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. Naïve Bayes Classifier
2. Decision Tree Classifier
3. k Nearest-Neighbour Classifier
4. Random Forest or Support Vector Machine* (*optional for extra points)

Naïve Bayes Classifier must be implemented from scratch, i.e., students are not allowed to use any tool or Python library functions.

For implementing other algorithms, students must use appropriate Python libraries. For obtaining extra points, students may also implement them in WEKA or other Data Mining tools and compare the results with Python implementation.

Students must submit their code and a report presenting 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 will do this project individually but present their findings in a pair. Details will be discussed in the class.

Each pair must choose a dataset from the following link:
https://www.mldata.io/datasets/

After choosing the dataset, students must submit it in written to the course instructor by 18 July 2019.

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 hardcopy)

Project Evaluation

The project has 10% weightage (10 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 (2 marks)
  iv. Discussion of the results achieved (4 marks)
  v.   Bonus Point (up to 3 marks)

Pair Formation and Dataset Selection (updated on 01:30PM, 17 July 2019

Pair No.	Student 1		Student 2		Dataset
1	2014-1-60-032	Md. Sakibur Rahman	2016-1-60-023	Al - Amin Islam Hridoy	
2	2014-2-60-087	Riajul Islam	2016-1-60-030	Md. Rifayet Azam Talukder	Bank Loan (Kaggle)
3	2014-2-60-098	Khadija Akter Trisha	2016-1-60-033	Sayma Obaida	Abalone
4	2015-1-60-029	Jahidul Hoque	2016-1-60-038	Taslima Yeasmin Emu	Black Friday (Kaggle)
5	2015-1-60-050	Sanjida Rahman	2016-1-60-042	Partha Dip Sarkar	Cars
6	2015-1-60-128	D. M. Kamruzzaman	2016-1-60-053	Shaykh Siddique	Forbes Billionaire
7	2015-1-60-148	Ramisa Hassan	2016-1-60-068	Humayra Khatun	Smart Phone
8	2015-1-60-155	Md. Mahmudul Hasan	2016-1-60-069	Syed Sahariar Hassan	Tic Tac Toe
9	2015-1-63-001	Ahamed Khan Zunayed	2016-1-60-079	Muhtasim - Ul - Alam	
10	2015-2-60-005	Golam Mostafa	2016-1-60-104	Farzana Rahman	School Grades
11	2015-2-60-044	Farhad Rahman	2016-1-60-124	Monica Yasmin	Heart Disease
12	2015-2-60-050	B. M. Saimur Rashid	2016-1-60-129	Fahim Abdul Kaium	
13	2015-2-60-080	Anamika Das Mou	2016-1-60-132	Asif Mahmud	
14	2015-2-60-082	Md. Mobinul Hoque	2016-1-60-134	Minhazul Hayat Khan	Soccer
15	2015-3-60-048	Fatema Nihar	2016-1-60-140	Nazmun Nahar Khanom	Breast Cancer
16	2016-1-60-005	Afsana Hossain	2016-2-60-015	Md. Kamrul Islam	Mushroom
17	2016-1-60-007	Shakil Mahmud	2016-2-60-047	Tasnim Ikra	
18	2016-1-60-011	Abeda Sultana Khanam	2016-2-60-081	Nishat Tasnim Mim	Chess King Rook

| 19 | 2016-1-60-012 | Md. Amit Khan | 2016-3-60-006 | Rakibul Alam | PIMA Native |
| 20 | 2016-1-60-017 | Jenifa Mousumi | 2016-3-60-038 | Muhammad Abu Bakar Siddik | Bank Marketing |

CSE 477(1) 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 tables, 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.