

## Project Details

Total Points: 100

The project constitutes an important part of this class. Students will be expected to work in groups of up to four members for the project, throughout the semester. You are expected to work on a machine learning project of considerable size. The project should aim to answer three (3) research questions for which you'll have to design experiments (which models to use, how to evaluate, etc.).

What can count as a research question? For example, using the same insurance dataset:

- Predict whether the person is a low-, moderate- or high-risk individual (classification)
- Given some data about a person and its risk classification, predict how much that person would cost to the insurance company per year to be able to determine how much of a premium should be charged (regression)
- Are there clusters of claims that might indicate fraud? (clustering)

What is NOT a research question?

- Performing data preprocessing
- Deciding which neural network architecture to use
- Tuning a model's hyperparameters
- Doing anything that falls outside the scope of machine learning. For example, building a website to showcase your project. If you are unsure, ask an Instructor.

We are not expecting you to achieve state-of-the-art performance or to create Artificial General Intelligence (AGI) in your work. However, it is required to demonstrate that if an approach did not work originally, that you were able to diagnose it and solve the problem or you tried other approaches that may or may have not worked.

The final group members and the topic should be submitted on **Monday, March 8th**. You are strongly encouraged to choose a topic close to your area of interest and something you are passionate about. No project proposal should be based on re-using data from a Kaggle competition unless previously approved by an Instructor. The project proposal and the progress report will be graded. We expect meaningful progress to be made by the time that the progress report is due.

There will be final presentations that will occur either during the final two weeks of classes, the last week of classes, or during final exams week, depending on the number of groups. Students will be graded on the content, work, presentation, report, and teamwork. Specific due dates and milestones are shown below.

- **Monday, March 8th:** Declare team and general topic.
  - Write a Piazza post to the Instructors and all team members (up to four members total). Select the 'projects' folder for the post.
  - Make a general statement about your team and what topic you plan to pursue. Provide any and all details that you have about the general project.
- **Monday, March 15th:** Submit project proposal via Canvas
  - We will provide a template
  - Details for the proposal requirements are below.

- **Week of March 15th:** meet with Instructor for proposal feedback
- Required minor modifications or rejected proposals will have to be re-submitted within one week
- **Monday, April 5th:** submit a progress report (and code) via Github
  - Provide a 1-2 page summary of your current progress
  - The summary should include detailed information of what you have done (e.g. data collection, programming, evaluation, ...), including information about what each team member is working on.
  - The summary should include information about upcoming milestones, along with any challenges that may be present
- **Tuesday, May 4th, 11:59 PM:** final submission (report + code) via Github
  - Submit a 5-page conference style report (templates will be provided) following the guidelines below. We recommend that you write a draft of your report as progress is made. Often, students write their reports in the last minute and submit documents that are full of errors. This is a graduate course, so we at least expect that you'll use a spellchecker before submitting.
  - Submit code and report via Github
- **Last week of class (4/27 and 4/29):** project presentations.
  - 7 minute presentation for each group (length depends on number of groups).
  - follow same guidelines as for the report. Note that all group members must present.

The breakdown of points for the assignment are as defined below:

- Proposal: 10%
- Progress report: 15%
- Code: 25%
- Report: 25%
- Presentation: 25%

Documents will either be submitted through GitHub or Canvas. **For Github submissions (more info below), please create a folder called 'project' (without quotes) and commit all files for the project to this folder.** You are strongly encouraged to submit early and often to avoid any last minute issues and to confirm that you are committing your work correctly. The submission should be in the form of a Python 3 Jupyter Notebook file, along with any special instructions for running the program. **Be sure to run the file before committing, so that we can directly see your results.** Programs that fail to run will not be graded and will result in a zero. All other files will be submitted through Canvas.

**Late submission policy: all files related to the final submission and presentation must be submitted on time. Late submissions will not be graded for these two components of the project.** For other milestones (e.g. proposal, progress report, group/topic), we expect all work to be finished and submitted on time as well, however, we do understand that there might be times when something unexpected comes up which delays you. Hence, as a late submission policy, we will allow late submissions up to 3 days late, each day carrying a 5 point penalty. If you submit 70 hours after the original deadline, the highest possible score for that assignment will be 85 points, out of 100. If the assignment is submit 1 minute after the deadline, then a 5 point penalty will be assessed. Late assignments submitted at 72:01 hours after the original deadline will

not be accepted. Please, make the arrangements to start and finish your assignments early. Trying to do the bulk of the project during the last few weeks is the surest way to fail the course.

### **Milestone 1.** [10 POINTS]

A project proposal needs to be submitted prior to starting the project, and the proposal is due via Canvas on the date provided above. Your proposal should be 1-2 pages long and should include:

- Your background: technical area, relevant coursework, programming skills, and pertinent research you have done, if any.
- Proposed project
  - Project description. What problem are you trying to solve?
  - Motivation for the project. Why is this project/problem important?
  - Technical description of your proposed project. What are you going to do? And How?
  - Description of the data that you will use, and how you will access it.
  - List of a few references that you have read or scanned on the subject.
- Potential challenges/hurdles: List any potential challenges that you feel you may encounter. This may include, things such as:
  - Lacking in data
  - Unsure how to implement a machine learning algorithm
  - Team is lacking in capabilities (e.g. do not understand Python, so cannot use Tensorflow)
  - Unsure about which machine learning algorithm to try
  - Unsure about how to pre-process the data

### **Milestone 2.** [15 POINTS]

Provide a 1-2 page summary of your current progress, submitted through Canvas.

- The summary should include detailed information of what you have done (e.g. data collection, programming, evaluation, ...), including information about what each team member is working on.
- The summary should include information about upcoming milestones, along with any challenges that may be present

Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

### **Milestone 3.** [25 POINTS]

You are to write a five-page report (excluding references) that provides specific project details, and this report should be submitted through Canvas. The report should include:

- an introduction, which summarizes the problem that your project strives to address and provides any motivations as to why this problem is important.
- The report should then discuss any relevant related work that has been done in the field, and discuss how your approach addresses components of the problem that are not addressed by related approaches.

- The report should then describe your approach to solving the problem. This is where you discuss the machine learning algorithm in detail.
- Next, experiments and experimental setup should be described. This is where you discuss your dataset, parameter settings, cross validation setup, evaluation measures, etc.
- The report should then summarize your results. There should be tables/figures to highlight your results, along with a detail discussion of the results.
- Next you summarize your approach and any future directions that you think are worth pursuing.
- Lastly, the report should contain a paragraph about what you liked about the project, learned from the project, any challenges that were encountered, differences between the final approach and the proposed, and any questions that you still may have.
- All references should then be placed at the end of the paper.

You MUST use the provided conference-style templates to produce your report, where Microsoft Word and LaTeX templates can be found on Canvas. DO NOT modify the template in any way. Your report must be submitted as a single pdf document on Canvas. This assignment must be submitted on time to receive credit. No late work will be accepted, unless you have a prior arrangement with the instructor.

Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

#### Milestone 4. [25 POINTS]

Submit all code via Github. Please create a folder called ‘project’ (without quotes) and commit all files for project to this folder. You are strongly encouraged to submit early and often to avoid any last minute issues and to confirm that you are committing your work correctly. The submission should be in the form of a Python 3 Jupyter Notebook file, along with any special instructions for running the program. **Be sure to run the file before committing, so that we can directly see your results.** Programs that fail to run will not be graded and will result in a zero.

#### Milestone 5. [25 POINTS]

The 7 minute presentation should generally be a concise and succinct version of your report. Everyone in the group is required to speak during a portion of the presentation. The presentation should include:

- a brief introduction that summarizes the problem that your project strives to address and provides any motivations as to why this problem is important.
- briefly discuss any relevant related work that has been done in the field, and discuss how your approach addresses components of the problem that are not addressed by related approaches.
- describe your approach to solving the problem. This is where you discuss the machine learning algorithm in detail.
- describe the experiments and experimental setup. This is where you discuss your dataset, parameter settings, cross validation setup, evaluation measures, etc.

- summarize your results. There should be tables/figures to highlight your results, along with a detail discussion of the results.
- Lastly, the presentation should discuss what you liked about the project, learned from the project, any challenges that were encountered, differences between the final approach and the proposed, and any questions that you still may have.

You can create your presentation using any platform that you like (e.g. PowerPoint, PDF, Keynote, Google).

Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.