East West University
Department of Computer Science and Engineering

CSE 477: Data Mining (Project Description)
Course Instructor: Dr. Mohammad Rezwanul Huq

CSE 477 Project

Comparative analysis of different classification algorithms

In this project, students must understand and implement different classification algorithms on a selected dataset. The classification algorithms are as follows:

1. Decision Tree Induction
2. Naïve Bayes Classifier
3. Random Forest
4. Support Vector Machine
5. Recurrent Neural Network or LSTM Network (optional)

Each of these algorithms must be implemented both in Python (using appropriate libraries if needed) and in Weka tool. Students must present a comparative analysis of these algorithms followed by a discussion section that is justifying their performance. A convincing discussion section is very important to achieve higher grades.

Group Formation and Dataset Selection
Students must form a group of three students. Each group will choose a dataset from the following sources:

- Kaggle: https://www.kaggle.com/
- UCI ML Repository: https://archive.ics.uci.edu/ml/index.php
- Skymind: https://skymind.ai/
- OpenML: https://www.openml.org/

After forming the group and choosing the dataset, students must submit it in written to the course instructor by 05 December 2018.

Project Deliverables
The deliverables of this project are as follows:

- Python Codes or Python Notebook files (must be in a zip file, upload link will be given)
- Final report (report template is enclosed herewith, must submit harcopy)

Project Evaluation

The project has 15% weightage (15 out of 100) to the final evaluation. The following parameters will be considered while evaluating the project.

i. Understanding the algorithms (2 marks)
ii. Implementing the algorithms correctly including data pre-processing phase (2 marks)
iii. Comparative analysis (3 marks)
iv. Discussion of the results achieved (5 marks)
v. VIVA (VIVA will take place on 17 December from 10:00 AM) (3 marks)

CSE 477 (2) Project

Comparative analysis of different classification algorithms

Submitted by:

Your Name (Your student ID)

1. Introduction

Write opening statements about the project focusing on what you want to do and what you have achieved.

2. Classification Algorithms

Briefly explain each classification algorithm including their general principle.

3. Dataset

Write about your dataset. How many rows are there? How many columns? Have you taken any pre-processing step or not?

4. Implementation

Write the configuration of your machine you have used to carry out the experiments.
DO NOT INCLUDE ANY PYTHON CODE HERE.
INCLUDE SCREENSHOTS OF EXECUTING ALGORITHMS FROM WEKA TOOL.

5. Performance Evaluation

Present your comparative study in terms of accuracy, TPR, FPR and other measures. Use appropriate graphs, charts and so on.

6. Discussion

Justify the reasons of the performance of these different algorithms. Suppose, why random forest is better than SVM? This section has to be comprehensive and most important section of your report.

7. Conclusion

Include concluding remarks here.