HD Task

About this task

Step-1

This task is designed to assess the Distinction level expectations. Please select the set based on your unit code. DO NOT SUBMIT BOTH SETS OF EVIDENCE.

Step-2

You will not receive any feedback for this task and the assessor will assess this task as part of portfolio assessment.

Feedback and submission deadlines

Feedback deadline: Not applicable

Submission deadline: Before creating and submitting portfolio.

Required documents.

1. **Submit a report (pdf format) in Ontrack (https://ontrack.deakin.edu.au)**
2. **Complete the problem credit task and submit your code file (.ipynb) separately in the OnTrack (https://ontrack.deakin.edu.au).**
3. **Submit your presentation either by uploading directly in OnTrack or by sharing the link to a cloud storage.**

Background

Crime and violations pose significant threats to justice and must be effectively controlled. Employing precise crime prediction and future forecasting trends can substantially bolster metropolitan safety through computational means. However, the inherent limitations of the human capacity to process vast amounts of complex information from big data impede early and accurate crime prediction and forecasting.

Datasets Description

- The Chicago Crime dataset contains a summary of the reported crimes occurred in the City of Chicago from 2001 to November 2019 (data link: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Dashboard/5cd6-ry5g)
- Dataset has been obtained from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.
- Dataset contains the following columns:
    - ID: Unique identifier for the record.
    - Case Number: The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.
    - Date: Date when the incident occurred.
    - Block: address where the incident occurred
    - IUCR: The Illinois Unifrom Crime Reporting code.
    - Primary Type: The primary description of the IUCR code.
    - Description: The secondary description of the IUCR code, a subcategory of the primary description.
    - Location Description: Description of the location where the incident occurred.
    - Arrest: Indicates whether an arrest was made.
    - Domestic: Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.
    - Beat: Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car.
    - District: Indicates the police district where the incident occurred.
    - Ward: The ward (City Council district) where the incident occurred.
    - Community Area: Indicates the community area where the incident occurred. Chicago has 77 community areas.
    - FBI Code: Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).
    - X Coordinate: The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection.
    - Y Coordinate: The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection.
    - Year: Year the incident occurred.
    - Updated On: Date and time the record was last updated.
    - Latitude: The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
    - Longitude: The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
    - Location: The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

Evidence of Learning- SIT307

1. Read the article and reproduce the results (Accuracy, Precision, Recall, F-Measure) for Chicago Crime dataset using following classification methods:
    - Logistic regression
    - SVM
    - KNN
    - Random forest
    - XGBoost

    These results can be found in **Table 1** (**Performance parameters for Chicago**) of the manuscript and should be used for comparison purposes, if required. Write a report summarising the dataset, used ML methods, experiment protocol and results including variations, if any. During reproducing the results:

    i) you should use the same set of features used by the authors.
    ii) you should use the same classifier with exact parameter values.
    iii) you should use the same training/test splitting approach as used by the authors.
    iv) you should use the same pre/post processing, if any, used by the authors.

    N.B.

    (i) If you find any issue in reproducing results due to incomplete description of model in the provided article, then make your own assumption and explain the reason. If your justification is correct, then your solution will be considered correct and assessed accordingly.
    (ii) Similarly, variation in results due to randomness of data splitting will also be considered during evaluation based on your explanation.
    (iii) Obtained marks will be proportional to the number of ML methods that you will report in your submission with correctly reproduced results.
    (iv) Make sure your submitted Python code segment generates the reported results, otherwise you will receive zero marks for this task.

2. Design and develop your own ML solution for this problem. The proposed solution should be different from all approaches mentioned in the provided article. This does not mean that you must have to choose a new ML algorithm. You can develop a novel solution by changing the feature selection approach or using different ML methods or different combinations of them. This means, the proposed system should be substantially different from the methods presented in the article but not limited to only change of ML methods. Compare the result with reported methods in the article. Write a technical report summarising your solution design and outcomes. The report should include:
    - Detail description of the model including all parameters so that any reader can implement your model.
    - Description of experimental protocol.
    - Evaluation metrics.
    - Present results using tables and graphs.
    - Compare and discuss results with respect to existing literatures.
    - Appropriate references (IEEE numbered).

3. Present your result in a 3 minutes video using PowerPoint slides/animation.

Evidence of Learning- SIT720

1. Read the article and reproduce the results (Accuracy, Precision, Recall, F-Measure) for Chicago Crime dataset using following classification methods:
   - Logistic regression
   - SVM
   - Naïve Bayes
   - MLP
   - KNN
   - Random forest
   - XGBoost

   These results can be found in **Table 1** (**Performance parameters for Chicago**) of the manuscript and should be used for comparison purposes, if required. Write a report summarising the dataset, used ML methods, experiment protocol and results including variations, if any. During reproducing the results:

   v) you should use the same set of features used by the authors.
   vi) you should use the same classifier with exact parameter values.
   vii) you should use the same training/test splitting approach as used by the authors.
   viii) you should use the same pre/post processing, if any, used by the authors.

   N.B.

   (v) If you find any issue in reproducing results due to incomplete description of model in the provided article, then make your own assumption and explain the reason. If your justification is correct, then your solution will be considered correct and assessed accordingly.
   (vi) Similarly, variation in results due to randomness of data splitting will also be considered during evaluation based on your explanation.
   (vii) Obtained marks will be proportional to the number of ML methods that you will report in your submission with correctly reproduced results.
   (viii) Make sure your submitted Python code segment generates the reported results, otherwise you will receive zero marks for this task.

2. Design and develop your own ML solution for this problem. The proposed solution should be different from all approaches mentioned in the provided article. This does not mean that you must have to choose a new ML algorithm. You can develop a novel solution by changing the feature selection approach or using different ML methods or different combinations of them. This means, the proposed system should be substantially different from the methods presented in the article but not limited to only change of ML methods. Compare the result with reported methods in the article. Write a technical report summarising your solution design and outcomes. The report should include:
   - Motivation behind the proposed solution.
   - How the proposed solution is different from existing ones.
   - Detail description of the model including all parameters so that any reader can implement your model.
   - Description of experimental protocol.
   - Evaluation metrics.
   - Present results using tables and graphs.
   - Compare and discuss results with respect to existing literatures.
   - Appropriate references (IEEE numbered).

3. Present your result in a 3 minutes video using PowerPoint slides/animation.