Term Project Progress, Presentation, and Report

Term Project Final Report

Your final report has to look **professional**, with the following format:

- Your report should be **4-8 pages** (including tables and graphs)
- Use font size 11
- Use either Calibri or Times New Roman font
- Use double-space

Your final report should start with a **one-page executive summary** (See below for more information about a good executive summary).

> "An executive summary (or management summary) is a short document or section of a document produced for business purposes. It summarizes a longer report or proposal or a group of related reports in such a way that readers can rapidly become acquainted with a large body of material without having to read it all. It usually contains a brief statement of the problem covered in the main report, background information, concise analysis and main conclusions.[1]"

> The structure of the executive summary:

- One page long
- Consist of short, concise paragraphs
- Begin with a summary
- Be written in the same order as the main report
- Only include material present in the main report
- Be readable separately from the main report

Then the **main body** of your report starts with an **introduction** about your problem. Here you should provide a brief background on the problem and then clearly define your problem.

Following the introduction, you should introduce your **data**, how you acquired it, and its characteristics. You should clearly define your target variable, and also reporting your features, and the number of records in your dataset

---

[1] https://en.wikipedia.org/wiki/Executive_summary

Then you need to conduct the required **data cleaning/pre-processing** steps

- Imputing potential missing values (and reporting it)

- Encoding potential categorical variable (and reporting it)

Next, you would conduct a **descriptive analysis** on your data (using describe() function) and report it

In the next section of your report, you would conduct the main analysis, which is **developing machine learning models** (or predictive models). This section should contain the following steps:

- Specifying the problem type (classification/regression)

- Using all algorithms we learned in class

  - **Classification algorithms:** k-NN, logistic regression, support vector machine, decision tree, random forest

  - **Regression algorithms:** k-NN, linear regression, lasso, ridge, decision tree, random forest

  - For the decision tree model, provide a screenshot of your visualized tree. Make sure to prune your tree so you do not have a big tree.

  - For each algorithm, using the *validation curve* function, you should try various values for parameters and see what parameter value for your data/problem is the best

- After developing all models, you need to compare their performance on the test set, and specify which one performs the best for your data

- Finally, you need to predict the target value for a couple of test data points (it is best if you could have real test data points)

- All of these have to be documented in neat tables (not just screenshots)

In the end, in the **discussion and conclusion** section, you should provide the potential insights that you got by doing descriptive and predictive analytics and provide your recommendations. Here you should also provide **variable importance** using decision tree and random forest models.

Make sure to **submit your Python code and dataset as well**. Put everything in a folder, then compress (zip) the folder, and upload only the compressed folder, including your report, code, and **dataset.**

I should be able to run your code using your data with absolutely no errors.

Presentation

You need to present the essential parts of your project during week 10.

Each group will have about 10-12 minutes to present their work. **All team members have to participate in the presentation**.

In your presentation:

- You do **not** need to talk about the details of algorithms, like what a decision tree is, etc. We believe that everyone in the class knows those details ☺
- You do **not** need to display your code in the presentation; you will submit that with your final report
- Include an introduction about your project/problem
- Tell what your target variable is
- Tell us about your features
- Include the specifications of your data, how many records, how many features
- Include data cleaning steps (if you had missing values, categorical variables)
- Include the results of your predictive modeling (accuracies of your models in different algorithms …)
- Include feature importance
- Include prediction for the target value using a couple of test data points (it is best if you could have real test data points)
- Include recommendation/conclusion

You need to have about 10-12 slides.

Your slides have to look professional.