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

**Introduction:** Osteoporosis is a significant health concern that results in increased fracture risk and reduced bone mass. The disease is very complex and influenced by a lot of factors like lifestyle choices, dietary habits, demographic variables, and medical history. There is an increased need for the development of predictive models that can identify individuals who are at risk of developing this disease. This project proposes the development of a predictive model that can accurately predict osteoporosis risk in individuals using data related to the causative factors of the disease. The project aims to facilitate early intervention of the disease and reduce the impact of osteoporosis on individuals and healthcare professionals.

* Team members:
* Domain: Healthcare and preventive medicine
* Problem statement: Developing a machine learning model that can predict osteoporosis risk in an individual based on affecting factors.
* Motivation: Early detection of osteoporosis can reduce the risk of fractures and other consequences. If osteoporosis is not detected on time, it can lead to disabilities and decreased mobility.

**Related work:** Kim et al. have developed and validated machine learning models that can identify the risk of osteoporosis in menopausal women. They have compared the developed model to traditional clinical decision tools and self-assessment tools. They collected medical records of Korean postmenopausal women to train and build models like SVM, RF, and ANN. Results show that SVM had a better accuracy when compared to other models (Kim et al, 2013). In another paper, Chiu et al. have developed and validated an Artificial neural network to identify elderly individuals with osteoporosis risk. Results indicated that ANN had the potential to act as a tool for identifying elderly people at osteoporosis risk (Chiu et al., 2006).

Even though these studies have contributed significantly to solving the problem, we aim to develop a model that takes into consideration all the factors that influence the diseases. Our project is different from other projects in terms of broader coverage of influencing factors.

**Data Plan:**

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

This dataset has comprehensive data on factors that influence the development of osteoporosis like demographic details, lifestyle choices, medical history, bone health indicators, age, gender, and hormonal changes.

* The dataset may require preprocessing. We will explore the data for any potential outliers or missing values. Missing values and outliers will be handled based on their nature. We will deploy feature selection techniques to select the most appropriate features for developing our model.

**Implementation Plan:**

* Data exploration and preprocessing
* Feature selection
* Splitting data into training and testing sets
* Training the model and evaluating it
* Hyperparameter tuning
* Model selection and deploying it

**Implementation Tools:**

For the data mining timeline: We will use Python, NumPy, Pandas, Scikit-learn, and Matplotlib for data manipulation, model development, and evaluation.

**Evaluation Plan:**

We plan to evaluate our model using performance metrics like accuracy, F1-score and ROC-AUC. To see if our model’s results are generalizable, we plan to implement the cross-validation technique.

**Plan for group collaboration:**

We plan to meet via video conferencing to discuss the project’s progress, and challenges and to co-ordinate tasks. We will assign tasks and responsibilities within the group. If possible, we will meet in person to collaborate and communicate effectively. We plan to use a shared cloud storage service that will be accessible to all the group members. We plan to document the preprocessing, any engineering steps, and the entire data mining timeline to avoid confusion.

**Timeline:**

Week 1: Project planning, reviewing existing literature, and exploring datasets.

Week 2: Data preprocessing and feature engineering

Week 3: Model selection

Week 4: Training model and evaluating it

Week 5: Hyperparameter tuning

Week 6: Final model selection, report preparation, and project presentation.

**References:**

Sung Kean Kim, Tae Keun Yoo, Ein Oh, & Deok Won Kim. (2013). Osteoporosis risk prediction using machine learning and conventional methods. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. <https://doi.org/10.1109/embc.2013.6609469>

Chiu, J. S., Li, Y. C., Yu, F. C., & Wang, Y. F. (2006). Applying an artificial neural network to predict osteoporosis in the elderly. *Studies in health technology and informatics*, *124*, 609.