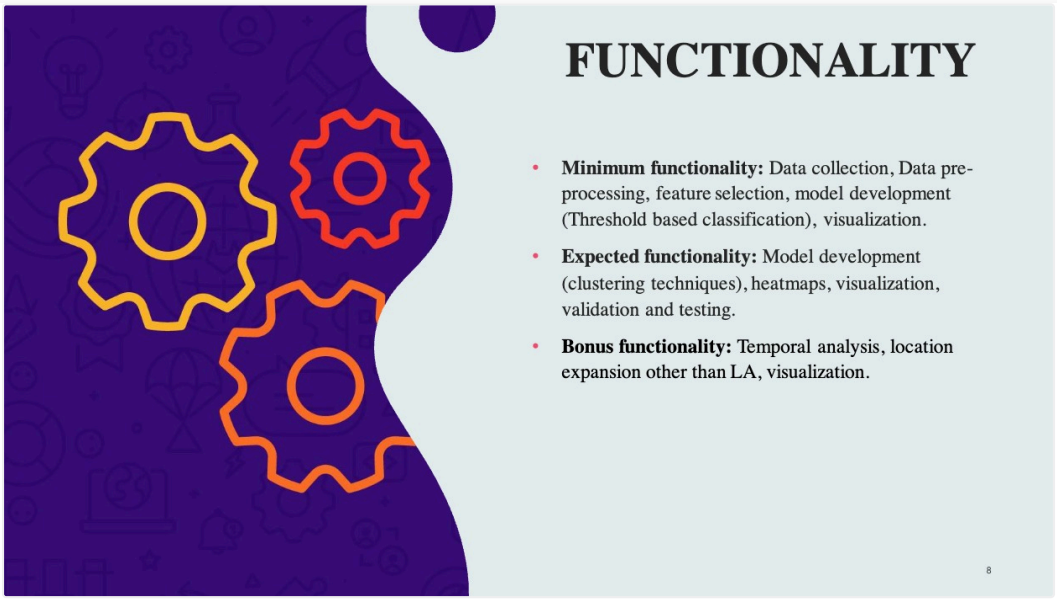
6

APPROACH

- **Threshold-based Classification**
- **Clustering Techniques**
- **Heat Maps**
- **Temporal Analysis**
- **Location expansion**



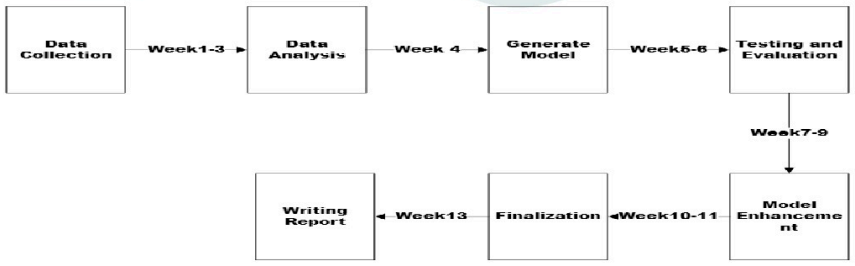
7 7


FUNCTIONALITY
Minimum functionality: Data collection, Data pre-processing, feature selection, model development (Threshold based classification), visualization.
Expected functionality: Model development (clustering techniques), heatmaps, visualization, validation and testing.
Bonus functionality: Temporal analysis, location expansion other than LA, visualization.
8

ENHANCEMENT

- Show a more detailed criminal category for each abstract criminal feature
- Evaluate the safety score for each neighbourhood by weighting for different criminal categories
- Add new features (Eg. Unemloyment rate, budget allocation) and determine the correlation between them and the crime hotspot.

PLANNING TIMELINE



10


QUESTIONS?
11