

Machine Learning Course - Final Project

Implement the following three problems. Last problem is bonus.

Check the requirements of each project and do the necessary implementation and deliverables. If you have any questions, don't hesitate to contact me in the Google classroom or contact the TAs.

Deliverables

- Submit the work as a colab link or a Jupyter notebook. Make sure the link is running properly. Include all the necessary external files such as external data or python files if needed. The structure is similar to what we did in assignment 1, assignment 2, and assignment 3
- Submit a one page pdf report explaining the steps you used for cleaning data if needed, the algorithms and techniques you used, and the links and references you used
- There is no one answer for this project. You will get your grade as long as you are doing correct implementation and are able to discuss your assumptions and results in a correct way.

Implementation requirements

- Provide the necessary cleaning for the raw data.
- Implement an evaluation for the hyperparameters whenever applicable
- Show how you test the model
- Use visualization techniques whenever applicable
- You can use any of the libraries that we used throughout the course such as Pandas, NumPy, etc.

Teams

- Every project will be composed of a team of 3 students. You can distribute the tasks among each other. For example, one student is responsible for data cleaning, one for building the model, and one for the visualization. Or one member handles one problem. However, in all cases, all the team must understand all the solutions and be able to answer questions related to their own part of the others.

Note

- The solution of the team should be totally theirs, can use code snippets with proper referencing but not a full solution. Full copies of the code will cancel the whole project for all the team.

Problem 1: Sales prediction**Introduction**

Retailers like Walmart, IKEA, Big Basket, Big Bazaar use sales forecasting for sale predictions of product requirements. Sales forecasting helps business owners get a clear idea of what products are in demand. Accurate sales forecasting will reduce wastage to a significant level and determine the incremental impact on future budgets. In this project, you are required to build a machine learning model for sales prediction.

Dataset

We will use the dummy Shampoo dataset which explained in the following link and can be downloaded from the same link

<https://www.kaggle.com/datasets/redwankarimsony/shampoo-saled-dataset>

Method

- Use simple linear regression to build your model.

Problem 2: Patient's Sickness Prediction System**Introduction**

Traditional healthcare systems became increasingly challenging to cater to the needs of millions of patients. But, with the advent of ML, the paradigm shifted towards value-based treatment.

Dataset

Use the cancer dataset we used in Assignment 2

Method

- Apply clustering, PCA, and t-SNE. Assume the last column to be a feature this time not a label
- Explain the patterns and comment on them

- You can use this link for code snippets if you would like to use Gaussian Mixtures. Or other code for using Kmeans

<https://jakevdp.github.io/PythonDataScienceHandbook/05.12-gaussian-mixtures.html>

Problem 3: Student University Recommendation (Bonus)

Introduction

Build a recommender system based on what is learned in the lecture which helps selecting a university since the criteria are usually confusing. Based on the student dataset and user profile, a list of 10 best universities will be suggested such that it maximizes the chances of a student getting admission into those universities.

Data

Download the data from this link, explanation is also there

Assumptions

You can simplify the data and use other techniques we will learned in the previous lectures to process the features and other parameters. But if you made any assumptions, make sure you list them in the report