

Columbia FinTech Bootcamp - Project 2 Proposals

Please submit this proposal no later than Monday March 20th, 2023

Group assignment methodology:

Students are allowed to choose their group members. Group size should be no more than 4.

Topic:

So far, you've learned several advanced analytical techniques that use machine learning. Accordingly, over the next two weeks, you'll work with a group to solve a real-world financial problem by using machine learning.

Here are a few ideas for getting started:

- Compare two or more machine learning models for solving a predictive task.
- Use natural language processing (NLP) to draw insight from written or spoken language.
- Deploy a machine learning model in Amazon SageMaker as an API.
- Deploy a robo advisor that's powered by Amazon Lex.
- Use machine learning to build a sophisticated algorithmic trading bot.

And, here are some real projects from previous classes:

- Use a deep learning machine model to evaluate X-ray images and predict whether a patient has pneumonia. Make the predictions with a high degree of accuracy to increase diagnostic efficiency, reduce costs, and impact healthcare in countries or regions that have higher death rates.
- Analyze historical 13F filings, and build a model that's based on the performance of those funds. Then recommend an optimized combination of investment sector weightings to create a portfolio that should outperform the annual return of the S&P 500.

Technical Requirements:

The following subsections list the technical requirements for Project 2.

Software Version Control (10 points)

- Repository created on GitHub. (2 points)
- Files frequently committed to repository. (3 points)
- Commit messages with appropriate level of detail included. (2 points)
- Repository organized, and relevant information and project files included. (3 points)

Data Collection and Preparation (10 points)



- Data collected from CSV files, APIs, or databases by using Python or a Python library. (5 points)
- Data cleaned and prepared for the application or analysis by using Python or a Python library. (5 points)

Machine Learning (40 points)

- Jupyter notebook, Google Colab notebook, or Amazon SageMaker Studio notebook created to prepare training and testing datasets. (5 points)
- One or more machine learning models created. (5 points)
- Models fit to the training data. (5 points)
- Trained models evaluated by using the testing data. Calculations, metrics, or visualizations that are needed to evaluate the performance included. (10 points)
- Predictions shown by using a sample of new data. Predictions compared if more than one model was used. (3 points)
- PNG images of your visualizations saved to distribute to the class and instructional team and to include in your presentation and the README.md file for your repo. (2 points)
- One new machine learning library, machine learning model, or evaluation metric used that the class hasn't already covered. (10 points)

Documentation (15 points)

- Code is well commented with concise, relevant notes. (5 points)
- GitHub README.md file includes a concise project overview. (2 points)
- GitHub README.md file includes detailed usage and installation instructions. (3 points)
- GitHub README.md file includes either examples of the application or the results and summary of the analysis. (5 points)

Presentation (25 points)

Each project group will prepare a formal 10-minute presentation that includes the following:

- An executive summary of the project and project goals. (5 points)
 - Explain how this project relates to fintech and machine learning.
- The selected model. (5 points)
 - Describe the machine learning model that your group selected and why.
- The data preparation and model training process. (3 points)
 - Describe the source of your data and why you chose it for your project.
 - Describe the collection, cleanup, and preparation process.
 - Describe the training process.
- The approach that your group took to achieve the project goals. (5 points)
 - Include any relevant code or demonstrations of the machine learning model.
 - Describe the techniques that you used to evaluate the performance of the model.
 - Discuss any unanticipated insights or problems that arose and how you resolved them.



- The results and conclusions from the machine learning model or application. (5 points)
 - Include relevant images or examples to support your work.
 - If the project goal wasn't achieved, share the issues and what the group tried for resolving them.
- Next steps. (1 points)
 - Take a moment to discuss the potential next steps for the project.
 - Discuss any additional questions that you'd explore if you had more time. Specifically, if you had additional weeks to work on your project, what would you research next?
- MODIFIED Investment/Project Cost(s) (1 points):
 - Start thinking about the following - how much would it cost to take a project like this to production/to-scale? At minimum, what would it cost for you and your team to be on the project (minus nuances of overhead and other costs). Research your "role", what it is currently called in the field, and the associated salary. Add this up across your team members and, if this project (getting more data, standing up the appropriate infrastructures, and getting the desired results for the client) was to run for 6 months, what would it cost to hire you and your team?

nlsdhn@gmail.com [Ander account](#)



Concept opgeslagen

***Vereist**

E-mailadres *

nlsdhn@gmail.com

Enter team members first and last names.

Marc Pocorni, Andre Johnson, Jason Steiner, N



Enter Project Title *

App to estimate Diabetes Risk Prediction using

Enter Project description.

Python-based project using neural network techniques to predict the risk of diabetes. The project app provides the users an interactive interface built with streamlit that can both be used to fine tune the neural network model and allows users to input their information and receive a prediction on their risk of developing diabetes.

Enter Project objective. *

Project Objective is to develop a machine learning model that can accurately predict an individual's risk of developing diabetes.

Enter links or describe datasets to be used. *

<https://www.kaggle.com/datasets/mathchi/diabetes-data-set>



Enter rough breakdown of tasks assignments for each team member.

1. data collection and preprocessing --- Group
 2. neural network development and training --- Niels
 3. streamlit app development --- Niels
 4. model performance and model usage --- ???
 5. model selection and individual user input --- ???
 6. Prediction output creation --- ???
 7. App testing --- ???
 8. Readme creation --- ???
 9. presentation --- Group
-

Een kopie van je antwoorden wordt gemaild naar het adres dat je hebt opgegeven.

Pagina 1 van 1

Verzenden

[Formulier wissen](#)

Verzend nooit wachtwoorden via Google Formulieren.



reCAPTCHA
[PrivacyVoorwaarden](#)

Dit formulier is gemaakt in 2U. [Misbruik rapporteren](#)

Google Formulieren



