CS471: Introduction to Artificial Intelligence
Final Project

Project Overview

Goal of the final project is to apply all the concepts you learned in the class. This is a team project. You are required to form a team of 3 students.

Your first task is to pick a project topic. To help you, a list of potential topics was attached at the end of this document. You are encouraged to pick some other topic that is not listed in the document. If you're looking for project ideas, please come to office hours, and I will be happy to brainstorm and suggest some project ideas. Pick something that you get excited and passionate about!

Once you have identified a topic of interest, an important aspect of designing your project is to identify one or several datasets suitable for your topic of interest. If you intend to collect the needed data yourself, keep in mind that this can often take considerable time.

Once you have chosen the dataset, you need to apply at least three different algorithms of your choice. Compare and discuss the results from these different algorithms. First of these algorithms have to be a rule-based approach: rules you would apply to solve the problem. The other two have to be machine learning algorithms.

Project Proposal

Your proposal should be a PDF document, including the title of the project, the full names of your team members, and a brief description of what you plan to do.

Your project proposal should include the following information:

- Motivation: What problem are you tackling? Discuss one prior research that used the same or similar dataset and discuss what you are planning to achieve with your approach.
- Dataset: What is the source of your dataset? Briefly explain the contents of the data. Do you need any pre-processing? If so, what pre-processing techniques are you planning to apply?
- Method: What algorithms are you planning to apply and why?
- Intended experiments: What experiments are you planning to run?
- Model evaluation: How do you plan to evaluate your algorithms?

Presenting one example of prior research on the topic is required.

Progress Report

By the end of the progress report deadline, you are expected to finish the pre-processing of the data and the rule-based approach.

Your progress report document should include the following information:

- Motivation: Describe in detail the problem you are trying to solve
- Dataset: Describe the data you have chosen along with its features?
- Exploratory data analysis: Perform exploratory data analysis and provide detailed description of each step
- Pre-processing: Detail the pre-processing techniques you used
    - Data cleaning
    - Feature engineering
- Algorithm 1 (rule-based approach): Describe the rule-based approach and share code snippets of your work. Also discuss the pros and cons of the rule-based approach

Final Project Report

- Motivation: Provide the motivation for choosing this problem?
- Prior research: Discuss the work that has been done using this dataset
- Goals: Share your goals of the things you are trying to achieve with this dataset and how is it different from the prior work/research

- Dataset: Provide the source of your data and explain the contents of it.
- Exploratory data analysis: Share the plots you generated to explore the dataset
- Pre-processing: Detail the pre-processing techniques you used for the dataset. This includes data cleaning and feature engineering.
- Algorithm 1 (rule-based approach): Describe the rule-based approach and share code snippets of your work. Also discuss the pros and cons of the rule-based approach
- Algorithms 2 and 3 (machine learning approaches): Describe the machine learning algorithm you used and explain the reasons behind your choice.
- Experiments: Describe the experiments that you've run, the outcomes, and any error analysis that you've done. Try to use plots wherever possible.
- Model comparison: Describe the metrics you used for comparing the models. Try to use plots wherever possible.
- Result analysis: Discuss your observations from the results and your thoughts about future improvement.
- Provide details on each team member's contribution.

Use Google Colab for your code, plots, and comments. When you finish editing, re-run all the cells to make sure they work. Share the code snippets at the required places. Copy the link of the Colab notebook and paste it in your project report. You can upload the project report as a pdf file on Canvas.

Grading

- Project proposal submission - 10 points
- Progress report - 10 points
- Project report - 60 points
    - Motivation + prior research + goals - 10
    - Dataset + exploratory data analysis + preprocessing - 10
    - Algorithm 1 (rule-based approach) experiments + results - 5

    - Algorithm 2 and 3 (machine learning approaches) experiments + results - 20
    - Model comparison + result analysis - 10
    - Member contribution - 5
- Presentation and giving feedback for others presentation - 20 points

Presentation guidelines:
1: Problem statement
2: Data description; what are your inputs and outputs
3: Prior research and explain your goals
4: Exploratory data analysis
5: Data cleaning + feature engineering
6: Describe the three algorithms you have chosen and why
7: Fine-tuning of these algorithms
8: Comparison of these algorithms
9: Conclusion and future goals

Project topics

1. Loan Prediction
   The idea behind this ML project is to build a model that will classify how much loan the user can take.

   Dataset: Loan Prediction Dataset

2. Sonar Prediction

   The Sonar Dataset involves the prediction of whether or not an object is a mine or a rock given the strength of sonar returns at different angles.

   Dataset: UCI

3. Banknote Prediction

   The Banknote Dataset involves predicting whether a given banknote is authentic given a number of measures taken from a photograph.

   Dataset: UCI

4. Abalone Prediction

   The Abalone Dataset involves predicting the age of abalone given objective measures of individuals.

   Dataset: UCI

5. Ionosphere Prediction

   The Ionosphere Dataset requires the prediction of structure in the atmosphere given radar returns targeting free electrons in the ionosphere.

   Dataset: UCI

6. Wheat Seeds Dataset

   The Wheat Seeds Dataset involves the prediction of species given measurements of seeds from different varieties of wheat.

   Dataset: UCI

7. Personal Key Indicators of Heart Disease

   Dataset: Kaggle

8. Fraudulent Transactions Prediction

   Dataset: Kaggle

9. Stroke Prediction Dataset

   Dataset: Kaggle

10. Car Evaluation Data Set
    Dataset: UCI

11. Rice (Cammeo and Osmancik) Data Set
    Dataset: UCI

12. Income Prediction
    Dataset: UCI

Helpful links:

Kaggle datasets: https://www.kaggle.com/datasets

UCI machine learning repository datasets:
http://archive.ics.uci.edu/ml/index.php