Mathematics for Data Science

Project Brief

Project Deliverable

- Your deliverable will be a python notebook that will contain your solution.
- You will need to submit the shareable link to your notebook

Problem Statement

A user added a new record to a dataset but forgot to indicate the value for the variable (y). This variable (y) contains either the value 1 or 2.

As a data scientist, you need to develop a way that uses the database to guess the missing value and autocomplete that field for them. You will need to write a function that can classify a new record into 1 or 2. You will simply find the closest observation and assign its value for the missing variable (y) as the missing value. The distance used to make the comparison between the two observations can be either Euclidean or Manhattan. This is the same principle by which **K-Nearest Neighbors (KNN) Algorithm,** a popular distance-based machine learning algorithm, works.

A few things to note:

- Write the **nearest_neighbor_predict()** function. It takes three arguments:
    - Training set features (**train_features**)
    - Training set target (**train_target**)
    - New observation features (**new_features**)
- You can use the euclidean or manhattan function to calculate the distance. This function should take in the **train_features** values and **new_features** as the parameters.
- Use the **argmin()** function to return the shortest distance.

Dataset

You can find the dataset for this project in the guiding notebook.