

WOC 4.0

Machine Learning BootCamp

**Project Report**

**Arushika Bansal**

**21JE0170**

**Mathematics and Computing**

***Description***

**Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn without being explicitly programmed. In the span of one month, I have implemented four algorithms from scratch using only basic python libraries like NumPy, Pandas and matplotlib.**

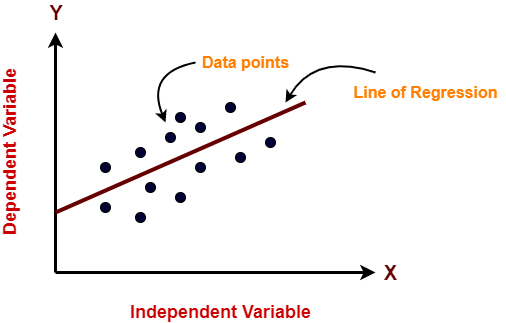
**I have then compared the y values I predicted and the one provided in the dataset to check the accuracy of the model.**

***Project Implementation***

* **Linear regression**

**Linear regression models are used to show or predict the relationship between two variables or factors. It is an algorithm based on Supervised learning.**

**For my model : RMSE = 0.0027827**

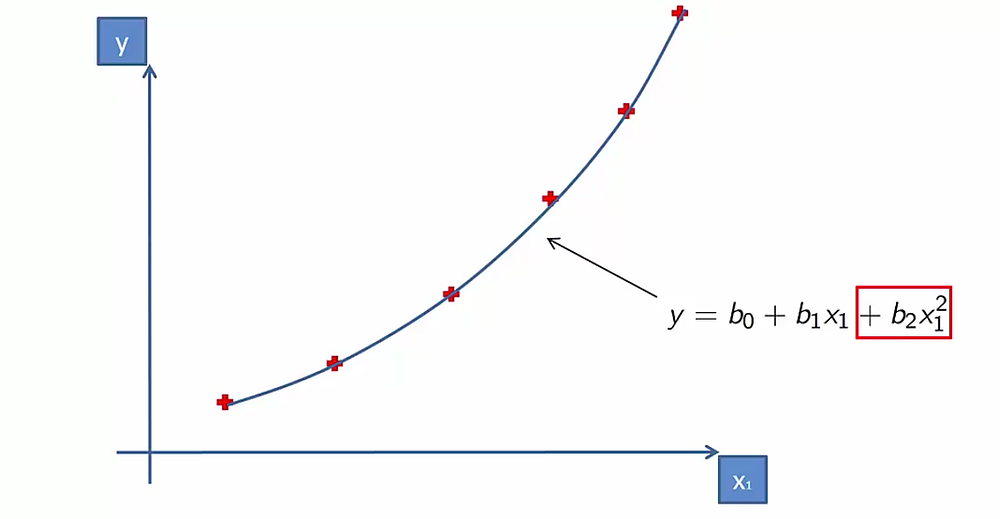


* **Polynomial regression**

**It is a special case of linear regression where only due to the Non-linear relationship between dependent and independent**

**Variables, we add some polynomial terms to linear regression to convert it into polynomial regression.**

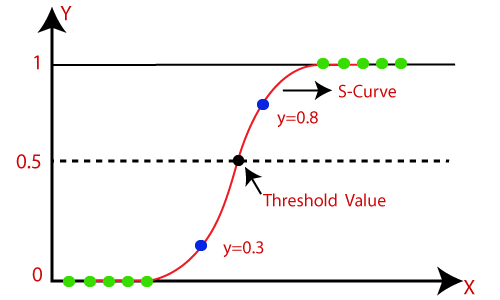
**For my model: RMSE = 0.00717698**



* ***Logistic regression***

**This algorithm uses the sigmoid function for predicting in its model. Logistic regression is based on supervised learning. It is used to understand the relationship between dependent and independent variables by estimating the probability using a logistic regression equation.**

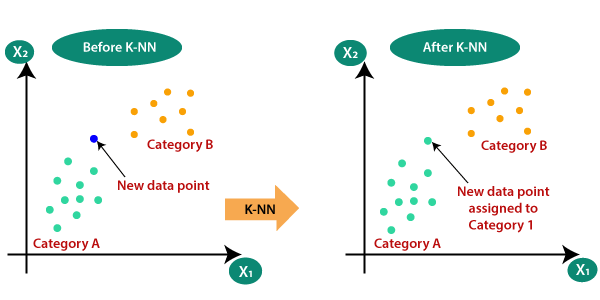
**For my model: Accuracy = 66.8918**



* ***KNN (K -Nearest Neighbour)***

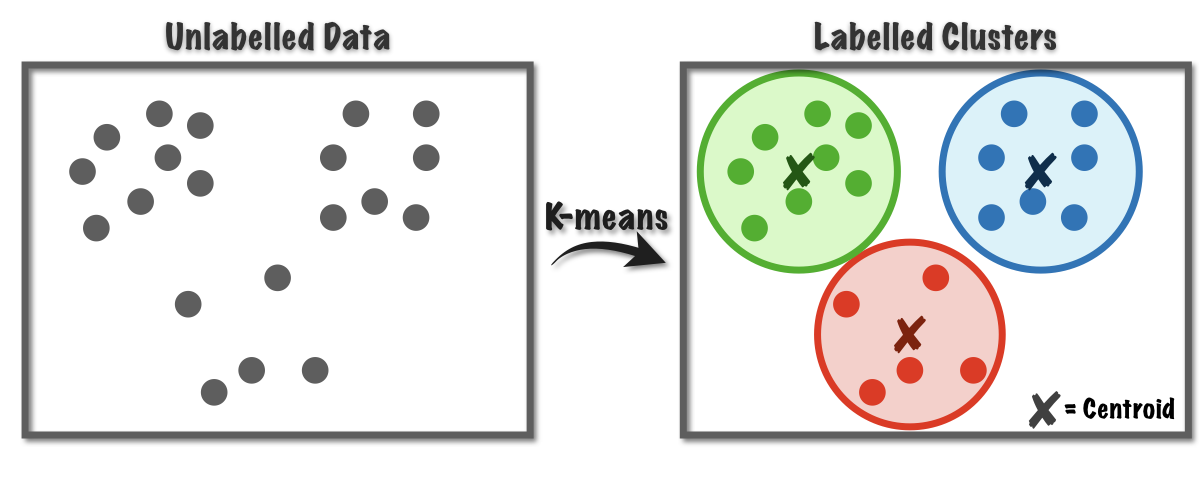
**It works by finding the distance between a query and all the examples in the data, selecting the specified number of examples (K) closest to the query, then votes for the most frequent label(in case of classification) or averages the labels (in the case of regression). It falls under the Supervised Learning category.**

**For my model: Can’t report due to time constraint**



* ***K-Means Clustering***

**This algorithm is used when you have an undefined dataset. The goal of this algorithm is to find groups in the data, with the number of groups represented by K. This algorithm falls under the category of Unsupervised learning.**

****

* ***Neural Networks***

**Neural Networks are a class of machine learning algorithms used to model complex patterns in datasets using multiple hidden layers and non-linear activation functions.**

**For my model: RMSE = 5175.699 [Linear data]**

**RMSE = 1205.9032 [Polynomial data]**

**Accuracy = Can’t report due to time constraint**

# Technology Stacks

* **Python: Python is a high-level programming language for overall programming. Besides being an open-source programming language, python is an extraordinarily interpreted, object-oriented, and interactive programming language.**
* **NumPy: NumPy stands for numerical python. It Is a Python Library used for working with arrays. It also has functions for working in the domain of linear algebra, Fourier transform, and matrices.**
* **Pandas: Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis".**
* **Matplotlib: Matplotlib is a low-level graph plotting library in python that serves as a visualization utility. It is mostly written in Python.**
* **Google collaboratory**

**About Me:**

* ***Personal Details:***
* **Name: Arushika Bansal**
* **Branch: Mathematics and Computing**
* **University: Indian Institute of Technology (Indian School of Mines), Dhanbad**
* **Year: First year**
* **Email:** [**21JE0170@iitism.ac.in**](mailto:21JE0170@iitism.ac.in)
* **Linkedin Id: Arushika Bansal**
* **Contact number: +91-8306890385**