CSE0442 Data Warehouses and Data Mining

Project Rules

General Information

Project Team	Required Program	Minimum Word Count	Deadline
Up to two students	MATLAB	2000	22.05.2024 11:55 p.m.

Project Description

Implementation

- Determine and use a public data set (can be downloaded from Machine Learning Repository).
- If there are records with missing values, delete them or replace them with average values.
- Test and use a feature extraction/dimensionality reduction technique (such as PCA).
- Choose at least 7 data mining/machine learning techniques for implementation.
- Use holdout validation (80% for training and %20 for test) and 10-fold cross validation methods for data dividing.
- Compare the performance of chosen data mining/machine learning techniques using following evaluation metrics: classification accuracy, precision, recall (sensitivity) and F-measure (F-score).

Report Preparation

- Prepare a project report using IEEE Conference Template (can be downloaded from IKU, Department of Computer Engineering, Graduation Project page: https://bm.iku.edu.tr/sites/bilm/files/inline-files/%283%29Graduation%20Project%20Report%20IEEE%20Conference%20Template.doc)
- Project report must include the following sections: Abstract (with Keywords), Introduction, Related Work, Methodology (including subsections "Methods", "Dataset Description" and "Evaluation Metrics"), Experimental Results, Conclusions and References.
- **In the Introduction Section**, give general information about the topic.
- **In the Related Work Section**, summarize at least 15 existing papers (conference or article).
- **In the Methodology Section**
  - **In the Methods Subsection**, briefly mention about chosen techniques and feature extraction/dimensionality reduction technique.
  - **In the Data Description Subsection**, describe the selected dataset (including # records, feature names, feature value types, feature value ranges, and # missing values).
  - **In the Evaluation Metrics Subsection**, explain the metrics and write their formulas.
- **In the Experimental Results Section**, write your system configurations (including operating system, CPU, RAM and used program) and compare the performance of selected classifiers in terms of these metrics (create tables for comparing).
- **In the Conclusions Section**, provide discussion of the overall coverage of the project.
- **In the References Section**, list the resources used in the report.

Report Delivery

- Send your report over the CATS system as a **Word** document before the deadline.

Evaluation	Weight
Project	%20