Haritha Akella

Data science

Project Proposal

Table of Contents

[Context Scenario 2](#_Toc434936228)

[Problem-Solution 2](#_Toc434936229)

[Project Metrics 3](#_Toc434936230)

[Features 3](#_Toc434936231)

[Appendix. 4](#_Toc434936233)

# Context Scenario

|  |  |
| --- | --- |
|  | * Consider real dataset – There are issues with real dataset:   + **Security & confidentiality**   + Privacy   + Scalable algorithms   + Ease of use * Consider synthetic dataset – Advantages of synthetic dataset   + protect the privacy and confidentiality   + used in the process of datamining |

# Problem-Solution

|  |  |
| --- | --- |
|  | Various techniques for obtaining synthetic dataset:   * + Linear regression model   + Clustering   + Outlier analysis   + By mean , standard deviation techniques   + Tools: TDS generators   Find the best technique to solve the issues such as finding frequent itemsets, Optimization problems.  Compare real and synthetic datasets results and formulate observations and conclusions. |

# Project Metrics

|  |  |
| --- | --- |
|  | Datasets considered for experiment:   * <https://archive.ics.uci.edu/ml/datasets/QtyT40I10D100K> * <http://recsyswiki.com/wiki/Grocery_shopping_datasets> * <https://www.kaggle.com/c/acquire-valued-shoppers-challenge/data> * <http://www.kde.cs.uni-kassel.de/bibsonomy/dumps> * <http://www.dai-labor.de/en/competence_centers/irml/datasets/> * <http://grouplens.org/datasets/movielens/> * <http://www.cs.sfu.ca/~sja25/personal/datasets/> * <http://recsyswiki.com/wiki/Recommendation_Datasets> * <http://www.kdnuggets.com/datasets/index.html> * <http://www.academia.edu/3096984/WebDocs_a_real-life_huge_transactional_dataset> |

# Features

|  |  |
| --- | --- |
|  | Features include:   * Avoids issues with real datasets * Utilizes the benefit of synthetic datasets * Experience the techniques to obtain synthetic datasets * Obtain the best techniques to solve certain problems on specific type of datasets. |

# Appendix.

.The link below is the reference for the project :

1. https://github.com/harithaakella/DSA\_Project