WEEK 7 DAY 3: XP

**🌟 Exercise 1 : Defining the Problem and Data Collection for Loan Default Prediction**

**Instructions**

1. Write a clear problem statement for predicting loan defaults.

Aaccording to personal history of an applicant, will they fail to repay their loan or not: Binary classification.

1. Identify and list the types of data you would need for this project

Personal details of applicants, credit scores, loan amounts, repayment history,

Years of employment, age, years as a customer at bank, monthly income, other assets available.

1. Discuss the sources where you can collect this data.

Financial institution’s internal records, credit bureaus, directly from applicant, public records of real estate

**🌟 Exercise 2 : Feature Selection and Model Choice for Loan Default Prediction**

**Instructions**

* From a given dataset (assume columns like age, income, loan amount, repayment history, credit score, etc.), identify which features might be most relevant for predicting loan defaults. Justify your choice of features.

Repayment history as the dependant value to be used as the label for supervised machine learning.The most beneficial independent variables might be the income. Based on the income we can identify if the applicant has the necessary measures to pay back the loan. The credit score is a good indicator how the applicant has been performing in the past. However, if this is the first time that the applicant is taking out a loan, they will have a perfect credit score and therefore not be a good indicator for their future performance. The loan amount in combination with the income is probably also a good indicator for predicting defaulting.

**🌟 Exercise 3 : Training, Evaluating, and Optimizing the Model**

**Instructions**

* Outline the steps to evaluate the model’s performance, mentioning specific metrics (like accuracy, precision, recall) that would be relevant for this problem.

After training the model, test it with unseen data

Evaluate the performance of the model with a confusion matrix to identify the number of true positives, false positives, true negatives and false negatives. Of the number of false positives and negatves is too high, the model will need some more work. Depending on what is being measured, the tolerance for false positives or negatives will vary. In a medica study the number of false negatives should be very low as not identifying a tumor as not cancerous while it is in reality, can be deadly. The tolerance for false negative in loan defaulting is less crucual, therefore the tolerance can be higher for wrong predictions.

Calculate accuracy to evalute the number of accurate predictions

* Review the precision, recall, f1 score from the classification report for further insights on the model effectivness.

**🌟 Exercise 4 : Designing Machine Learning Solutions for Specific Problems**

**Instructions**

* For each of these scenario, decide which type of machine learning would be most suitable. Explain.
* Predicting Stock Prices : predict future prices:

Supervised machine learning – regression analysis

* Organizing a Library of Books : group books into genres or categories based on similarities

Unsipervised machine learning. Clustering algorithms  
- Program a robot to navigate and find the shortest path in a maze.

Reinforcement learning

**🌟 Exercise 5 : Designing an Evaluation Strategy for Different ML Models**

**Instructions**

* Select three types of machine learning models: one from supervised learning (e.g., a classification model), one from unsupervised learning (e.g., a clustering model), and one from reinforcement learning. For the supervised model, outline a strategy to evaluate its performance, including the choice of metrics

Supervised ML: Linear Regression

🡪 accuracy, precision, recall, F1-score, like cross-validation, ROC curves

Unsupervised ML: K-means clustering

Reinforcement learning: Q-learning

* For the unsupervised model, describe how you would assess the effectiveness of the model

Silhouette score, elbow method, or cluster validation metrics.

* For the reinforcement learning model, discuss how you would measure its success.

Cumulative reward, convergence, and exploration vs. exploitation balance.