



IT5006 Project Description

Fundamentals of Data Analytics

AY 2024/25 Sem 2



Introduction to Predictive Policing

Predictive policing represents the application of analytical techniques, particularly quantitative techniques, to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions [1]. The concept emerged in the late 2000s, building upon earlier approaches like CompStat, and has since evolved with advances in data analytics and machine learning [2].



Core Components

Based on some comprehensive studies [1, 2, 11], predictive policing can fall into four main categories:

1. 🌐 **Predicting Places of Increased Crime Risk**
2. 👤 **Predicting Potential Offenders**
3. 👥 **Predicting Group/Population Crime Patterns**
4. ⚠️ **Predicting Potential Victims**



Technical Foundation

Modern predictive policing systems employ various analytical techniques:

1. 🤖 **Machine Learning Algorithms**
Utilizes supervised and unsupervised learning methods to identify patterns and make predictions. Commonly employs random forests, neural networks, and support vector machines.
2. 📊 **Statistical Analysis**
Applies regression models, time series analysis, and statistical inference methods. Essential for establishing statistical significance and reliability of predictions.
3. 📖 **Other Analysis**
Includes, but not limited to, spatial analysis such as hot spot identification or examining relationships between individuals, groups, and locations to understand crime networks and social dynamics [5].



Notable Predictive Policing Efforts

1. 🎯 **PredPol (Now Known as Geolitica) [2]**

2. 🖥️ Real-Time Crime Centers [7]
3. 🗺️ Risk Terrain Modeling [9]

1 Project Overview

In this project, you will develop and evaluate predictive policing models using historical crime data, implementing various analytical approaches across various categories of predictive policing.

2 Team Composition

👥 Teams of maximum five members with complementary skills in:

- 📊 Data analysis
- 🖥️ Coding
- 📝 Documentation
- 🎤 Presentation
- 👑 Leadership
- 🤖 Ethical analysis

3 Datasets

The project can utilize two comprehensive crime datasets as stated below. You may also justify (applicability and relevance) and use any additional dataset.

1. 🗺️ Chicago Crime Dataset (2001-Present)

Official dataset from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.

- Access: <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2>
- Format: CSV, API available
- Size: over 7 million records as of 2024 (continuously updated)
- Update Frequency: Daily

2. 📄 Generalization Dataset: NIBRS (National Incident-Based Reporting System)

Used specifically to test model generalizability across different jurisdictions and contexts FBI's comprehensive database for crime reporting across the United States.

- Access: [Data Download](#)
- [NIBRS Data Dictionary and Data Diagram](#)
- Format: CSV, bulk downloads available
- Coverage: National data from participating agencies

Dataset Usage Strategy:

1. Model Development Phase:

- Use Chicago dataset for training and initial validation

- Develop and tune models using local patterns and characteristics

2. Generalization Testing Phase:





- Apply trained models to NIBRS data
- Test model performance across different:
 - Geographic regions
 - Demographic contexts, etc.
- Evaluate model robustness of developed model

Note: Students must complete any required data usage agreement forms before accessing these datasets.

Project Tasks




1. Literature Survey (2 pages max)

Review current predictive policing literature focusing on:

-  Summary of existing systems and their effectiveness
-  Review of modeling approaches
-  Feature engineering techniques
-  Evaluation metrics and their appropriateness





2. Exploratory Data Analysis

Using any tool of choice (Tableau or Power BI), perform comprehensive analysis of patterns in the data. For example:

-  Temporal pattern analysis
-  Spatial distribution study
-  Crime correlation analysis

3. Preprocessing




Based on your EDA insights, implement necessary preprocessing steps:

-  Data cleaning and normalization
-  Missing value handling
-  Temporal or/and Spatial aggregation
-  Feature standardization:

4. Feature Engineering

This phase involves creating relevant features that capture various aspects of crime patterns:

-  Temporal features





-  Spatial features
-  Environmental features
-  Interaction features

5. Model Development

Implement multiple (maximum 4) models for your proposed problem. Your analysis and choice of models can go beyond the material covered in the class.

6. Model Evaluation

Implement evaluation strategies based on available data and proposed problem. Some examples include:



-  **Standard Classification Metrics**
-  **Spatial Accuracy Evaluation**
-  **Temporal Accuracy Measures**
-  **Model Robustness Analysis**

Note: Performance metrics should focus on available data fields (location, time, crime type) and avoid metrics requiring unavailable data (response times, patrol information, resource allocation).

7. Ethical Analysis

Conduct a brief analysis of ethical implications based on available data.

For example:

-  **Geographic Distribution Analysis:**
 - Examine prediction rates across different beats/districts
 - Analyze if certain areas or crime types are over/under-predicted
 - Compare model performance across different areas
 - Document any systematic differences in prediction accuracy
-  **Model Fairness Analysis:**
 - Compare prediction accuracy across:
 - Different times of day or days of week
 - Different crime types or locations
 - Document any systematic performance variations
 - Identify potential sources of prediction bias

Key Considerations:

1. Data Limitations:
 - Working only with provided fields
 - No individual-level analysis
 - No demographic information available

- Limited contextual information

2. Privacy Requirements:

- Minimum geographic area sizes
- Minimum time windows
- Data aggregation rules

3. Documentation Needs:

- Clear methodology description
- Explicit limitation statements
- Usage guidelines

4. Review Process:

- Regular bias checks
- Performance audits across categories
- Periodic methodology review


Note: This analysis focuses on what can be evaluated using available data fields only. More comprehensive ethical analysis would require additional information.

5 Final Report Structure (13 pages)

- 📄 Cover page (1 page)
- 📊 Dataset and preprocessing (2 page)
- 🔍 Exploratory analysis (2 pages)
- 🤖 Modeling approaches (Max 4 methods; 3 pages)
- 📈 Results and evaluation (3 pages)
- ⚖️ Ethical analysis (1 page)
- 📖 References (1 page)

6 Deliverables and Deadlines

| Milestone | Task | Deliverable | Due Date | Weight |
|-----------|------------------------|--|-------------------------------|--------|
| 0 | 👥 Team Formation | Canvas signup | Week 2 | - |
| 1 | 📖 Literature Survey | Two-page report | Week 5 | 15% |
| 2 | 📊 EDA & Preprocessing | Code & report | Week 8 (Milestone opens) | 25% |
| 2 | 💻 Model Development | Implementation & report (continued) | Week 11 | 30% |
| 2 | ⚖️ Evaluation & Ethics | Analysis & report (continued) | Week 12 (Milestone closes) | 10% |

| Milestone | Task | Deliverable | Due Date | Weight |
|-----------|---|--|----------|--------|
| 3 |  Final Report & Evaluation | Complete report and Project Evaluation | Week 13 | 20% |

Note:

- Milestone 2 opens with a running (or live) document that you keep on updating till week 12. Once you update your document, please submit the updated version on Canvas.
- Code repositories need to be shared separately via Canvas folder - *[Project] Milestone 2 (Codes)*. Similar to earlier point, please update this code repository as and when needed for a particular milestone.
- Project evaluation may start after the guest lecture in Week-12.
- Refer Canvas folders for more details on submission deadlines.



Contact Information

- Prakash Sukhwai
- Senior Lecturer, DISA SOC
- **Email:** pcsdisa@nus.edu.sg



References & Suggested Readings

1. Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and public policy*, 5(1), 1-6.
2. Ferguson, A. G. (2017). The rise of big data policing: Surveillance, race, and the future of law enforcement. In *The Rise of Big Data Policing*. New York University Press.
3. Willis, J. J., Mastrofski, S. D., & Weisburd, D. (2007). Making sense of COMPSTAT: A theory-based analysis of organizational change in three police departments. *Law & society review*, 41(1), 147-188.
4. Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American statistical association*, 110(512), 1399-1411.
5. Brayne, S. (2020). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press, USA.
6. Wang, B., Yin, P., Bertozzi, A. L., Brantingham, P. J., Osher, S. J., & Xin, J. (2019). Deep learning for real-time crime forecasting and its ternarization. *Chinese Annals of Mathematics, Series B*, 40(6), 949-966.
7. Ratcliffe, J. H., Taniguchi, T., Groff, E. R., & Wood, J. D. (2011). The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology*, 49(3), 795-831.

8. Farrell, G., Farrington, D. P., & Johnson, S. D. (2012). Preventing repeat victimization: A systematic review.
9. Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting. *Justice quarterly*, 28(2), 360-381.
10. Richardson, R. (2021). Racial segregation and the data-driven society: How our failure to reckon with root causes perpetuates separate and unequal realities. *Berkeley Tech. LJ*, 36, 1051.
11. Perry, W. L. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation.