Title of Your Course Project

Name - Group members

COMPSC 445: Machine Learning in Applied Data Science.

Abstract

Briefly write a summary of your work. You have to write it in three parts in a single paragraph. First, discuss the background of the problem. Second, what you have done in the project; third, briefly write the results/outcome of your project. (**10 percents**)

1 Introduction

The introduction will be a significant part of your project report. It will have three paragraphs. In the first paragraph, you have to state what problem you are solving. Provide a basic background of the problem. Why this problem is important to solve, and what impact it might have. (**5 percents**)

In the second paragraph, you provide a literature review of the work. What are similar works? Here you have to comment on their effectiveness in solving the same problem. Try to tell about the strengths and limitations of their work. You have to find out at least three relevant work of your problem. And must cite them. Example of citation [1] [2] [3] [4]. (**5 percents**)

In the third paragraph, you establish why your work is important. Referring to the second paragraph where you have discussed the literature of the work, here you have to justify the need for your own work. (**5 percents**)

1.1 Our Contribution

In this section you use bullet points or enumerations to say what are the basic contributions of your project.

You may also write down the contribution of each team members.

(**5 percents**)

2 Material and Methods

First, provide a block diagram of your work showing a graphical description of your work. You have to describe the figure here. (**5 percents**)

2.1 Dataset Description

Describe the datasets that you have used/collected. You must provide url / citations to the dataset if not collected by you. Use a table to show the properties or distributions of the dataset. If you have collected your own dataset, you must put it in a GitHub repository and put the link in this section. If required, put histograms, density distribution, or samples from your data here; also, state why you have chosen the data. Describe all the features.

What are the attributes/ features (explain them)? What is/ are the target variable(s). Using domain specific knowledge, explain why you chose the features. Describe the size of your dataset, its source and age.(**10 percents**)

2.2 Algorithms Used or Methodology

Describe the details of the models or architectures that you have used. You have to use multiple models. For each model, you may draw the separate diagrams (depending on your project) and describe them in the text. Why have you chosen them? What were the parameters and hyperparameters that you have used? Why have you used them?

Also provide the Github link, colab link, or provide codes (as a attached file) for your project(**10 percents**)

2.3 Performance Evaluation

Here, you have to state the function or metrics [2] [5] [6] you have used and how you divided the dataset for training, validation, and test purpose.(**5 percents**)

3 Experimental Analysis

Here provide tables and graphs and describe them. There must be results with multiple values or a set of values of the hyper-parameters for each model. You need to compare the results in this section. Also, show convergence graphs (if possible) for the train, validation/test sets. The link to the code in colab / GitHub must be here.(**15 points**)

4 Conclusion

In this section you must again summarize your work. Tell about the limitations and possible future work. (**5 points**)

5 Presentation

You need to present the project/research during the specified time. I will talk about that during the class (**20 percents**)

6 For Master's/ Ph.D student

Master's or Ph.D students can submit a report similar to the IEEE/ACM paper format or identical to the other journal's layout.

References

[1] Peter Harrington. Machine learning in action (vol. 5). *Greenwich, CT: Manning*, 2012.

[2] Md Faisal Kabir and Simone A Ludwig. Enhancing the performance of classification using super learning. *Data-Enabled Discovery and Applications*, 3(1):5, 2019.

[3] Ashraful Haque, Jessica Engel, Sarah A Teichmann, and Tapio Lönnberg. A practical guide to single-cell rna-sequencing for biomedical research and clinical applications. *Genome medicine*, 9(1):1–12, 2017.

[4] Malte D Luecken and Fabian J Theis. Current best practices in single-cell rna-seq analysis: a tutorial. *Molecular systems biology*, 15(6):e8746, 2019.

[5] Md Faisal Kabir and Simone A Ludwig. Classification models and survival analysis for prostate cancer using rna sequencing and clinical data. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2736–2745. IEEE, 2019.

[6] Md Faisal Kabir, Tianjie Chen, and Simone A Ludwig. A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction. *Healthcare Analytics*, 3:100125, 2023.