Machine Learning

Scenario: Credit Scoring

To predict whether a person is likely to **qualify for a loan**, based on basic financial and personal information.

Features:

- Age: age of individual (e.g. 28)
- Income: monthly income (e.g. 10000)
- Loan Amount: the requested loan amount (e.g. 5000)
- Credit Score: individuals current credit score (the higher the better) (e.g. 570)
- Current Dept: individuals current dept (e.g. 2500)
- Employment Status: encoded as 1 (employed) or 0 (unemployed)

Label will be the person applying for the loan.

Training Data will be the banks client history and data on previous applicants that have applied for a loan.

Expected Output:

- 0 (will qualify) for low-risk applicants
- 1 (not qualify) for high-risk applicants if certain risk factors are present like low income, high loan, poor credit score, etc.

Formula:

y (output 1 or 0) = $b + w_1 \cdot age + w_2 \cdot income + w_3 \cdot loan\_amount + w_4 \cdot credit\_score + w_5 \cdot dept + w_6 \cdot employment\_status$

Scenario: Predicting Future Product Sales for Inventory Management

To predict the **number of units** of a product that will be sold in the next week or month. It helps businesses avoid **overstocking** or **stockouts**.

Features:

- Product ID: unique id of the item (e.g. 1065)
- Week: number week of the year (1-52) (e.g. 45)
- Price: the current price of the item (e.g. 20.45)
- Discount: the percentage discount applied to the item that week (e.g. 15)
- Items Sold: number of items sold that week (e.g. 250)
- Foot traffic: the number of customers that visited the store that week (e.g. 5700)
- Promoted: if the product was being promoted that week (1=yes, 0= no)
- Previous Weeks Sales: the sales amount of the previous weeks sales (e.g. 800)

Label will be the number of units of an item is sold per week.

Training Data will be the stores previous sale history.

Expected Output: will be based on the predicted units sold that week which will conclude to the following:

- Products with higher discounts and promotions tend to sell more.
- The number of units sold for a specific item.
- The more foot traffic they get means more potential customers.

The output will also help to predict:

- Number of reorder stock to match forecasted demand.
- To avoid inventory shortage
- Helps reduce waste from unsold stock.

Formula:

$y\ (\text{sales}) = b + w_1 \cdot \text{price} + w_2 \cdot \text{discount} + w_3 \cdot \text{last\_week\_sales} + w_4 \cdot \text{items\_sold} + w_5 \cdot \text{discount} + w_6 \cdot \text{foot traffic} + w_7 \cdot \text{promoted}$

Scenario: Crop Yielding Prediction

Farmers and professionals in the agriculture industry rely on predictive insights into crop growth, as it directly impacts their business performance. This example of machine learning demonstrates how to predict crop yield based on environmental factors, helping agricultural professionals make informed decisions about resource allocation, planning, and risk management.

Label: Crop yield in kilograms per hectare (kg/ha)

Features

- Sunlight: the average number of sunlight hours per day.
- Water: weekly rainfall or water supply in millimeters.
- Temperature: average daytime temperature in °C.
- Soil pH: acid or alkaline level of the soil.

Training Data

The training data will consist of records from previous harvesting seasons, including recorded sunlight, rainfall, temperature, soil pH, and crop yields per hectare. These examples assist the model in learning the relationship between conditions and yield.

Expected Output

The two expected outcomes would be that:

- higher crop yields are predicted under optimal conditions, such as sufficient sunlight, water, temperature and a balanced soil pH.
- lower crop yields are predicted under suboptimal conditions, such as too little sunlight, too much water, poor soil, or extreme temperatures.

Formula:

y (crop yield) $= b + w_1 \cdot \text{sunlight} + w_2 \cdot \text{water} + w_3 \cdot \text{temperature} + w_4 \cdot \text{soil\_pH}$

Scenario: Predicting How Long a Phone Battery will Last

Goal: To predict the number of hours a phone's battery will last before it needs to be charged, based on the current usage behaviour and battery condition.

Features:

- Screen brightness– the average brightness level setting on the phone (e.g. 80%)
- App usage – the number of hours the phone is used for apps daily (e.g. 5 hours)
- Battery health – the overall health of the battery as a percentage (e.g. 85%)

Label: Battery life in hours

Training Data uses historical battery usage statistics from the same phone model. These logs include details like average screen brightness, hours of app usage, battery health, and the actual battery lifespan before the next charge. This helps the model learn patterns based on real user behaviour and device characteristics.

Expected Output:

- Predict how many hours a phone battery will last before requiring a recharge.
- Show the impact of brightness, app usage, and battery health on battery duration.
- Help users adjust settings to extend usage time.

Formula:

$y\ (\text{battery\_life}) = b + w_1 \cdot \text{brightness} + w_2 \cdot \text{app\_usage} + w_3 \cdot \text{battery\_health}$

Scenario 2: Predicting Delivery Time for Packages

Goal: To predict how many days it will take for a package to be delivered, based on shipping conditions and order details.

Features:

- Distance – between the sender and the receiver in kilometers (e.g. 307km)
- Package weight – in kilograms (e.g. 3kg)
- Shipping method – whether the method is standard (1) or priority (2)

Label: Delivery time in days (e.g. 5)

Training Data uses historical shipment and tracking data from past orders.

Expected Output:

- Predict how many days a package will take to reach its destination.
- Differentiate expected delays based on shipping type (standard vs priority).
- Allow customers to choose the best shipping option for faster delivery.

Formula:

$y\ (\text{delivery\_time}) = b + w_1 \cdot \text{distance} + w_2 \cdot \text{package\_weight} + w_3 \cdot \text{shipping\_method}$