**Esercitazione Progetto: Analisi Predittiva dei Crimini a Chicago** **Obiettivo dell’esercitazione**

L’obiettivo dell’esercitazione è la realizzazione di un progetto software in Python che analizzi il dataset disponibile al seguente link:

🔗 [Chicago Crime Dataset – Kaggle](https://www.kaggle.com/datasets/chicago/chicago-crime)

L’analisi dovrà fornire una rappresentazione dettagliata della distribuzione dei crimini nella città di Chicago, suddivisi per tipologia, e includere un modello di intelligenza artificiale in grado di stimare la probabilità che un determinato crimine si verifichi in ciascun quartiere della città.

**Compiti richiesti**

Ogni studente dovrà lavorare individualmente per realizzare un progetto che includa le seguenti componenti:

**1. Analisi preliminare del dataset**

* Scaricare il dataset da Kaggle e caricarlo in un ambiente di analisi dati (es. Jupyter Notebook, Google Colab o IDE locale).
* Pulire i dati rimuovendo le informazioni inutili o inconsistenti (es. valori nulli, record duplicati, ecc.).
* Individuare e descrivere:
  + Le colonne considerate **rilevanti** per l’analisi predittiva.
  + Le colonne considerate **irrilevanti** e giustificare l’esclusione.
* Selezionare la **label** (cioè, la variabile da predire). La scelta della label è a discrezione dello studente (ad es. tipo di crimine, probabilità di accadimento, ecc.) ma deve essere motivata.

**2. Visualizzazione dei dati**

* Rappresentare graficamente la distribuzione dei crimini per:
  + **Tipologia** (es. furto, omicidio, rapina, ecc.)
  + **Zona geografica/quartiere** (es. mappa di calore, clustering, ecc.)
  + **Andamento temporale** (opzionale ma consigliato)
* Utilizzare strumenti di visualizzazione come: matplotlib, seaborn, plotly, folium, ecc.

**3. Modello di intelligenza artificiale**

* Costruire un modello predittivo che stimi, per ciascun quartiere, la **probabilità** che si verifichi un determinato tipo di crimine.
* Selezionare un algoritmo di machine learning tra quelli studiati (es. Random Forest, Logistic Regression, SVM, KNN, Reti Neurali, ecc.)
* Giustificare la scelta dell’algoritmo in funzione dei dati e del tipo di problema.
* Valutare le prestazioni del modello con metriche appropriate (accuratezza, precision, recall, F1-score, AUC, ecc.).

**4. Output del progetto**

Ogni studente dovrà produrre:

1. **Report in formato PDF o DOCX** che includa:
   * Descrizione dell’analisi esplorativa
   * Motivazione delle scelte effettuate
   * Descrizione del modello AI scelto
   * Discussione dei risultati ottenuti
2. **Codice Python commentato**, organizzato in moduli chiari, che permetta:
   * L’importazione e la pulizia dei dati
   * La visualizzazione e analisi dei crimini
   * L’addestramento e valutazione del modello
   * La previsione della probabilità dei crimini per quartiere
3. **Presentazione PowerPoint (max 10 slide)** che illustri sinteticamente:
   * Il problema affrontato
   * Il dataset
   * Le principali visualizzazioni e scelte
   * Il modello predittivo e i risultati

**Consegna**

* Tutti i materiali (report, codice Python e presentazione PowerPoint) dovranno essere consegnati in formato elettronico (cartella compressa .zip o repository GitHub).
* La scadenza per la consegna verrà comunicata dal docente.

**Valutazione**

La valutazione terrà conto dei seguenti criteri:

|  |  |
| --- | --- |
| **Criterio** | **Peso** |
| Pulizia e comprensione dei dati | 20% |
| Scelte motivate nella selezione delle variabili e della label | 15% |
| Completezza e chiarezza delle visualizzazioni | 15% |
| Qualità del modello AI e motivazione della scelta | 25% |
| Presentazione e documentazione | 15% |
| Originalità dell’approccio individuale | 10% |

**Suggerimenti**

* Assicuratevi di comprendere il significato di ogni colonna del dataset.
* Non è necessario usare tutto il dataset completo (oltre 7 milioni di righe); potete filtrarlo o campionarlo per l'addestramento iniziale.
* Annotate nel codice ogni fase dell’elaborazione.
* Potete consultare documentazione esterna, ma il lavoro deve essere personale.