**ST311 Summative Project**

**Content**

[**Instructions: overview and key dates**](#instructions)

[**Documents to submit**](#docs2submit)

[**Project ideas**](#ideas)

[**GenAI policy**](#genai)

[**Marking criteria**](#marking)

[**Individual marks**](#individualmark)

[**Appendix: some project titles**](#appendix)

**Instructions: overview and key dates**

Working on a deep learning research project as a team, you will explore an idea related to the material in the course.

**Work in teams of 2 to 4.** You are responsible for forming a team. Submit the five-digit candidate numbers of your team member(s) through this link:

<https://docs.google.com/spreadsheets/d/1cAFieMQbNuc141DwrAAFh-XUvJsb4044RniZ_-3aVfk/edit?usp=sharing>

**Key dates:**

17 February noon (Week 5): Group details in the spreadsheet.

24 March noon (Week 10): Submitted a proposal about the project. I encourage you to submit this before the deadline. Ideally you will start during WT. Maybe wait till Week 9 to see Transformer or start looking at [Hugging Face](https://huggingface.co/) site.

6 May noon (1st week of ST): Project deadline. Documents to submit:

* The paper
* Slides
* Code

All documents should be submitted via Moodle.

**Communication**

Use the forum for questions related to your project.

**Do not contact me by email regarding your project,** unless it does not contain details of your project I can identify you by.

**Documents to submit**

All submissions must be in PDF format, except for the code which should be in html, and **only by one member of the group**.

The name of your files should be in the format:

<Group number>-<name>.pdf

where name is one of : proposal, paper, slides, code.

Example: Group 1 would submit:

01-proposal.pdf,01-paper.pdf, 01-slides.pdf, 01-code.html

The code may in included as an appendix of the paper.

**Do not include any names** in documents you submit.

**Proposal**: The purpose of the proposal is for me to assess the feasibility of your project. Provide an abstract, introduction, methods and algorithms, references. There are no marks for the proposal. Keep it to no more than 2 pages. I will write back via Moodle.

**The paper**: Your paper should be in [[NeurIPS paper format](https://nips.cc/Conferences/2021/PaperInformation/StyleFiles)](https://neurips.cc/Conferences/2023/PaperInformation/StyleFiles#:~:text=All%20submissions%20must%20be%20in%20PDF%20format.%20Submissions,additional%20pages%20containing%20only%20the%20references%20are%20allowed.). The online editor [Overleaf](https://www.overleaf.com/latex/templates/neurips-2023/vstgtvjwgdng) makes this task simple. Assume the audience reading your paper has taken ST311.

The report should consist of a title, abstract, introduction, methodology, numerical evaluation using real data example, a conclusion section and a reference section.

The paper should be **no longer than 10 pages plus reference**.

**Slides:** **No more than 15 slides** presenting your work that could be delivered to a general audience who have taken an introductory course in statistics and maths. Your slides should not be too technical. They should inform and possibly entertain the audience.

**Code**: You must use PyTorch and/or Hugging Face only.

The code should run in Colab. Do not use any d2l module. Upload your source code to Colab and provide a link to it at the top of the notebook. **Ensure the link is not restricted access.**

**Project ideas**

The course projects should build on top of the course materials. Here are three ideas:

* Implement an existing method and re-evaluate the implementation against some standard benchmarks. An example of this kind of study is [here](https://arxiv.org/pdf/1506.02078.pdf).
* Adapt current algorithms for a new application.
* Compare the performance of some architectures on a task.

A list of projects ideas is in the appendix.

**GenAI policy**

We will follow the guideline to the [International Conference on Machine Learning 2023 policy](https://icml.cc/Conferences/2023/llm-policy) about the usage of Large Language Models (LLMs), such as ChatGPT. The LLM policy prohibits text produced entirely by LLMs (i.e., “generated”). This does not prohibit authors from using LLMs for editing or polishing author-written text.

Any work, code, or ideas that are not your own must be appropriately acknowledged with a citation.

**Marking criteria**

The project counts for **70% of the marks**. The mark is out of 100, and is split as follows:

* Paper: 70 marks
* Slides: 20 marks
* Code: 10 marks

The marks are distributed according to following criteria:

**Paper (70 marks)**

|  |  |  |
| --- | --- | --- |
| **Quality** | 30 marks | Is the report technically sound? Are claims well-supported by theoretical analysis or experimental results? Are the authors careful and honest about evaluating both the strengths and weaknesses of the work? To get full mark in this category, you will need to include at least one of:   * An algorithm box. * Equations describing your model. |
| **Clarity** | 20 marks | Is it in NeurIPS format?  Is the report clearly written?  Is it well-organized?  Does it adequately inform the reader? |
| **Originality** | 15 marks | What is original about your study? problem or approach new?  Is this a novel combination of familiar techniques? |
| **Significance** | 5 marks | Are the results important?  Does it advance the state of the art in a demonstrable way?  Does it provide unique data, unique conclusions on existing data, or a unique theoretical or pragmatic approach? |

**Slides (20 marks)**

|  |  |  |
| --- | --- | --- |
| **Content** | 12 marks | Inform the audience?  Interesting? |
| **Presentation** | 8 marks | Well organised?  Suitable level for the audience? |

**Code (10 marks)**

Is the code well annotated and structured?

**Individual marks**

You are expected to split the work on your project among yourself. At the end of the report, you must add a section ***Statement about individual contributions***, in which you need to provide the percentage of contribution of each group member and **summarise individual contributions.**

The grade for each group member will be a function of the contribution of each group member using the following equation.

member\_grade =

report\_grade × (member\_contribution/max\_contribution)

For example, for a group with 2 members contributing, 60%, 40% and the report grade is 75 (out of 100), the individual grades are

75 x 60/60 = 75, and 75 × 40/ 60 = 50.

**Appendix**

A diagram of a research paper

AI-generated content may be incorrect.

**Appendix: some project titles**

COVID-19 detection using deep learning techniques.

Option pricing using deep learning.

Detecting Cyberbullying using Neural Networks for Text Classification.

Sentiment classification of Covid-19 related posts on Twitter.

Classify similar products into groups using product images and titles. Large Movie Review Sentiment Analysis using various Deep Neural Networks.

LSTM with Emotional Analysis for Stock Price Prediction.

Image classification for diagnosing Covid-19 & Pneumonia by Chest X-ray.

Conversational ChatBot using deep learning.

Generation of financial time series using GANs.

Predicting poverty level from satellite imagery using deep neural networks.

predicting Meta stock prices using LSTM and DNN.

Using neural networks to predict the Myers-Briggs Personality Type (MBTI) of users from their social media posts.

Heart anomaly detection.

Multi-class classification of news category with BERT.

Music genre recognition using neural networks.

A step of a paper with text on it

AI-generated content may be incorrect.

<https://www.turing.com/kb/how-to-write-research-paper-in-machine-learning-area>