

Analisi predittiva dei crimini a Chicago

Rappresentazione dettagliata della distribuzione dei crimini nella città di Chicago suddivisi per tipologia, implementando un modello di AI in grado di stimare la probabilità che un determinato crimine si verifichi in ciascun quartiere della città.

PROGETTO FINALE DI “ACADEMY AI ENGINEERING”

Laura Perulli - 09/07/2025



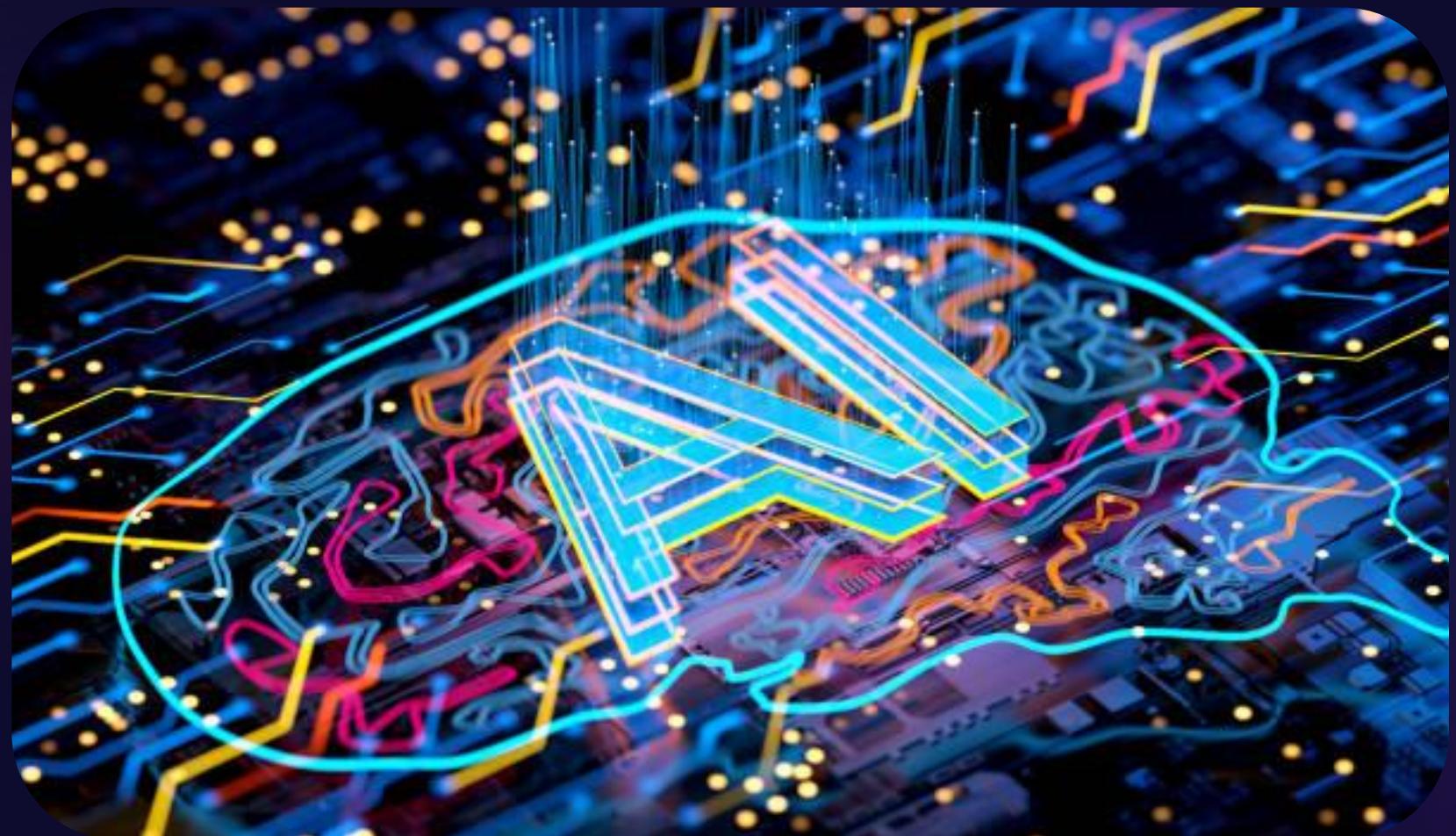
⌚ Obiettivo del Progetto

1 Prevedere la probabilità di furto

Determinare la probabilità che un crimine di tipo "THEFT" si verifichi in aree specifiche di Chicago.

2 Utilizzare il Machine Learning

Applicare tecniche avanzate di machine learning per analizzare dati storici e identificare pattern comportamentali dei crimini.



📁 Il Dataset (crimini a Chicago 2001-Oggi)

Il dataset utilizzato è una risorsa pubblica denominata “**Crimes in Chicago**”, disponibile sulla piattaforma Kaggle, che raccoglie informazioni dettagliate sui crimini registrati nella città.

- **Dimensioni originali:** Oltre 7 milioni di record di crimini.
- **Periodo:** Dati storici dal 2001 al 2025.
- **Dataset utilizzato:** Sottoinsieme di 500.000 righe per ottimizzare performance e memoria.
- **Campionamento:** 75.000 righe selezionate per visualizzazioni geografiche ottimizzate.
- **Fonte:** Chicago Police Department
- **Feature principali:**
 - Tipo di crimine
 - Data
 - Stato dell'arresto
 - Informazioni geografiche (latitudine, longitudine, area di comunità, distretto)



Analisi preliminare del Dataset

1. Pulizia dei dati:

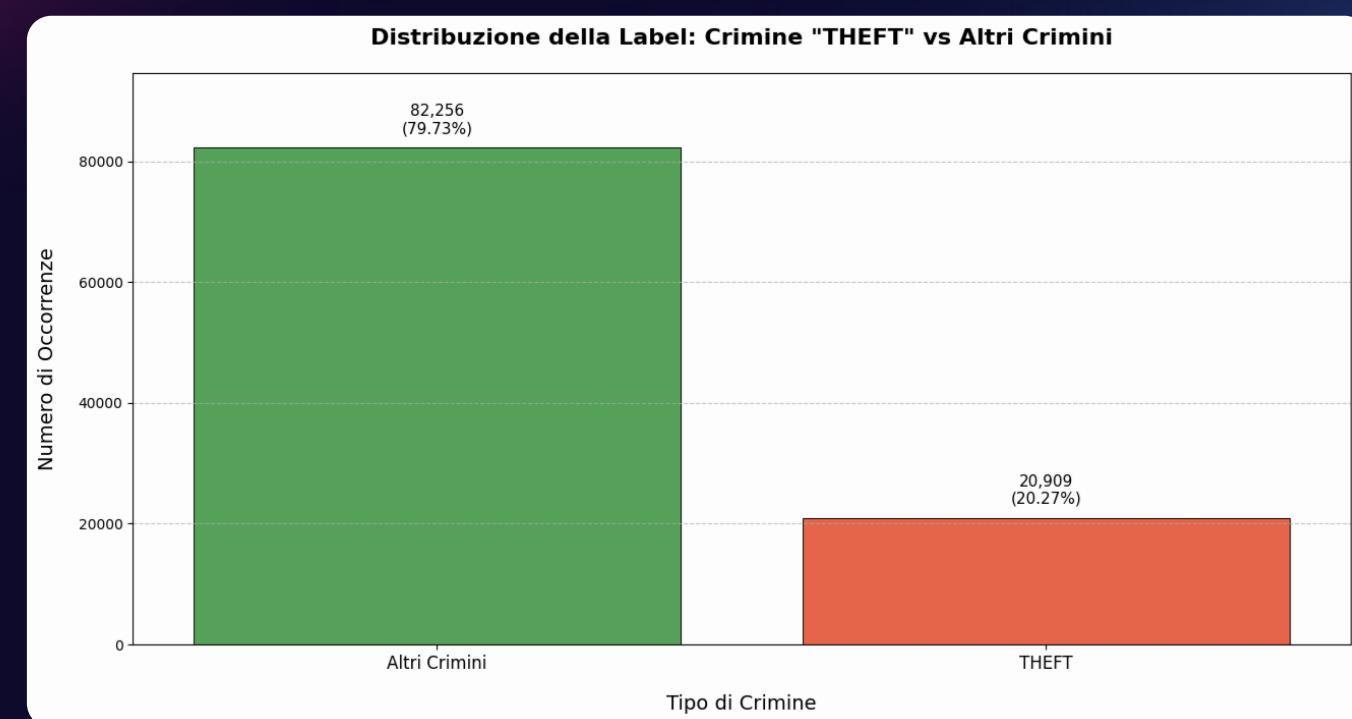
Identificazione ed eliminazione di record duplicati e gestione dei valori nulli per garantire l'integrità del dataset.

2. Selezione colonne rilevanti/irrilevanti:

Focalizzazione sulle feature più significative per l'analisi spaziale e temporale del crimine (es. *Data*, *Area Comunitaria*, *Latitudine*, *Longitudine*) e rimozione di quelle irrilevanti (es. *ID*, *Numero caso*).

3. Creazione della Label:

Definizione di una variabile target binaria *Is_Crime_Type* (1 per "THEFT", 0 per altri crimini). Nel grafico si può osservare una **distribuzione sbilanciata**, per questo è stata adottata la strategia di impostare *class_weight='balanced'*.





Visualizzazione dei dati

L'analisi esplorativa ha rivelato pattern significativi nei dati sui crimini a Chicago. I reati più diffusi includono **THEFT, BATTERY** e **HOMECIDE**.

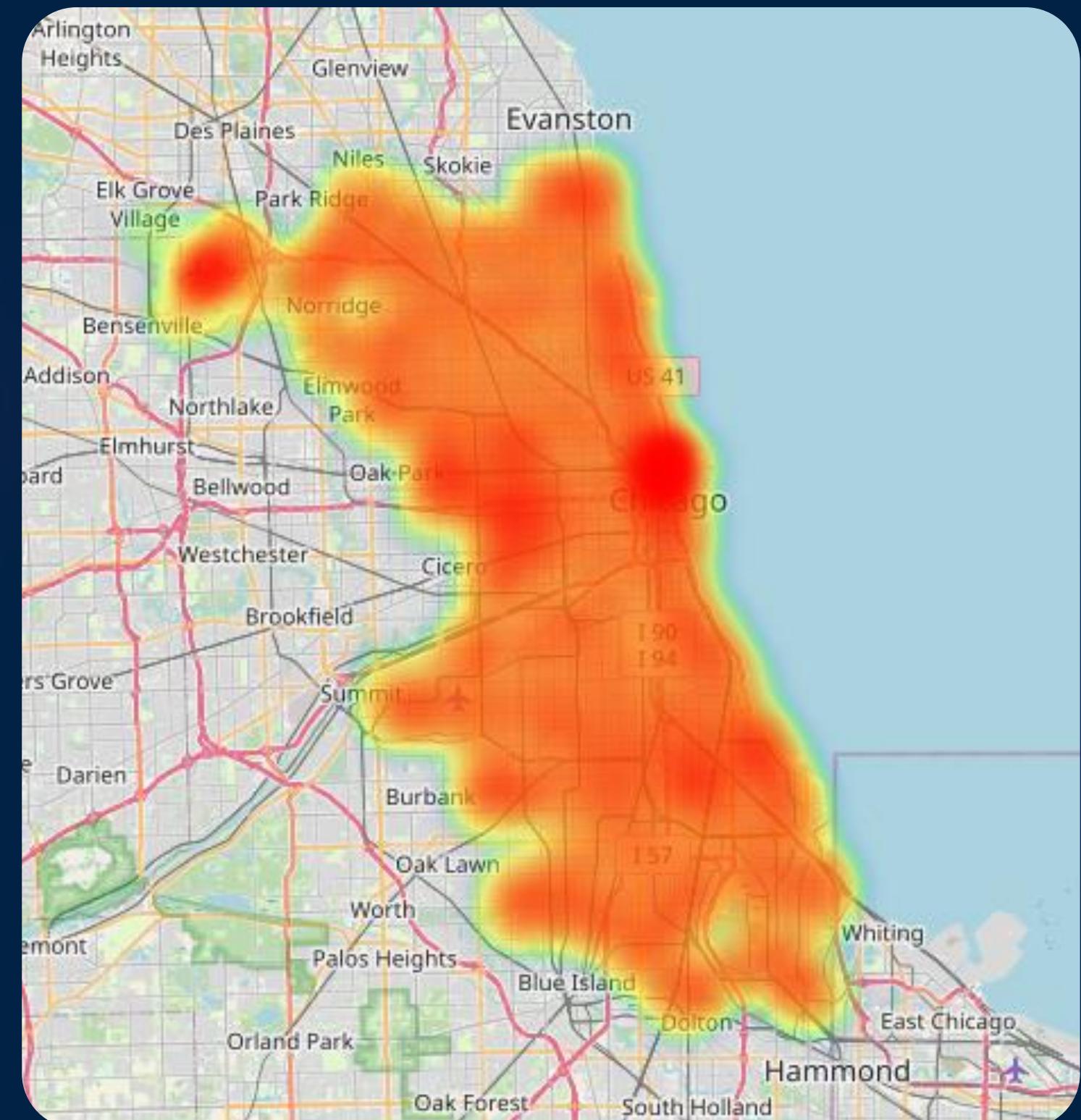
Nell'analisi **geografica/quartiere** è possibile visualizzare la densità dei crimini a Chicago.

- Le aree più **scure/rosse** indicano una maggiore concentrazione di eventi criminali.
- Maggiore comprensione visiva dei quartieri più a rischio.

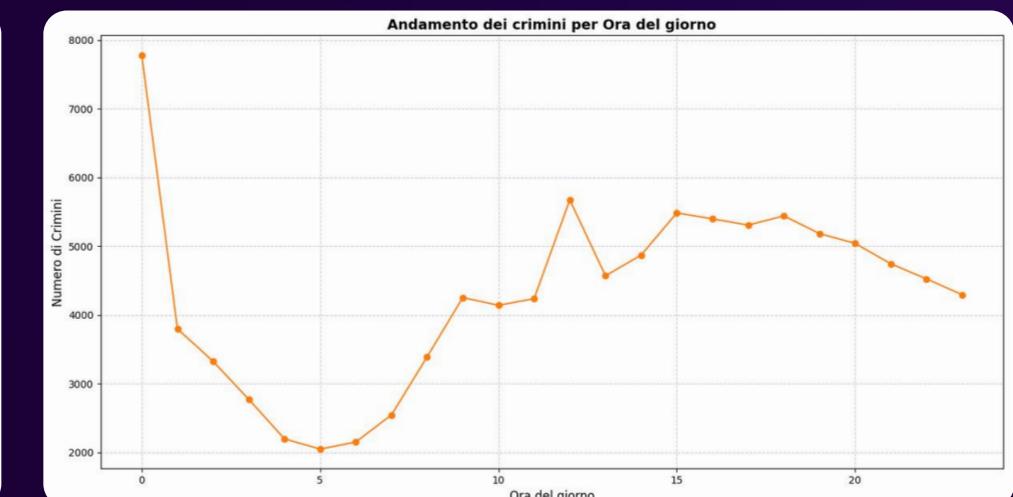
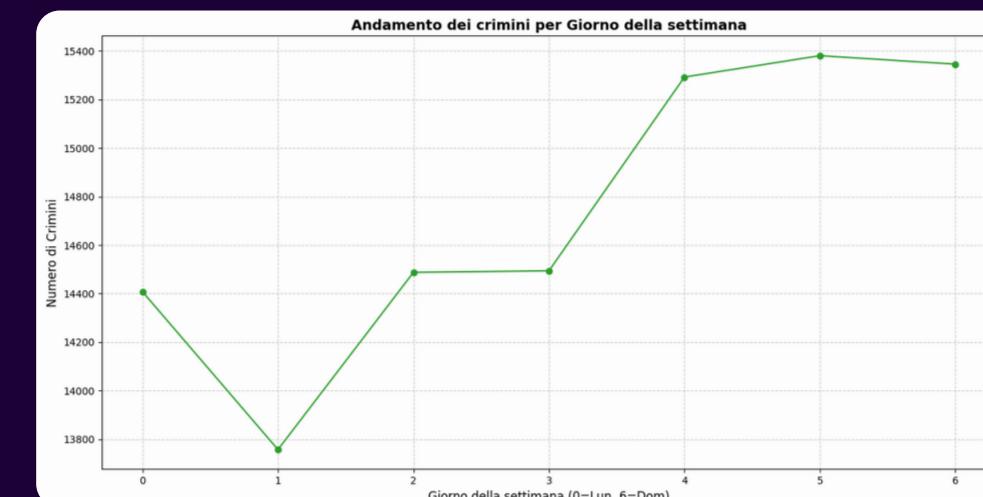
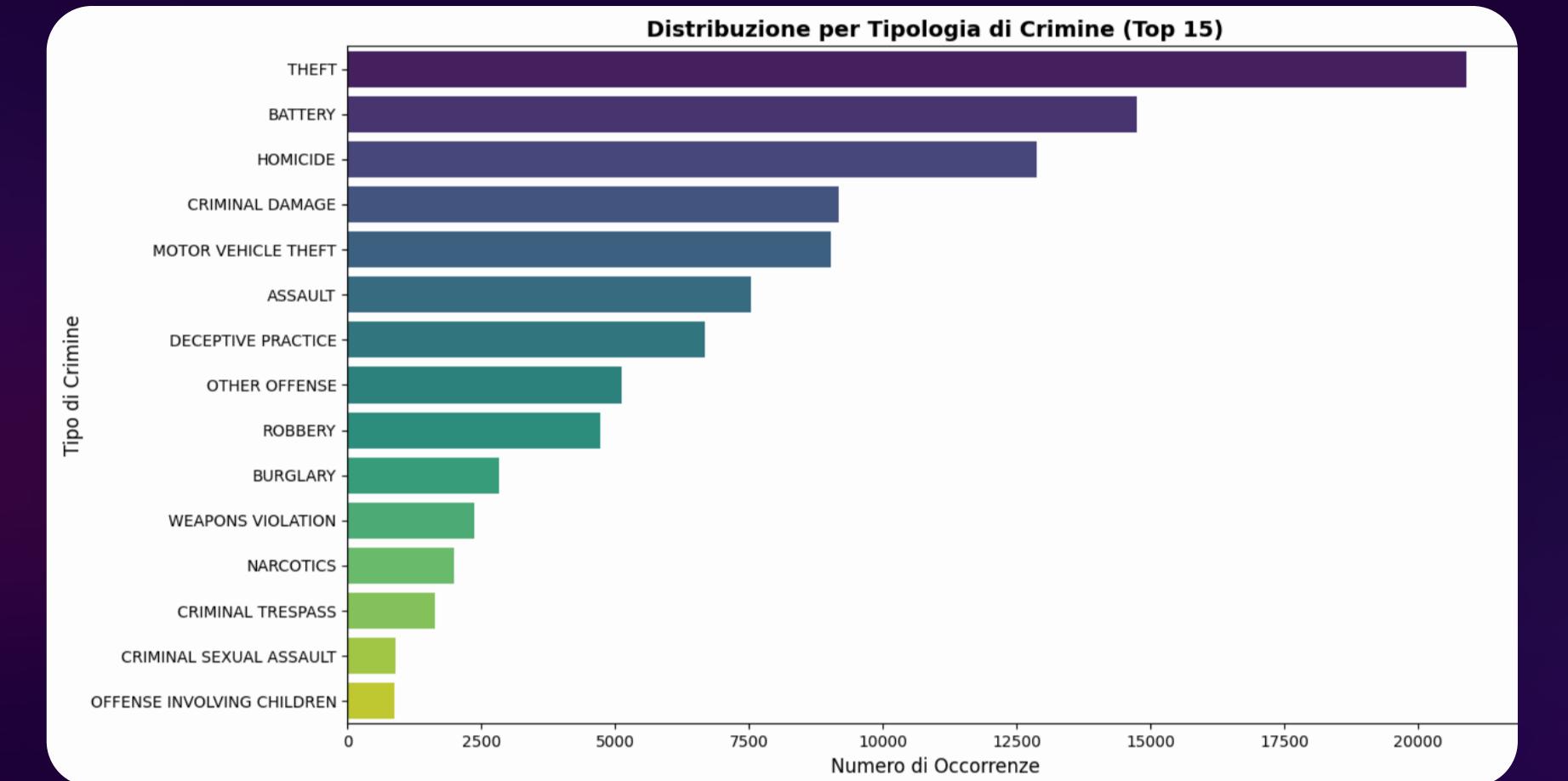
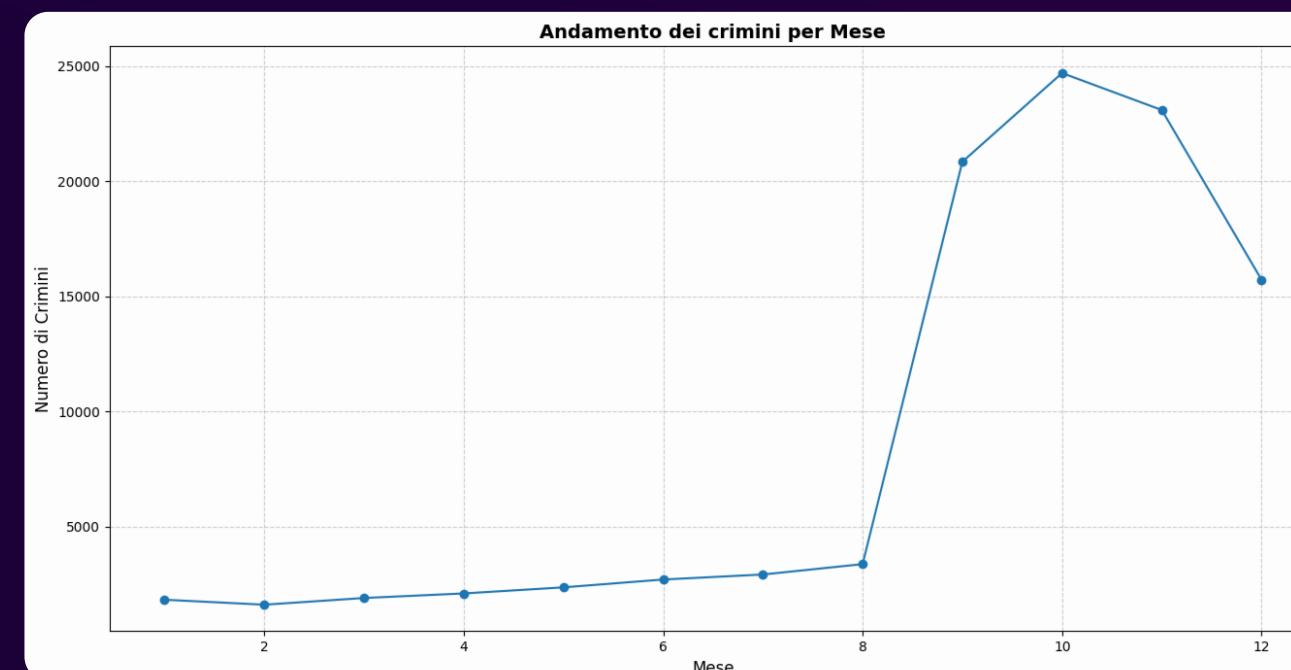
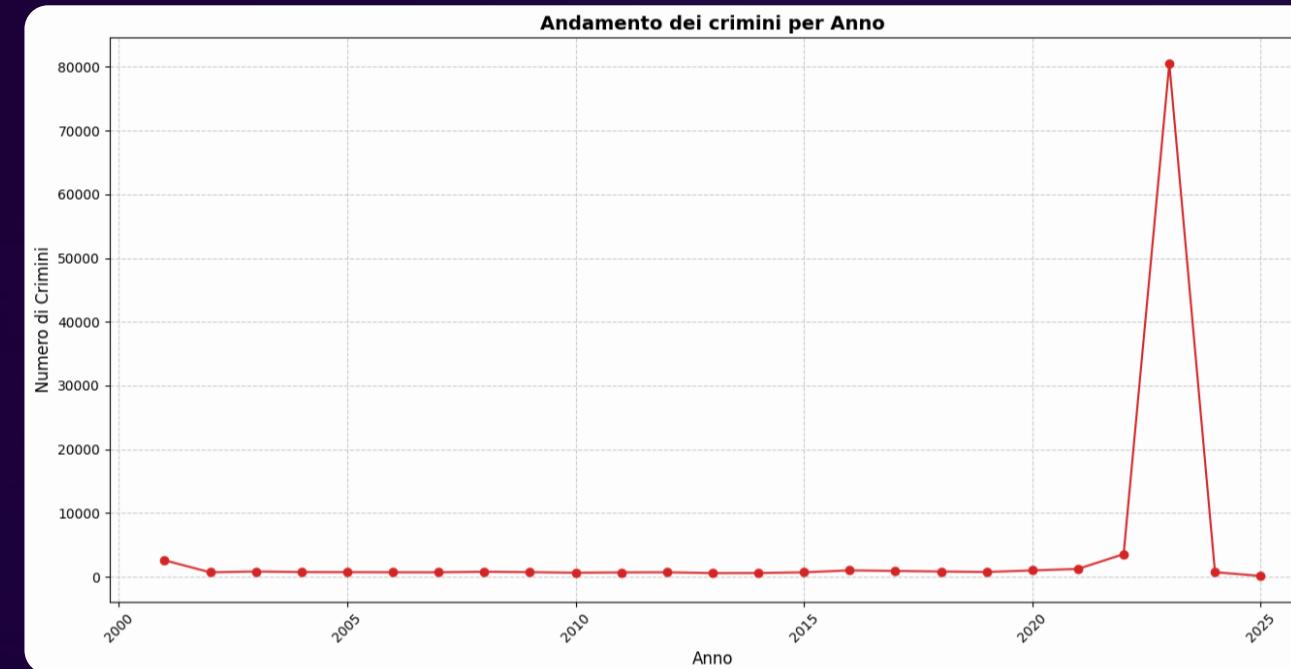
Analisi temporale:

- Il picco di attività criminale è stato osservato nel **2023**.
- Il mese di **ottobre** mostra un aumento costante dei crimini.
- Il **sabato** e la **mezzanotte** (0:00 AM) sono i periodi con il maggior numero di crimini.

Mappa di calore



Grafici esplorazione Pattern e andamento temporale





Il Modello predittivo di AI



Algoritmo

Regessione Logistica



Addestramento del modello

Valutazione delle prestazioni e Classification Report



Split Dati

Training/test 80-20% con stratificazione



Pipeline

Pre-processing per feature numeriche (StandardScaler)
ed encoding per feature categoriche (OneHotEncoder)

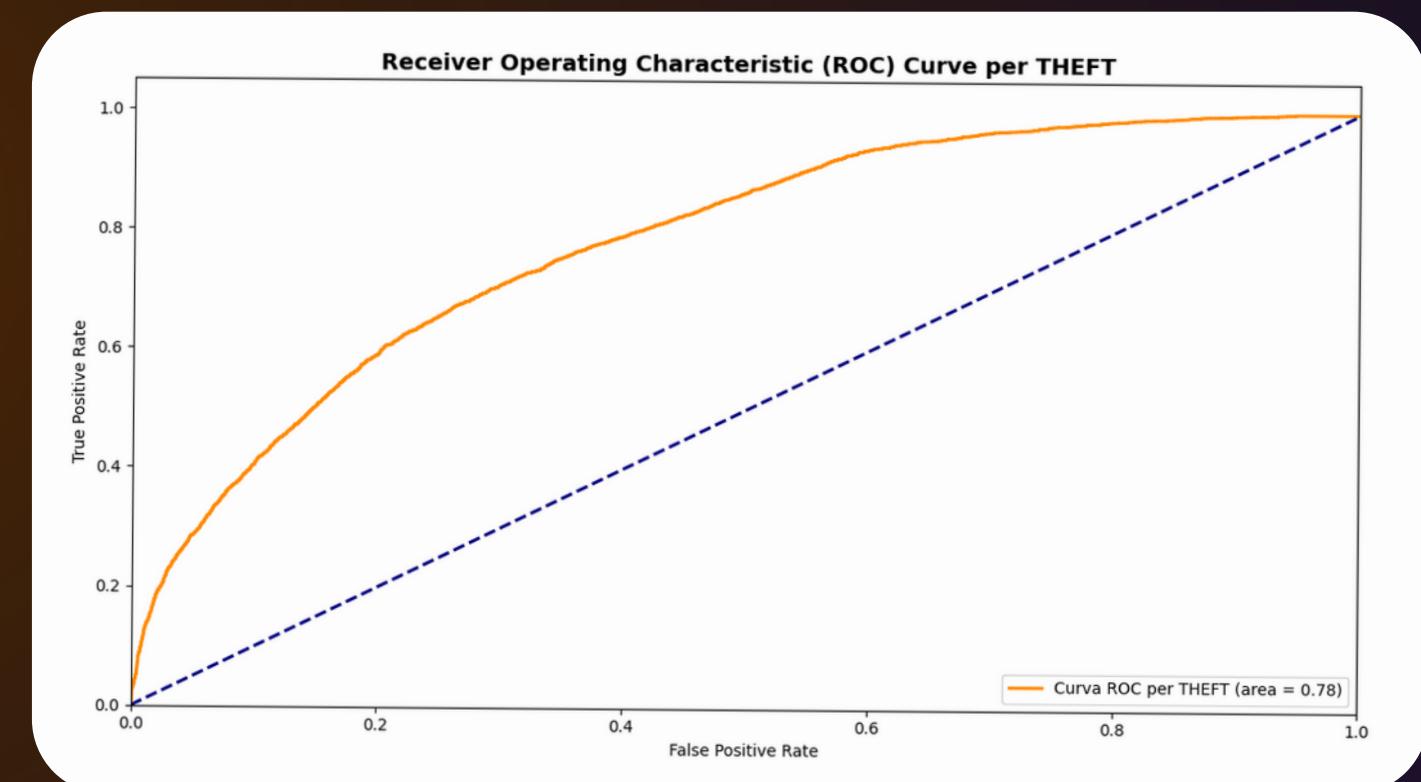


Performance del Modello

Il modello di Regressione Logistica ha dimostrato una buona capacità predittiva, con le seguenti metriche di performance:

- **Accuracy:** 0.32 - In questo caso avendo una presenza di classi molto sbilanciate, l'accuratezza da sola può essere fuorviante.
- **ROC AUC per "THEFT":** 0.78 - Un valore elevato che suggerisce una buona capacità discriminativa tra la classe "THEFT" e le altre.

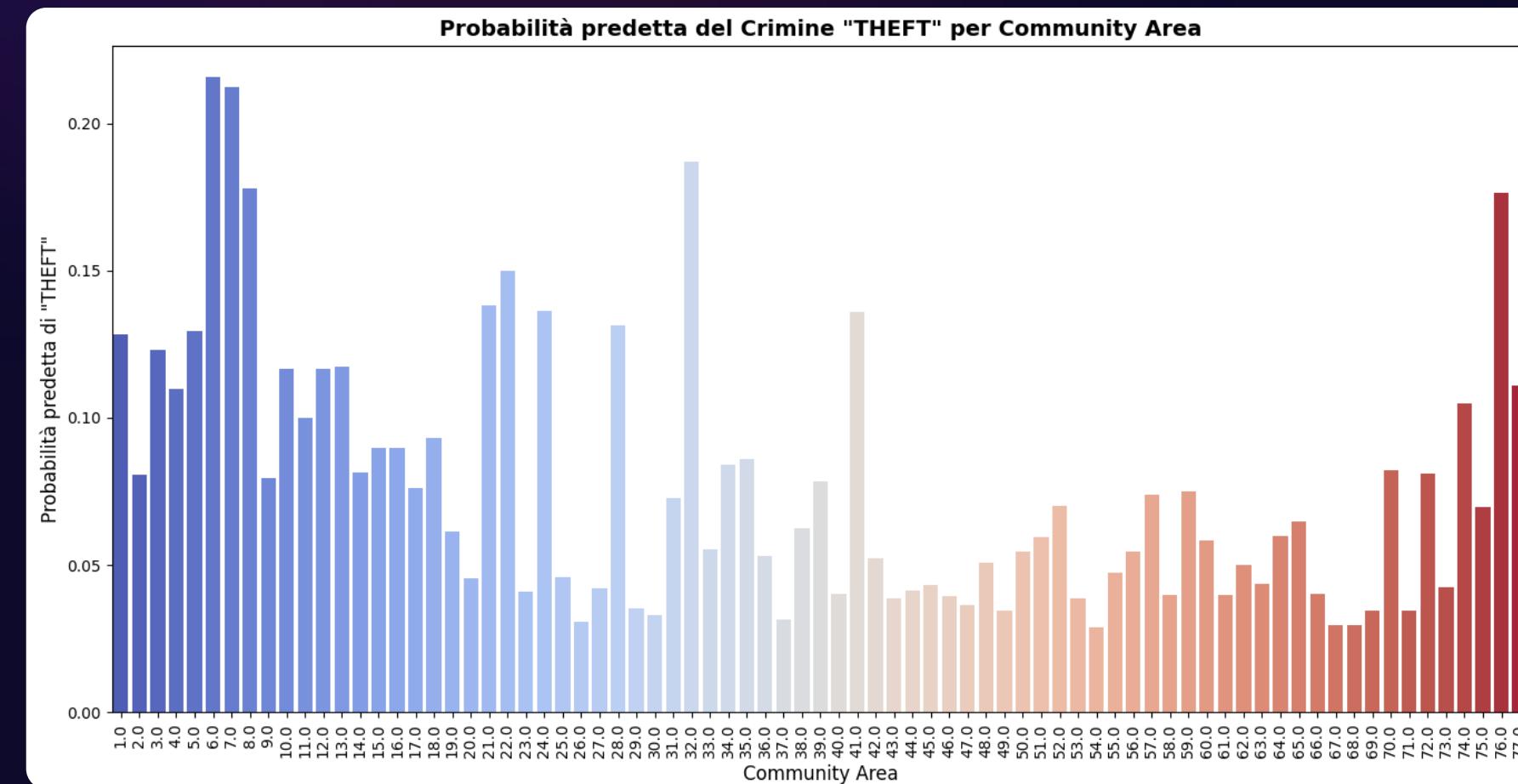
Questi risultati confermano la validità del modello nel distinguere il crimine di furto dagli altri tipi di reato, fornendo una base solida per future analisi e interventi. Il **classification report completo** fornisce metriche dettagliate (Precision, Recall, F1-score, Support) ed è disponibile per un'analisi dettagliata delle performance di ogni tipo di crimine.



💡 Probabilità predetta di Furto per quartiere

Il modello predittivo è in grado di stimare la probabilità di furto (*THEFT*) per ciascuna **Community Area** di Chicago. Questa analisi geografica è fondamentale per identificare le zone a rischio.

- **Analisi di Top e Bottom 10 Community Area:** indica le 10 aree comunitarie con la più alta e la più bassa probabilità di furto.
- **Grafico delle probabilità per tutte le Community Area:** il grafico a barre evidenzia le probabilità predette del crimine '*THEFT*', visualizzando le aree più a rischio e quelle più sicure.



Conclusioni

Obiettivo raggiunto:

Il progetto ha raggiunto l'obiettivo. Attraverso l'analisi esplorativa di un vasto dataset di crimini a Chicago, è stato creato un robusto modello predittivo di Regressione Logistica con cui è possibile identificare e visualizzare le tendenze spaziali e temporali della criminalità con un focus specifico sui furti.



Completezza del Progetto

Acquisite competenze durante il corso, con teoria ed esercizi, utilizzando tecniche di Machine Learning.



Precisione del Modello

ROC AUC per 'THEFT' di 0.78, indicatore di eccellente capacità discriminatoria del modello.

Prospettive Future

Migliorare l'accuratezza con altri modelli avanzati di ML (Random Forest, KNN, Reti Neurali, ecc.).
Se possibile integrare dati aggiuntivi come quelli socio-economici e sviluppare un'interfaccia utente interattiva per le previsioni in tempo reale.



Comprensione dei Dati

Approfondita visualizzazione e analisi di crimini per tipo, distribuzione geografica e pattern temporali.



Applicabilità

Generazione delle probabilità di furto per Community Area e identificazione delle zone a rischio.

GRAZIE!!