# Project Title (fill in)

Data Set(s) (edit to specify which): Credit Card Default; Mushroom; NBA Players

Author(s), email contacts

April 25, 2023

**Please organize your report along the lines of this template**; you may use any word processing software you like, as long as you submit your report as the required pdf file described below.

**Your report must be typewritten and submitted as a pdf document**, in machine readable form (no scans or screen shots).

**The report should not exceed 15 pages.** Please note that this is **not** the recommended length **nor is it a hard limit**. You should favor readability (proper font and image sizes) over possible page issues. However, be aware that part of your grade depends on the written quality of your report, and this includes both clarity and conciseness.

**You must properly cite your references.** If you are following ideas from papers or from online discussion forums (this includes discussion forums such as Kaggle), be sure to clearly distinguish between what is someone else's work and what is yours and to cite the source. Wherever you include material that came from elsewhere, include a citation at that location (e.g., [1], or [author's last name(s)]). This includes figures that were taken from elsewhere; include the citation (e.g. [1]) in the figure caption. This also applies to figures taken from EE 559 lecture or discussion.

At the end of the report, include the full reference. We suggest the IEEE format (easily found in Word or LaTeX) like this [1] or this [2]. If it's from an internet resource, then include a link. Failure to properly acknowledge sources of what you report amounts to plagiarism and can result in substantial penalties.

## 1   Abstract

A brief (typically 200 words), informative description of your project. Include the problem, dataset(s) you used, approach (naming the machine learning methods you used and how you compared them), and key results (be sure to include at least your best result and the method with which you obtained it, for each dataset).

The abstract should be considered a "stand alone" section – it should be understandable on its own and include only information that is described (and supported) elsewhere in the report. The report should also be able to "stand alone" without the abstract.

**Tip**: many people find it works better to write the abstract last, even though it will be read first.

## 2   Introduction

### 2.1   Problem Assessment and Goals

A brief description of the data set(s) you chose, and the goals you want to achieve.

### 2.2   Literature Review (Optional)

If you have investigated previous or existing approaches to your problem, briefly describe them here.

# 3 Approach and Implementation

Report your approach and implementation details in the following subsections. You should mention which libraries you used, and which functions you coded yourself in section 5, but avoid including code anywhere else in your report. Your description of what your system does should be readable and understandable to a reader that isn't familiar with the functions and libraries you used but is familiar with the algorithms and techniques that were covered in EE 559. (For example, stating "we standardized all real-valued features", and also stating the functions used in your code for this, is fine; stating only the functions used in your code is not fine.)

**If your project includes both regression and classification:**

- If your approaches for regression and classification are similar, we suggest that you follow the outline below, describing your work on both regression and classification in each subsection. This way, elements in common can be described only once; then you can follow up to describe what is different for each problem (classification and regression) separately. However, if your approaches don't have much in common, or if you think it would make more sense for your project to follow this sequence of subsections separately for each problem (e.g., 3.1-3.5 for classification, then 3.6-3.10 for regression; or create subsubsections such as 3.1.1: Preprocessing for Classification 3.1.2 Preprocessing for Regression), you may do that instead.

- It is helpful to provide a table describing which classifiers or regressors, and which preprocessing methods, were used for each problem.

## 3.1 Dataset Usage

Describe the procedure you followed in the use of your dataset. You should clearly state how many data points were used for training, validation set(s), and testing. For cross validation, also specify the number of folds and number of runs.

Describe how validation sets were used, and where in the process cross validation was used. If cross validation was used multiple times, specify how they were arranged (sequential loops or nested loops), and where decisions were made based on validation-set results.

The description must include the dataset usage in all stages of your work, from feature preprocessing, feature engineering, and dimensionality adjustment, to training and model selection including training and validation steps. Feel free to create subsections for each part if it seems appropriate to do so.

Also describe where in the process the test set was used, and how many times the test set was used.

## 3.2 Preprocessing

Describe in detail any preprocessing techniques you used. This could include for example, data normalization, and re-casting different variable types. Also justify or explain your choice of methods.

If you used different preprocessing for different classifiers or regressors, or if you tested the same classifier or regressor with different preprocessed inputs, state so. A table or plot can be useful in these cases.

## 3.3 Feature engineering (if applicable)

If you developed any new features, describe them here. If you developed a set of them but then refined the set afterwards, describe your methods, state any intermediate results that helped to choose among the features, and give the final set that was chosen.

If you did not develop any new features, state "Not applicable" for this section.

If your refinement of features you developed was combined with your feature dimensionality adjustment, you may combine both Sections 3.3 and 3.4 into one section, entitled "Feature engineering and dimensionality adjustment".

## 3.4 Feature dimensionality adjustment (if applicable)

Explain your approach and implementation. If you optimized over some parameter(s), or compared different methods, show and describe your work and outcomes.

If you did not do any dimensionality adjustment, provide a brief justification of why there was no need for it.

## 3.5 Training, Classification or Regression, and Model Selection

Describe how you trained your model(s), the classifiers or regressors you used, how and where you did model selection, and the parameters you chose. You may find a block diagram or flow chart is useful in describing some of this.

For each classifier/regressor or learning algorithm you used (below referred to as "ML method"), include the following. (You can optionally use a separate subsection (3.5.1, 3.5.2, etc.) for each ML method you used.)

- **Describe your model and your algorithm**. It generally isn't necessary to repeat equations given in EE 559 just to describe the model and algorithm; however, you must give enough information to clearly define which model and algorithm (and which version of the model and algorithm) you are using. Also, if you want to refer to any equations (e.g., for your interpretation or analysis), you must include those equations in your report. If applicable, also explain what you did to adapt the ML method to your case.

- If you used any **models or algorithms that are outside the scope of EE 559**, then give a more complete description of what they are. If the description would be long, you can give a summary of it in your report and cite reference(s) that give more complete descriptions.

- **State the parameters of the model and how they were chosen.** If a parameter is chosen by heuristics, state so. If a parameter is chosen by some model selection, optimization, or validation process, state so and describe the method.

- **Consider and describe degrees of freedom with the number of constraints** or data points you have.

- If you have **sets of results to show for this ML method**, include them here. (For a comparison of results from different ML methods, use the next subsection.)

- **Include any analysis and interpretation** of these (intermediate) results here.

  Notes on presentation:

- **You do not have to present in-detail intermediary results for each method.** For example, no need to show cross-validation performance for each value of each tested parameter. However, be sure to state the range of values that was tested, the increment or number of values tried, along with the final parameter choice.

- **Do not copy and paste print-screen (or screenshots)** of results. Include numerical results in tables or as part of the text; save images as files and import or insert them into the report.

# 4 Analysis: Comparison of Results, Interpretation

Present performance comparison of all your different models and methods here. Include a comparison with the required trivial and baseline systems, and, optionally, with any results you found in the literature or internet. For each result, be sure to clearly state whether it is from training, validation, cross-validation, or test set. Use table(s) and/or plots. Once again, **do not paste screen shots or print-screen images**. If you worked on more than one problem or solved multiple missions,

you can either have different subsections for different problem/mission, or you can directly compare how each classifier performed on each problem or mission in the same section.

Include all the required performance measures for your trivial system, baseline system, and final chosen system, on a validation set (with or without cross validation).

Finally, report all the required performance measures for your final chosen system, and for the required trivial and baseline systems, on the test set. If you also evaluated your final system on other performance measures that aren't required, also report those results.

Clearly state what your best performing system was, including normalization, preprocessing, feature set, classifier/regressor used and values of parameters; enough to uniquely specify your final system.

Include your analysis and interpretation of your results. Compare the d.o.f. and constraints from model to model. Can you explain what you observed? If not, any conjectures? Did you observe anything particularly unexpected? Note that your analysis shows your understanding of what you have done and is an important part of your project grade. Analysis of intermediate results can be stated along with the reporting of those results, in the section that reports those results.

# 5   Libraries used and what you coded yourself

State what libraries you used. You do not need to list classes and functions individually. Also state what you coded yourself. For example, sklearn provides multiple options for running cross-validation. If, instead, you coded the CV loop yourself, this might count as extra workload, so please mention it here. Similarly, include whether you used pandas for preprocessing or other functions outside of basic csv file operations. You do not need to provide implementation details.

Also, if you used any libraries or functions/classes that aren't on the allowed list for EE 559 methods [see "Computer languages and available code" in the Project Assignment] (for example, libraries or functions/classes for which you were given special approval, or that were necessary or appropriate for non-EE 559 methods in your project), state them here, and include the following.

- A clear description of everything (the models you used, the methods and approach, parameters, etc.) in your report; do not assume that we will look through your code to understand what you did. If your approach is too complicated or lengthy to adequately describe in the report, cite a reference or link that gives a good description, that we can easily access.

- A readme.txt file that describes what additional libraries you used (and links to download them), and what is needed to run your code. "Additional libraries" means libraries or functions that aren't allowed for the EE 559 methods in your project, but you used for the non-EE 559 parts of your project. (See Appendix A, item 2 for a complete description.)

- Clearly state with your results which results are from systems that followed EE 559 topics and requirements, and which used other methods and/or libraries. This is so we can compare performance or systems fairly among different students.

# 6   Contributions of each team member

If a team project, state here what the contribution of each team member was (i.e., who did what).

# 7   Summary and conclusions

Briefly summarize your approach and key results, and optionally state what would be interesting or useful to do as follow-on work. Optionally, summarize some of the key things you learned while doing the project.

# References

[1] *Bibliography management with bibtex*, available at `https://www.overleaf.com/learn/latex/bibliography_management_with_bibtex`

[2] Vitor Cerqueira, et al., *Combining Boosted Trees with Metafeature*, in Advances in Intelligent Data Analysis XV: 15th International Symposium, Stockholm, 2016.

# A   Code and report file instructions

This section provides information to you; it is not part of your template

1. **Similar to your project report, your pdf code file must cite sources of any code that you did not create yourself.**

   - This can be done informally, by stating in a comment where it came from (e.g., "thanks to Kaggle kernels for the following code for function norm() [www.kaggle.com/.../...]"). Failure to do this could result in plagiarism penalties.

2. **Upload two files containing your code, plus possibly one readme file**, to D2L:

   - A single pdf file (computer readable, with no screenshots, scans, or print-screen images) that includes all your code (similar to homework code files)

   - Also a single zip file that contains all your code (.py or .ipynb) files, so that we can run your code if necessary

   - If you used any non-standard libraries or functions (non-standard means weren't included on the allowed list of libraries and functions for using EE 559 methods in your project), include a requirements.txt, requirements.json, and/or readme.txt with sufficient information so that we can perform needed installations and run your code.

3. Also submit **a single pdf file of your report**, including all figures, tables, references, etc. Must be computer readable and have no screenshots, scans, or print-screen images: upload to D2L

   - Check report submission folder instructions. Team projects will submit to the team projects folder.