****

****

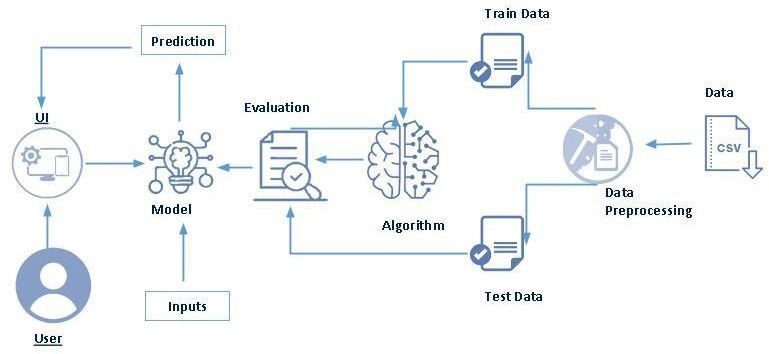
****

**Harnessing AI to Predict Best Selling Manga**

**Machine Learning**

**Harnessing AI to Predict Best Selling Manga Prediction by using Machine Learning**

Since the 1950s, manga has become an increasingly major part of the Japanese publishing industry By 1995, the manga market in Japan was valued at [¥](https://en.wikipedia.org/wiki/Japanese_yen)586.4 billion ([$](https://en.wikipedia.org/wiki/United_States_dollar)6–7 billion) with annual sales of 1.9 billion manga books and [manga magazines](https://en.wikipedia.org/wiki/Manga_magazines) (also known as manga anthologies) in Japan (equivalent to 15 issues per person). In 2020 Japan's manga market value hit a new record of ¥612.6 billion due to the fast growth of digital manga sales as well as increase of print sales in 2022 Japan's manga market hit yet another record value of ¥675.9 billion. Manga have also gained a significant worldwide readership. Beginning with the late 2010s manga started massively outselling American comics. Now the Manga want to know the Best-Selling Mangas from Past to Now.



**Project Flow:**

● User interacts with the UI to enter the input.

● Entered input is analysed by the model which is integrated.

● Once model analyses the input the prediction is showcased on the UI To accomplish this, we have to complete all the activities listed below,

**● Define Problem / Problem Understanding**

○ Specify the business problem

○ Business requirement

○ Literature Survey

○ Social or Business Impact.

**● Data Collection & Preparation**

○ Collect the dataset ○ Data Preparation

**● Exploratory Data Analysis**

○ Descriptive statistical ○ Visual

**● Model Building**

○ Training the model in multiple algorithms

○ Testing the model

**● Performance Testing & Hyperparameter Tuning**

○ Testing model with multiple evaluation metrics

○ Comparing model accuracy before & after applying hyperparameter tuning

**● Model Deployment**

○ Save the best model

○ Integrate with Web Framework

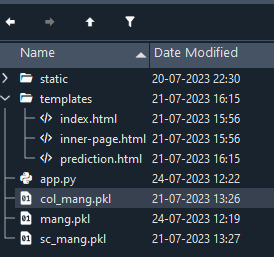
**● Project Demonstration & Documentation**

○ Record explanation Video for project end to end solution

○ Project Documentation-Step by step project development procedure

**Project Structure:**

Create the Project folder which contains files as shown below



● We are building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scripting.

● model.pkl is our saved model. Further we will use this model for flask integration.

● Training folder contains a model training file.

**Milestone 1: Define Problem / Problem Understanding**

Activity 1: Specify the business problem

Refer Project Description

**Activity 2: Business requirement**.

The accurate estimation of Best-Selling Manga is crucial for Company as it allows them to company to know which manga is selling highest. By this Analysis the company will gain the complete information about their manga’s selling. The company easily identify their low selling Manga’s and High selling Mangas. And identify their weakness and strengths and make solutions to improve in their weakness.

**Activity 3: Literature Survey (Student Will Write)**

A literature survey for a Best-Selling Manga project would involve researching and reviewing existing studies, articles, and other publications on the topic of Manga Series. The survey would aim to gather information on current classification systems, their strengths and weaknesses, and any gaps in knowledge that the project could address. The literature survey would also look at the methods and techniques used in previous Manga projects, and any relevant data or findings that could inform the design and implementation of the current project.

**Activity 4: Social or Business Impact.**

**Social Impact:**  By providing accurate and up-to-date information on Selling Manga’s project can help Company make more informative about the Top Selling Mangs and their strengths & weakness

**Business Model/Impact:** By providing information on the properties and interactions of Selling Mangas, The company will know what are the low selling manga’s and try to improve their low selling manga’s

**Milestone 2: Data Collection & Preparation**

ML depends heavily on data. It is the most crucial aspect that makes algorithm training possible. So, this section allows you to download the required dataset

**Activity 1:** Collect the dataset There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.

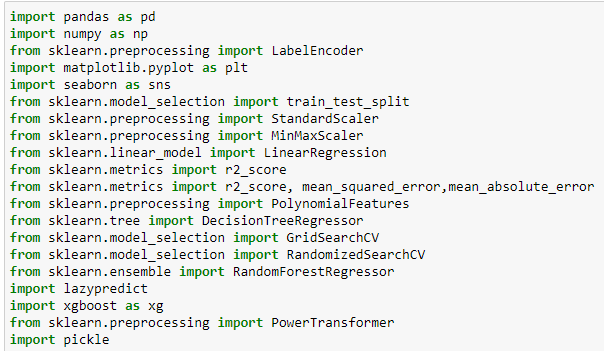
In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to download the dataset

Link: - https://www.kaggle.com/datasets/drahulsingh/best-selling-manga

As the dataset is downloaded. Let us read and understand the data properly with the help of some visualisation techniques and some analysing techniques.

Note: There are a number of techniques for understanding the data. But here we have used some of it. In an additional way, you can use multiple techniques.

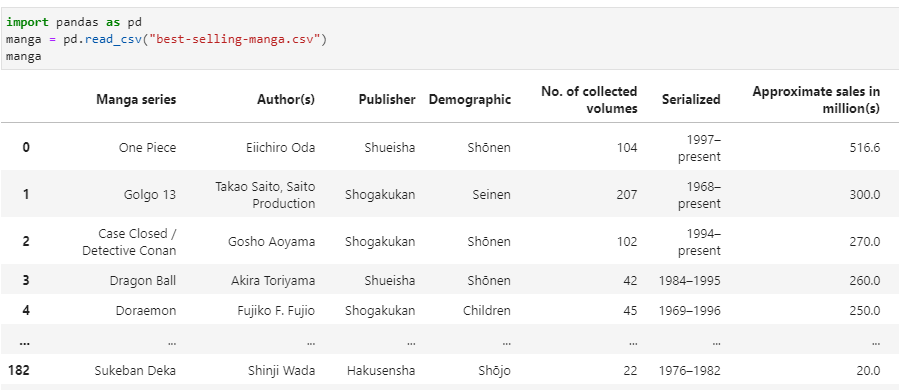
**Activity 1.1:** Importing the libraries Import the necessary libraries as shown in the image.



**Activity 1.2: Read the Dataset**

Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.

In pandas we have a function called **read\_csv ()** to read the dataset. As a parameter we have to give the directory of the csv file.



**Activity 2: Data Preparation**

As we have understood how the data is, let's pre-process the collected data. The download data set is not suitable for training the machine learning model as it might have so much randomness so we need to clean the dataset properly in order to fetch good results.

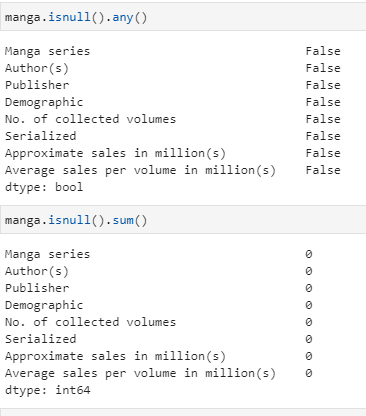
This activity includes the following steps.

● Handling missing values

● Handling categorical data

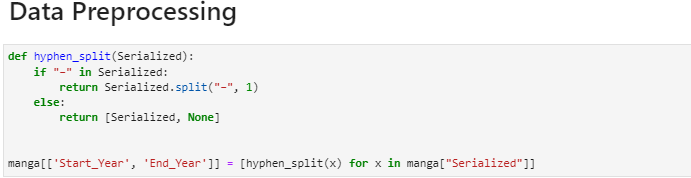
● Handling Outliers

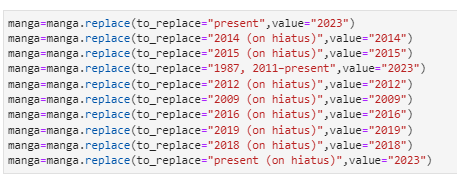
Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps



**Activity 1:**

In this Manga Series data there is one column with combination of numeric and string Data.so, I have divided the column in to columns and replaced the string data into numeric.

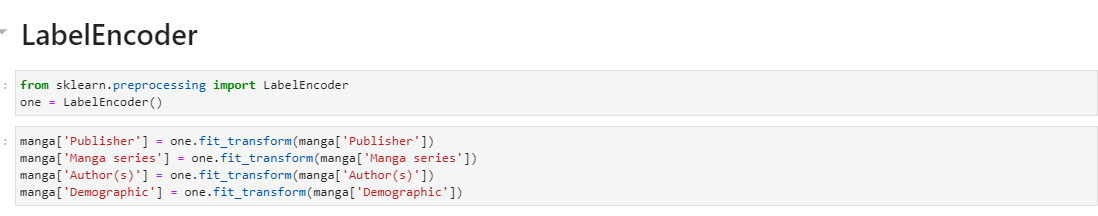




**Activity 2.2: Handling Categorical Values**

As we can see our dataset has categorical data, we must convert the categorical data to integer encoding or binary encoding.

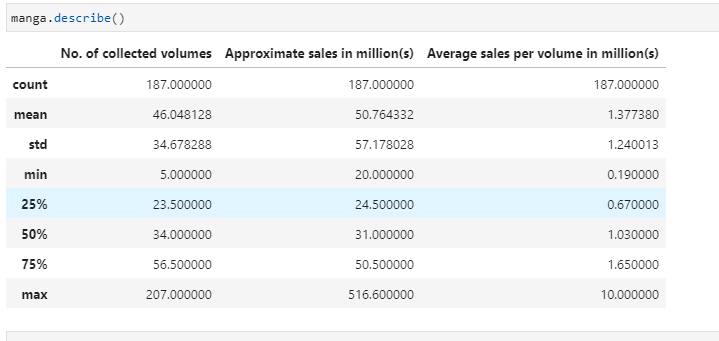
To convert the categorical features into numerical features we use encoding techniques. There are several techniques but, in our project, we are using manual encoding with the help of list comprehension.



**Milestone 3: Exploratory Data Analysis**

**Activity 1**: Descriptive statistical Descriptive analysis is to study the basic features of data with the statistical process.

Here pandas have a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.

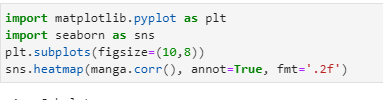


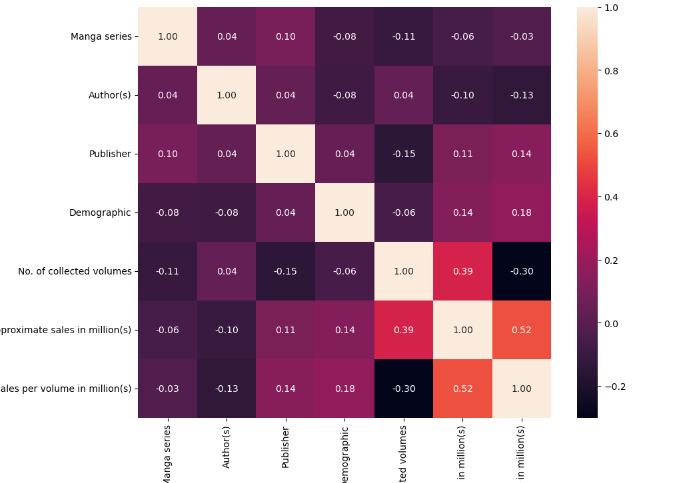
**Activity 2: Visual analysis**

Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data. It is a way to quickly identify patterns, trends, and outliers in the data, which can help to gain insights and make informed decisions.

**Activity 2.1: Heatmap analysis**

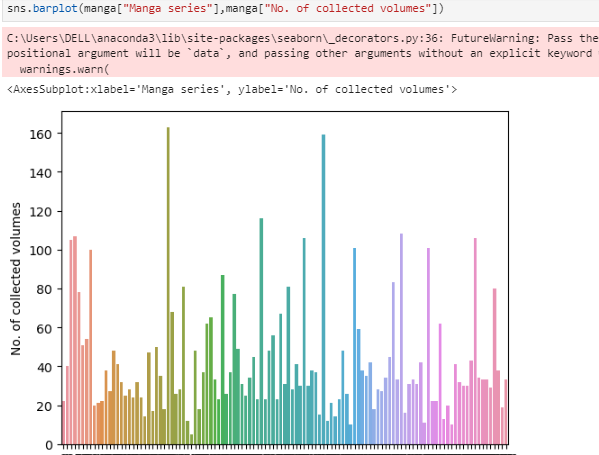
A heatmap is a graphical representation of data that uses a system of color-coding to represent different values.





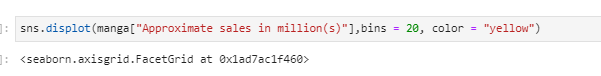
**Activity 2.1: Bar chart analysis**

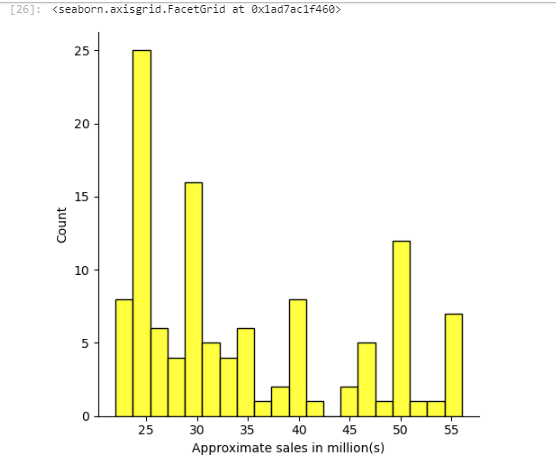
By this bar chart we can see the no of collected volumes by Manga series Wise.



**Activity 2.2: Bar chart analysis**

By this displot we can see the range of the Approximate sales in millions

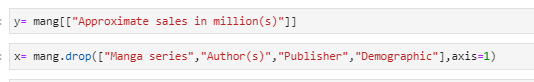




**Splitting data into train and test**

Now let’s split the Dataset into train and test sets. First split the dataset into x and y and then split the data set

Here x and y variables are created. On x variable, df is passed with dropping the target variable. And on y target variable is passed. For splitting training and testing data we are using train\_test\_split () function from sklearn. As parameters, we are passing x, y, test size, random state.

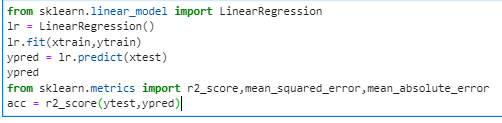


**Milestone 4: Model Building**

**Activity 1:** Training the model in multiple algorithms Now our data is cleaned and it’s time to build the model. We can train our data on different algorithms. For this project we are applying three classification algorithms. The best model is saved based on its performance.

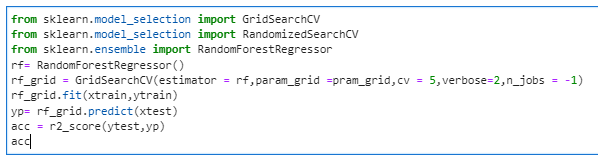
**Activity 1.1: Linear regression model**

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.



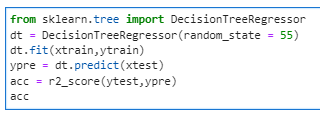
**Activity 1.2: Random Forest model**

A function named random Forest is created and train and test data are passed as the parameters. Inside the function, Random Forest Classifier algorithm is initialised and training data is passed to the model with. Fit () function. Test data is predicted with. predict () function and saved in a new variable. For evaluating this analysis r2Score is used.



**Activity 1.3: Decision Trees model**

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final reult is a tree with decision nodes and leaf nodes.

****

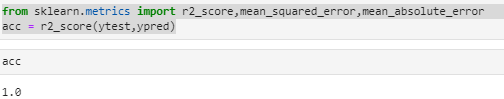
**Milestone 5: Performance Testing & Hyperparameter Tuning**

**Activity 1:** Testing model with multiple evaluation metrics Multiple evaluation metrics means evaluating the model's performance on a test set using different performance measures. This can provide a more comprehensive understanding of the model's strengths and weaknesses. We are using evaluation metrics for classification tasks including accuracy, precision, recall, support and F1-score.

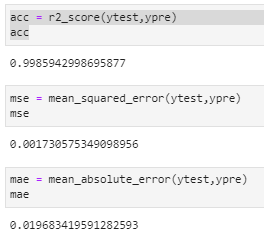
**Activity 1.1: Compare the model**

For comparing the above three models, the compare Model function is defined.

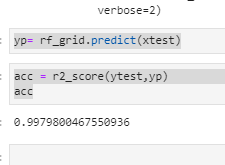
**Linear Regression**

****

**Decision Tree**

****

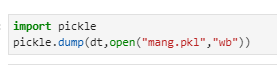
**RandomForest**

****

**Milestone 6: Model Deployment**

**Activity 1:** Save the best model

Saving the best model after comparing its performance using different evaluation metrics means selecting the model with the highest performance and saving its weights and configuration. This can be useful in avoiding the need to retrain the model every time it is needed and also to be able to use it in the future.



**Activity 2: Integrate with Web Framework**

In this section, we will be building a web application that is integrated to the model we built. A UI is provided for the uses where he has to enter the values for predictions.

The enter values are given to the saved model and prediction is showcased on the UI. This section has the following tasks

● Building HTML Pages

● Building server-side script

● Run the web application

**Activity 2.1:** Building Html Pages: For this project create two HTML files namely

• home.html

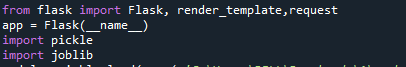
• predict.html

• submit.html

and save them in the templates folder. Refer this link for templates.

**Activity 2.2:**

Build Python code:



Import the libraries Load the saved model. Importing the flask module in the project is mandatory. An object of Flask class is our WSGI application. Flask constructor takes the name of the current module (\_\_name\_\_) as argument.



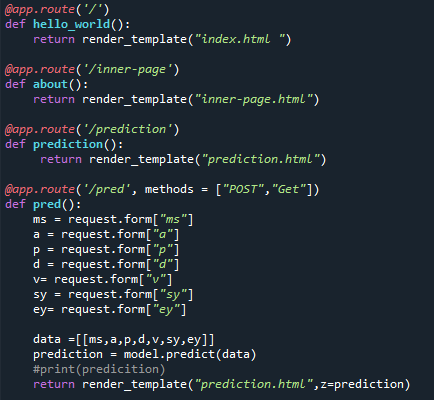
**Render HTML page**



Here we will be using a declared constructor to route to the HTML page which we have created earlier.

In the above example, ‘/’ URL is bound with the home.html function. Hence, when the home page of the web server is opened in the browser, the html page will be rendered. Whenever you enter the values from the html page the values can be retrieved using POST Method.

Retrieves the value from UI



Here we are routing our app to predict () function. This function retrieves all the values from the HTML page using Post request. That is stored in an array. This array is passed to the model. Predict () function. This function returns the prediction. And this prediction value will be rendered to the text that we have mentioned in the submit.html page earlier.

**Main Function:**



**Activity 2.3: Run the web app application**

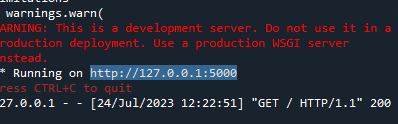
● Open anaconda prompt from the start menu

● Navigate to the folder where your python script is.

● Now type “python app.py” command

● Navigate to the localhost where you can view your web page.

● Click on the predict button from the top left corner, enter the inputs, click on the submit button, and see the result/prediction on the web.



Now, Go the web browser and write the localhost URL(http://127.0.0.1:5000)to get the below result

**Home Page:**



**About Page:**



**Prediction Page:**

****

****

**Milestone 7:**

Project Demonstration & Documentation Below mentioned deliverables to be submitted along with other deliverables

**Activity1: -** Project Documentation-Step by step project development procedure Create document as per the template provided