**Introduction to Machine Learning**

**Exercise 1 – Defining the problem and data collection for loan default prediction**

Many people with different backgrounds go to their banks and request a loan for various reasons. It would be useful for the banks to be able to use a prediction model system to have more of a clear answer whether a customer should be given a loan or not.

For this to work large amounts of data isn’t needed, and in the case of a bank they already have a lot of historical data and personal data about customers.

I would collect large amounts of data on all current and historical customers, to give more of an insight into the type of customer the bank has. The type of data I would like to have are:

* Personal Information (age, employment, gender, relationships, location, education)
* Banking Information (savings, balances over time, income, previous loans, overdrafts, if they have defaulted on a previous loan, if they have paid off a previous loan, credit cards, assets)

I would gather this data from the bank itself and from other sources (credit card companies, government agencies).

**Exercise 2 – Feature selection and model choice for loan default prediction**

The most important data for the selection to determine if a loan should be given will be the customer’s income, bank balance, and loan history (including amount, defaults, and repayments).

I will then start an analysis of the selected data to try to find a pattern of why they maybe requesting a loan.

**Exercise 3 – Training, evaluating, and optimizing the model**

To train the model, I would input large amounts of data to ensure the model is continuously looking at many customers information and learning from each one. This will help with the evaluation of True Positive. For this type of prediction model it is very important to predict the positive class if a customer is going to default on a loan or not.

**Exercise 4 – Designing machine learning solutions for specific problems**

For each of these scenarios, decide which type of machine learning would be most suitable.

- Predicting Stock Prices: predict future prices 🡪 Reinforcement – because depending on the future stock price you will receive a reward or penalty  
- Organizing a Library of Books: group books into genres or categories based on similarities 🡪 Supervised – because you will already know the categories  
- Program a robot to navigate and find the shortest path in a maze 🡪 Unsupervised – because the robot will need to learn the patterns themselves without prior knowledge of the shortest path.

**Exercise 5 – Designing and evaluation strategy for different ML models**

Supervised Model: Pictures of food on a plate – I would have a database of thousands of pictures of foods. I would train the model to look at pictures of different foods on a plate to trey to recognize each individual food. Then I would give the model two pictures – the same food but the pictures would be different by the foods being in different locations on the plate – this would see if the model can recognize the individual foods in both pictures, by evaluating the performance and accuracy.

Unsupervised Model: Customer segmentation. Interesting buyer persona profiles can be created using unsupervised learning. This helps businesses to understand their customers' common traits and purchasing habits.

Reinforcement Model: Games – Getting the model to play through all the levels of a game to test it and find bugs. It will have the reward of completing a level successfully.