Chicago Crime analysis and prediction

Group Name:

1. **Group Info**
   Group Members: Tommaso Tognoli, Gokhulnath Thirumaran

2. **Introduction**
   In the fields of data mining and machine learning there are many algorithms and models that can be used to extract useful information from large datasets. Each of these tools comes with certain advantages and disadvantages and, today more than ever, it is crucial to select the right one to efficiently and effectively handle huge amounts of data. The goal of this project is to apply different algorithms to a very large dataset regarding crimes committed in Chicago from 2001 to the present, in order to find interesting insights on the data and compare the performance of different solutions.

3. **Problem and Goal**
   The selected dataset has 7 millions instances and over 20 features. This means that operating on the full dataset would require a lot of computational power and time. For this reason some data preprocessing might be necessary to reduce the complexity of our task (i.e. removing some redundant). Furthermore, some features might be irrelevant for our goals. Removing those features will further decrease the size of the dataset while increasing the quality of the data for our specific scenario.

   Once the data has been preprocessed, we will use different approaches to extract hidden patterns from the data. For example, are certain types of crime more common in certain areas? How has the situation evolved in the past 20 years? Are certain police departments more/less likely to conclude the investigations with an arrest?

   Finally, we will evaluate the quality of the obtained results based on the appropriate metrics of each implemented algorithm.

4. **Formalization into a ML task**
   Data type: the data we are working on is a table with 22 features including dates, strings, boolean and integer values. The dataset has no missing values but has some redundant information.

   Function: Chicago Crime Analysis and Prediction using machine learning. This function is designed for analyzing crime trends and predicting future crime occurrences in the city of Chicago. By using historical crime data, it identifies patterns based on variables such

as time of day, type of crime, location, and other contextual factors. The model will process this data to predict potential crime hotspots and high-risk periods. This prediction can be enhanced by incorporating regression models to identify the correlation between various socio-economic factors and crime rates. In real-time use, the function could analyze current data on crime and proactively predict where future incidents are most likely to occur, potentially helping law enforcement agencies to allocate resources more efficiently and improve public safety.

5. **Data Plan**
    - https://www.kaggle.com/datasets/chicago/chicago-crime/data
6. **Project Schedule**

    The tentative project schedule is the following:

Week	Task	Description

7. **References**

- S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma and N. Yadav, "Crime pattern detection, analysis & prediction," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 225-230.
- R. Yadav and S. Kumari Sheoran, "Crime Prediction Using Auto Regression Techniques for Time Series Data," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2018, pp. 1-5.
- B. Sivanagaleela and S. Rajesh, "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 595-599.
- B. K. Hemalatha, K. R. Naganandhini, R. Selvapriya, and T. P. Vinotha, "Crime pattern detection using machine learning algorithms," *ESP Journal of Engineering & Technology Advancements*
- *S. G. Lilhare, Y. Kumavat, G. Banait and A. Kurkelli, "Crime Hotspots Mapping and FIR Data Interface," 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET), Nagpur, India, 2024, pp. 1-7, doi: 10.1109/ICICET59348.2024.10616366.*
- *B. Manohar, I. Nikhil, B. Ramakrishna, D. Nagaraj, and S. J. Vivekanandan, "Crime prediction and analysis using random forest algorithm," International Journal of Progressive Research in Engineering Management and Science (IJPREMS), vol. 4, no. 5, pp. 326-331, May 2024. DOI: 10.58257/IJPREMS33981.*