**A picture containing application

Description automatically generated**

**Project Proposal:**

**Predicting Likes Quantity For Data Sets In Kaggle**

**SDAIA ACADEMY (T5)**

**By Shaikha Bin Ateeq**

**Overview** :

Kaggle is a data science and machine learning online community. Users can use Kaggle to acquire and publish data sets, explore and build models in a web-based data-science environment, collaborate with other data scientists and machine learning experts, and compete in data science challenges.[1] However the number of likes on the dataset may be considered a significant point, as the selection of the dataset may depend on it, to develop a solution that aims to predict the number of likes on data set based on a machine learning algorithm; specifically, we will be using a linear regression model. To do so, we will use a scraped date from the Kaggle dataset in ‘computer since’ To accomplish the prediction, The working flow will be as the following:

**Data Description**

The data set for the model will be scraped from Kaggle website: https://www.kaggle.com/. The below table outlines the features that will be used to develop the prediction model.

* Dataset Name: The name of the dataset
* Created By: the person how published the dataset
* Rank: There are three levels of award, Bronze, Silver and Gold
* Like: number of like in the dataset
* Download Size: download size of the dataset
* Notebook: number of Jupyter notebooks of the dataset
* View:The number of people who had viewed the data

**Tools:**

To explore and analyze the data, I will be using Jupyter notebook to use python language and Python libraries, such as:

* Pandas and NumPy packages for data manipulation.
* Matplotlib, seaborn library for data visualization.
* BeautifulSoup library for web scraping.
* Selenium is a Python library and tool used for automating web browsers
* LinearRegression from the sklearn.linear\_model package for applying the linear regression model

**References :**

[1] Kaggle - Wikipedia. (2021). Retrieved 29 November 2021, from https://en.wikipedia.org/wiki/Kaggle