



Business Analytics papers (IS)

1. Telco Customer Churn Dataset

Used In: Customer Churn Prediction in Telecom Using Machine Learning

Mining Type: Classification

Paper Link: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6>

Dataset Link: <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

2. Credit Card Fraud Detection Dataset

Used In: Credit Card Fraud Detection Using Data Mining Techniques

Mining Type: Classification

Paper Link:

https://www.researchgate.net/publication/344788652_CREDIT_CARD_FRAUD_DETECTION_USING_DATA_MINING_TECHNIQUES

Dataset Link: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

3. Amazon Reviews for Sentiment Analysis

Used In: Sentiment Analysis of Consumer Reviews Using Deep Learning

Mining Type: Classification / Text Mining

Paper Link:

https://www.researchgate.net/publication/363181339_Sentiment_Analysis_of_Consumer_Reviews_Using_Deep_Learning

Dataset Link: <https://www.kaggle.com/datasets/bittlingmayer/amazonreviews>

4. CMS DE-SynPUF Data

Used In: Healthcare Insurance Fraud Detection Using Data Mining

Mining Type: Association / Anomaly Detection

Paper Link:

<https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02512-4>

Dataset Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs>



5. Groceries Dataset for Market Basket Analysis

Used In: Market Basket Analysis of Basket Data with Demographics

Mining Type: Association Rule Mining

Paper Link:

https://www.researchgate.net/publication/352806803_Market_Basket_Analysis_of_Basket_Data_with_Demographics_A_Case_Study_in_E-Retailing

Dataset Link: <https://www.kaggle.com/datasets/rashikrahmanpritom/groceries-dataset-for-market-basket-analysismba>

6. Wholesale Customers Dataset

Used In: Customer Segmentation Using K-Means Clustering

Mining Type: Clustering

Paper Link:

https://ijaem.net/issue_dcp/Customer%20Segmentation%20using%20K%20means%20Clustering.pdf

Dataset Link: <https://archive.ics.uci.edu/ml/datasets/wholesale+customers>



Software Analytics papers (CS)

7. NASA PROMISE Datasets

Used In: A Comparative Analysis for Software Fault Detection Using Machine Learning Techniques

Mining Type: Classification

Paper Link:

https://www.researchgate.net/publication/384588144_A_Comparative_Analysis_for_Software_Fault_Detection_Using_Machine_Learning_Techniques

Dataset Link: <https://github.com/ApoorvaKrisna/NASA-promise-dataset-repository>.

8. Bugzilla Bug Reports

Used In: Predicting the Severity of a Reported Bug

Mining Type: Classification

Paper Link:

https://www.researchgate.net/publication/224137662_Predicting_the_severity_of_a_reported_bug

Dataset Link: <https://bugzilla.mozilla.org/>

9. Code Smells Dataset

Used In: Code Smell Detection Using Ensemble Machine Learning Algorithms

Mining Type: Classification

Paper Link: <https://www.mdpi.com/2076-3417/12/20/10321>

Dataset Link: <https://zenodo.org/record/6555241>