**Predicting Osteoporosis**

**Project title:** A predictive model for predicting osteoporosis risk.

**Progress so far:**

**Completed Tasks**

* **Importing all the required libraries**

In this the libraries that we have used are NumPy, Pandas, Matplotlib. All these libraries can be used to perform various operations on the dataset. Here, NumPy library is used to perform all the mathematical operations on the numerical data that is present in the dataset. Pandas library is used for data preprocessing tasks, all the missing data can be handled. Matlplotlib library is used for the process of data visualization to check for the behaviour of the data.

* **Importing the dataset**

The dataset is imported using the Pandas library. The dataset is in the CSV format. The columns in the dataset are divided into independent and dependent variables using the index lock functions.

* **Dataset Information**

The information regarding the dataset can be known using the info method to understand the datatypes of the attributes present in the dataset. By this we can understand, if all the data is in numerical format that can be understood by the machine learning model. If there is any categorical data it can be converted into numerical data.

* **Data Preprocessing**

The head method is used to see the top five rows in the dataset. This gives an idea about the data present in the dataset. We can understand about all the attributes present in the dataset.

To check if there are any missing values present in the dataset,.isnull().sum() method that gives us the number of null values present in the data of each attribute of the dataset. If there are any missing values in the dataset, it can be replaced with mean of all other data present in that attribute. There are no null values in the dataset, I have checked it suing the function. Therefore, we can move further.

**Data**

The osteoporosis dataset is taken from Kaggle.

Link: <https://www.kaggle.com/datasets/amitvkulkarni/lifestyle-factors-influencing-osteoporosis>

**Challenges:**

All of us have collaborated for the competition of the project. We communicated with each other regularly and divided the work among ourselves.

The challenge for us is checking if there are any null values present in the dataset, but there are no null values so we proceeded further.

Then choosing the independent variables that are responsible for predicting using the model. And then converting the categorical data into numerical data using one hot encoding.

**Collaboration:**

We are connecting in Zoom and in-person every week and sharing our work equally.

**Next steps:**

Further steps are one-hot encoding the categorical data, and visualizing the dataset. The dataset is split into training set and test set. Then the Machine learning model is built. An finally the accuracy of the models is calculated and the model with maximum accuracy is selected.