

1. Title:

Unsupervised Anomaly Detection on Healthcare Providers data.

Dataset Link: <https://www.kaggle.com/datasets/tamilisel/healthcare-providers-data>

2. Project Statement:

Healthcare fraud stands as a pressing issue in today's society. Diverting funds away from crucial medical needs such as medication, emergency services is considered as Healthcare fraud. The actions of unethical practitioners or patients contribute significantly to the escalating costs of healthcare.



Outcomes:

This project aims to create Machine learning models to detect the fraudulent transactions from healthcare data

Modules to be implemented

1. Data Collection and Exploration
2. EDA
3. Data Preprocessing
4. Explore various ML algorithms
5. Deep Learning approach using auto encoders
6. Presentation and Documentation

3. Week-wise module implementation and high-level requirements with output screenshots

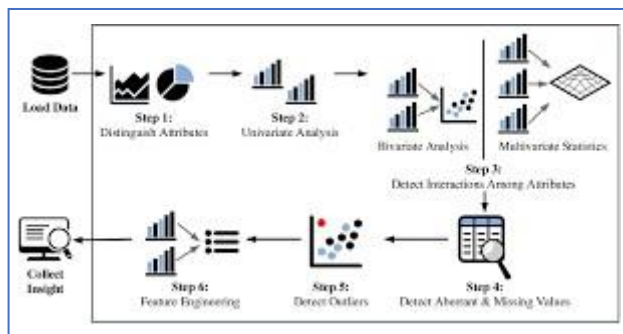
Milestone 1: Weeks 1-2

Module 1: Data Collection & Exploration

- Understand the problem statement
- Collect data and understand the importance of the feature set

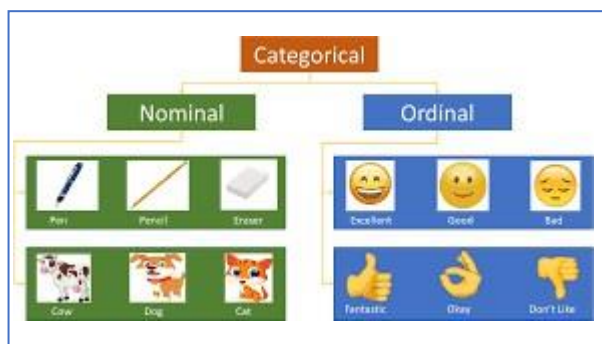
Module 2: EDA

- Exploratory data analysis of the data (Univariate, Bivariate analysis)
- Visualization of data using different plots.



Module 3: Data Preprocessing

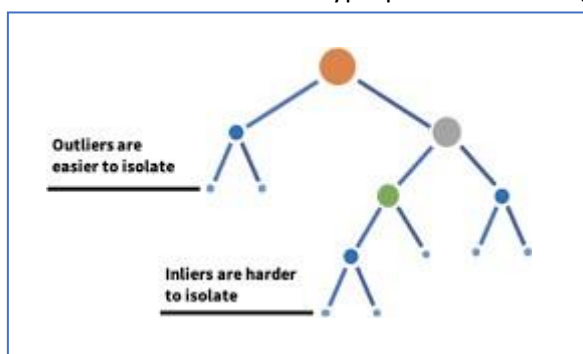
- Missing Value Treatment
- Encoding techniques for categorical values
- Normalization and standardization of numerical columns.



Milestone 2: Weeks 3-5

Module 4: ML Algorithms & Hyperparameter tuning

- Explore multiple anomaly detection algorithms
- Apply ML algorithms on transformed data
- Visualize results and finalize the best.
- Finetune final ML model
- Visualize results after hyperparameter tuning

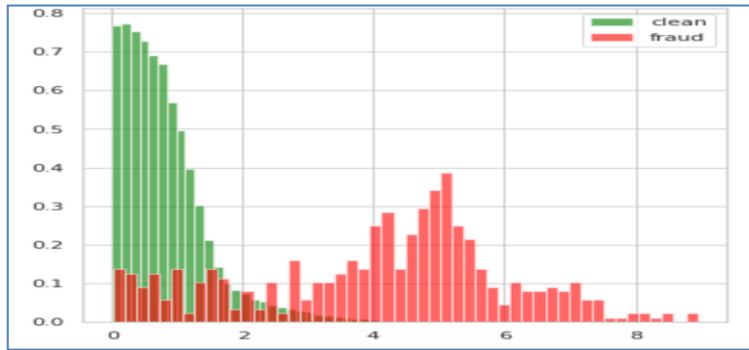


Milestone 3: Weeks 6-7

Module 5: Deep Learning Approach using Auto Encoders

- Understand working of Auto Encoder Model
- Apply Autoencoder model on the data
- Visualize results

The final output histogram plot should be similar to below plot



Milestone 4: Weeks 8

Module 6: Presentations and Documentation

Prepare a presentation which must include the details of the problem statement, details of the data collected, data preprocessing methods and its outcomes, model building methodology, Hyperparameters, performance metrics and recommendations based on the outcome.

- Project document which should capture the same topics mentioned above in more detailed format.

Evaluation Criteria:

Milestone 1 Evaluation (Week 2):

- Understand the problem statement and the dataset
- Approval on the EDA findings
- Approval on the encoding techniques used
- Approval on visualizations

Milestone 2 Evaluation (Week 3-5):

- Approval on final dataset created after pre-processing
- Approval on applying ML algorithms
- Approval on finetuning the hyperparameters
- Approval on visualizations created on ML algorithms results

Milestone 3 Evaluation (Week 6-7):

- Approval on usage of Deep Learning methods on data
- Approval on visualizations

Milestone 4 Evaluation (Week 8):

- Approved Final Model.
- Approved Presentation and Project Documentation.
- Approved Remediation/Action plans for the Business.
- Final Code Submissions on GitHub.