**Project Title: TechR Project 0:** **UCI Repository AI4I 2020 Predictive Maintainance**

**Project Overview**:

The AI4I 2020 Predictive Maintenance Dataset is a synthetic dataset that reflects real predictive maintenance data encountered in industry.

The objective of the project was to complete the three steps essential in building a Machine Learning Model, based on a given dataset.

In this project, I carried out tasks like **EDA** (Exploratory Data Analysis), **Feature Engineering** and built a **Logistic Regression Model** with an accuracy of 89.3%.

**Team Members**:

Individual.

**Project Timeline**:

# December 22, 2023 - Start Date.

# December 23, 2024 – Research on various methods for project completion.

# December 26, 2024 – Actual coding complete.

# December 27, 2024 – Documentation and submission of project complete.

**Tools and Technologies Used**:

Jupyter Notebook, UCI Repository

**Dataset Description**:

Dataset Characteristics Subject Area Associated Tasks

Multivariate, Time-Series Computer Science Classification, Regression, Causal-Discovery

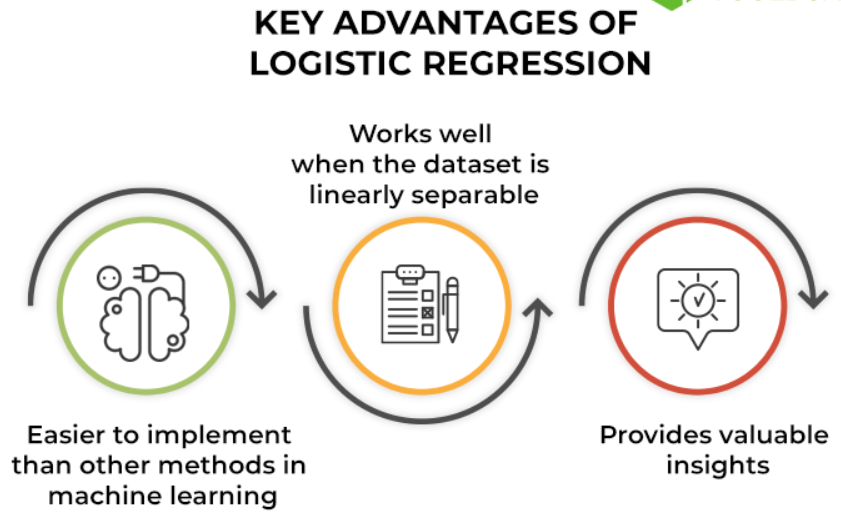
Feature Type # Instances # Features

Real 10000 6

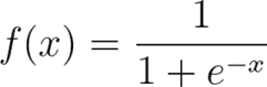
**Model Architecture**:

The type of model that I have built is a Logistic Regression Model.

Logistic regression is a supervised machine learning algorithm that accomplishes binary classification tasks by predicting the probability of an outcome, event, or observation. The model delivers a binary or dichotomous outcome limited to two possible outcomes: yes/no, 0/1, or true/false.

It analyses the relationship between one or more independent variables and classifies data into discrete classes. It is extensively used in predictive modelling, where the model estimates the mathematical probability of whether an instance belongs to a specific category or not.

The sigmoid function is referred to as an activation function for logistic regression and is defined as:

where,

e = base of natural logarithms

value = numerical value one wishes to transform

**Training Process**:

For building the Logistic Regression Model, **Train-Test Split** Method was used.

The first step done is this method was the dropping of dependent features, which might provide problems in building a classification-based regression model.

Train-Test Split involves taking a dataset and dividing it into two subsets. The first subset is used to fit the model and is referred to as the training dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset.

The Train-Test split in this project was 70:30.

**Model Evaluation**:

Apart from the Train-Test Split, a concept known as **Confusion Matrix** is also implemented.

A confusion matrix is a matrix that summarizes the performance of a machine learning model on a set of test data. It is a means of displaying the number of accurate and inaccurate instances based on the model’s predictions. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance.

The matrix displays the number of instances produced by the model on the test data.

**True positives (TP):** occur when the model accurately predicts a positive data point.

**True negatives (TN):** occur when the model accurately predicts a negative data point.

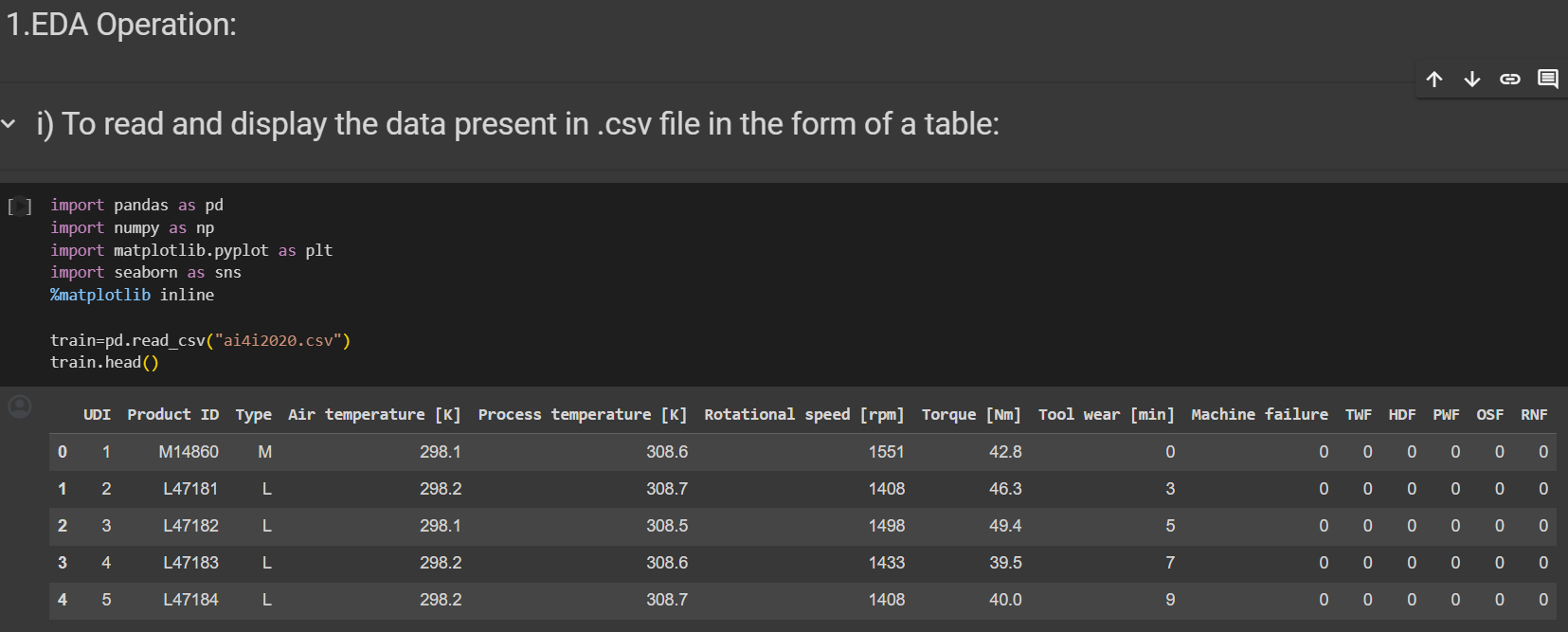
**False positives (FP):** occur when the model predicts a positive data point incorrectly.

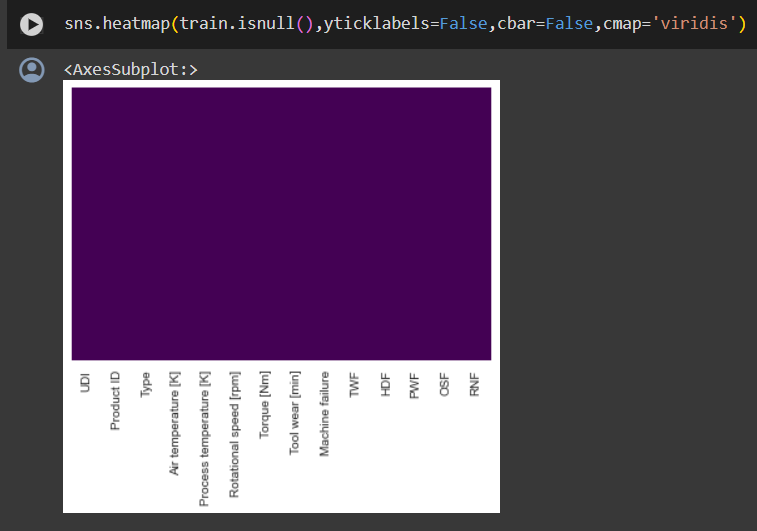
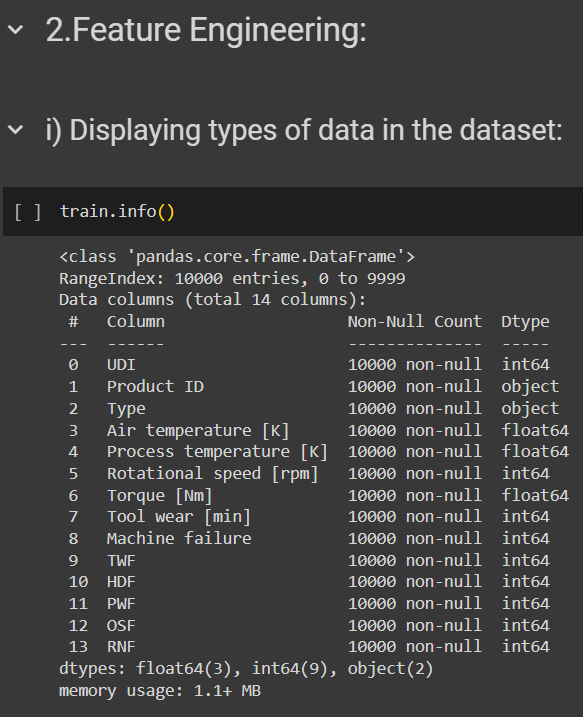
**False negatives (FN):** occur when the model misinterprets a negative data point.

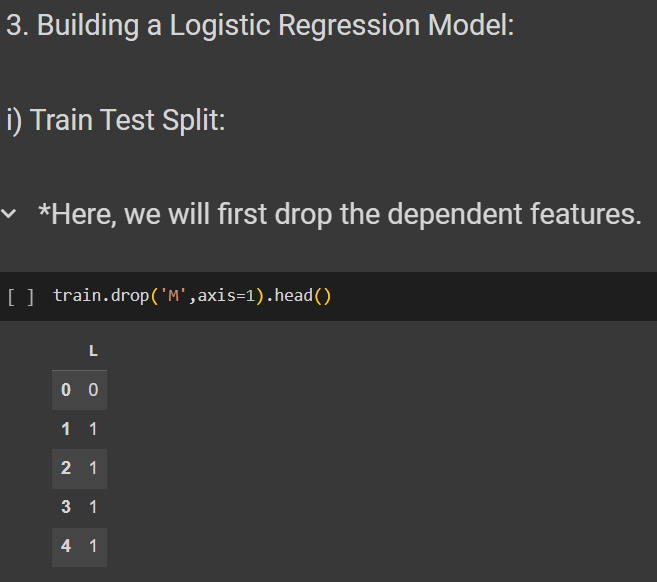
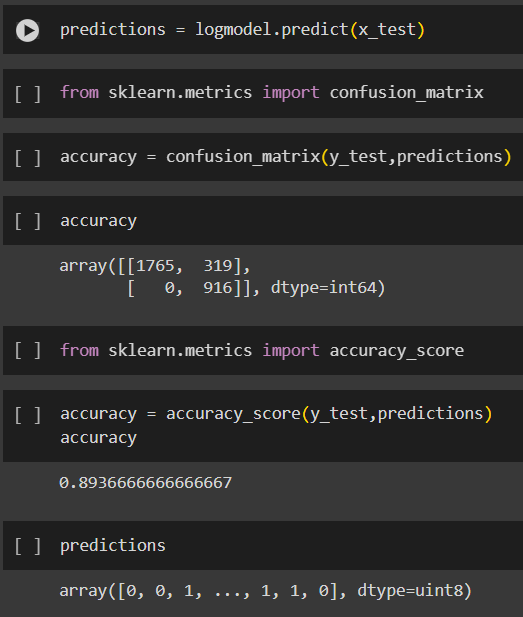
The accuracy of the model was also found, with the help of these two concepts.

Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome. You can calculate accuracy by dividing the number of correct predictions by the total number of predictions.

The accuracy of given model was found to be 89.3%, which proves that the model is widely successful.

**Results**:



****

**Challenges Faced and Resolutions**:

The project was a simple one overall, but since it was the very introduction to Machine Learning, it proved to be a bit challenging to me.

But in the process of researching for this project, I learnt many new concepts in the field of Machine Learning, and was able to complete the project successfully.

**Lessons Learned**:

The lessons I learnt throughout the duration of this project was to keep a **learning attitude**. I have learnt many new concepts throughout the duration of this project, and will continue learning throughout the duration of the internship.

**Future Work**:

The future scope for a predictive maintenance project using the UCI Repository AI4I 2020 Predictive Maintenance dataset can be vast. Here are some ways in which this project can be expanded:-

Advanced Machine Learning Techniques: Experiment with advanced machine learning algorithms such as deep learning (e.g., recurrent neural networks, convolutional neural networks) to improve predictive accuracy and identify more complex patterns in the data.

Real-Time Monitoring: Develop systems for real-time monitoring of machinery health using streaming data. This could involve integrating predictive models into industrial control systems to provide early warnings of potential failures.

Fault Diagnosis: Extend the project to include fault diagnosis capabilities, where the model not only predicts when maintenance is needed but also identifies the specific type of fault or anomaly occurring in the machinery.

Edge Computing: Explore the feasibility of deploying predictive maintenance models on edge devices located near the machinery. This could reduce latency and enable real-time decision-making without relying on cloud infrastructure.

Collaborative Maintenance: Explore collaborative predictive maintenance approaches where data from multiple similar machines or across different sites are combined to improve the accuracy and robustness of predictive models.

**Conclusion**:

In conclusion, the project was a simple one, in which I learnt various new concepts like EDA, Feature Engineering, Logistic Regression, Train-Test Split, Confusion Matrix, Accuracy, etc.

This project was a great introduction to the field of Machine Learning for me.

**References**:

<https://archive.ics.uci.edu/dataset/601/ai4i+2020+predictive+maintenance+dataset>

<https://www.semanticscholar.org/paper/Explainable-Artificial-Intelligence-for-Predictive-Matzka/b609c8e9ec6a2b8c642810953ef6dffe5766f7c1#related-papers>