

CSCI323 Project Topics

1. Modern Neural Network

- 1) **Spam email detection:** This project is to develop a model to conduct spam email detection. Students need to systematically investigate how good the model works. The following website is a good source for you to understand this topic. You can use the datasets therein or any other publicly available datasets.
<https://link.springer.com/article/10.1007/s10462-022-10195-4>
- 2) **Abalone age prediction:** This project is to develop a machine learning model to predict the age of abalone from physical measurements. The following website is a good source for you to understand this topic. Students should use the provided dataset and systematically investigate how good the developed model works.
<https://archive.ics.uci.edu/ml/datasets/Abalone>
- 3) **Facial expression recognition:** This project is to develop a machine learning model to identify and categorize emotional expressions depicted on a human face. The following website is a good source for you to understand this topic. Students can use the dataset provided on the website and shall systematically investigate how good the developed model works.
<https://paperswithcode.com/task/facial-expression-recognition>
- 4) **Wine quality modelling:** This project is to train a machine learning model to model wine quality based on physicochemical tests. The following website is a good source for you to understand this topic. Students should use the provided dataset and systematically investigate how good the developed model works.
<https://archive.ics.uci.edu/dataset/186/wine+quality>
- 5) **House price prediction:** This project is to train a machine learning model to predict the house price based on the information related to house. Students should find interesting house-related datasets and train a suitable model. The following website is a good source for you to understand this topic.
<https://arxiv.org/pdf/1901.01774>

Note that, for the above five topics, students may have two options:

- (1) Option 1: students only need to evaluate at least three baselines, which may be from the other existing work/solutions.
- (2) Option 2: students can choose one from the latest existing work/solution and implement another one themselves to beat the selected one.

2. Search Algorithms

- 1) **Travelling salesman problem:** This project is to implement at least two algorithms (which are not limited to the algorithms in the following website) to solve the travelling salesman problem and systematically investigate the advantages & disadvantages of each algorithm. The following website is a good starting point for you to understand this topic.
https://medium.com/@siddhantmittal_78686/travelingsalesman-a-survey-paper-c30f5edf1af8
- 2) **k shortest simple path problem:** This project is to implement at least two algorithms (which are not limited to the algorithms in the following article) to solve the k shortest simple path and analyse the advantages & disadvantages of each algorithm. The following website is a good starting point for you to understand this topic.
<https://hal.science/hal-02465317/document>

3. Markov Decision Process & Reinforcement Learning

- 1) **Implement AlphaGo game:** This project is to implement an AlphaGo game. The following websites are good starting points for you to understand this topic. Students shall actively find additional information and resource for them to effectively conduct this project.
<https://medium.com/betacom/the-game-of-go-and-alphago-6fc3d10b821>
<https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf>
- 2) **Drive a taxi:** This project is to design a stimulation of self-driving taxis. The goal of self-driving taxis is to pick-up passengers and drop them off at the destination in the least number of moves. The following website is a good starting point for you to understand this topic. Students shall actively find additional information and resource to effectively conduct this project.
https://www.gymnasium.dev/environments/toy_text/taxi/
- 3) **Autonomous maze solver:** This project is to design an autonomous maze solver. The aim is to train an RL agent to navigate through a maze and find the shortest path to the goal. The agent receives rewards for reaching the goal and penalties for hitting walls or taking too many steps. Students shall actively find additional information and resource to effectively conduct this project.
- 4) **Robotic Arm Control.** This project is to train an RL agent to control a robotic arm to perform tasks like picking and placing objects or drawing shapes. The agent learns to move the arm efficiently to achieve the goal. Students shall actively find additional information and resource to effectively conduct this project.

4. Constraint Satisfaction Problems

- 1) **Design a multi-city lighting project:** This multi-city lighting project involves several steps. Firstly, randomly place n cities, each represented by a point, on a 2D blank map. Then, maximally connect these cities with roads, represented by straight lines, ensuring the roads do not intersect. The objective is to illuminate all n cities using k different colours, with each city assigned a unique colour. However, cities connected by a road cannot share the same colour. This project requires you to develop a program capable of:
 - a. Generating and plotting a 2D map with the specified number of cities and roads, according to a user input of n .
 - b. Illuminating all cities (i.e., colour the n points) using only three colours: Red, Green, and Blue.
 - c. Illuminating all cities (i.e., colour the n points) using only four colours: Red, Green, Blue, and Yellow.
 - d. Employing three distinct algorithms to achieve the aforementioned tasks and comparing their performance relative to the value of n .
 - e. Plotting the 2D map the illuminated cities for the specific value of n and the number of available colours k .
- 2) **Automatic sudoku puzzle solver:** A Sudoku board consists of 81 squares, some of which are initially filled with digits from 1 to 9. The puzzle is to fill in all the remaining squares such that no digit appears twice in any row, column, or 3×3 box. This project is expected to display the solution of a sudoku puzzle when an image of this puzzle is provided. Your program needs to have the following modules at least:
 - a. Automatically recognise a specific sudoku puzzle from the given image and extract relevant information. (You need to use image recognition here)
 - b. Use an algorithm to automatically solve the recognized sudoku puzzle.
 - c. Display the solution to the recognized sudoku puzzle.
- 3) **Sentiment analysis with Naïve Bayes classifier:** Sentiment analysis is a technique to determine the sentiment or emotion expressed in a piece of text, for example, whether it is positive,

negative, or neutral. It is typically conducted by using machine learning and natural language processing techniques to automatically classify the sentiment of text into one of the categories. Sentiment analysis has been widely used to understand public opinion, customer feedback, and market trends to improve products, services, and customer experiences. This project is to implement a text classification system for sentiment analysis with the Naïve Bayes classifier. Your program needs to have the following modules at least:

- a. Organise training, validation and test data and load them into your system.
- b. Develop a Naïve Bayes classifier (including optimising the hyper-parameters when needed) with the training set.
- c. Conduct a systematic evaluation of the trained classifier on the test set.

On top of this, students are welcome to explore more advanced deep learning-based methods. The following website is a good starting point for you to understand this topic: <https://paperswithcode.com/task/sentiment-analysis>.

--- END ---