# **Artificial Intelligence and Machine Learning (AIML) – Project**

**Names: Sec: 2A**

Hema Mayukha Vasireddy – 2320040020, Sec A

Anishka Sree Rapolu – 2320040041, Sec A

Sowmya Sree I – 2320040031, Sec A

**Problem Statement: 3D Structure Prediction from Protein Sequences Using Computational Methods**

Proteins, essential molecular machines within biological systems, derive their function from their three-dimensional (3D) structures. Experimentally determining these structures is often labor-intensive and costly. This project aims to develop a computational approach to predict the 3D structure of a protein directly from its amino acid sequence, utilizing a combination of bioinformatics tools and machine learning algorithms.

**Dataset:**

1. **Title:** Protein Data Bank (PDB)

**Source:** RCSB Protein Data Bank

1. **Title:** SCOP (Structural Classification of Proteins)

**Source:** SCOP

**Algorithm:**

1. **Data Collection**:

* **Datasets**: Protein sequences and corresponding structural data from online repositories.

2. **Sequence Analysis**:

* **Secondary Structure Prediction**: Utilize tools like PSIPRED to predict secondary structures (alpha-helices, beta-sheets).
* **Homology Modeling**: Apply homology modeling techniques using MODELLER to generate 3D models based on known structures of similar proteins.

3. **Model Refinement**:

* **Energy Minimization**: Use tools such as PyRosetta to refine the predicted structure, minimizing energy to achieve a more stable and accurate model.

4. **Validation**:

* **Structural Validation**: Employ validation tools like PROCHECK to assess the quality of the predicted 3D model.

5. **Visualization**:

* **3D Model Visualization**: Visualize the final 3D structure using molecular visualization tools like PyMOL or Chimera.

**Expected Outcome:**

This project aims to produce accurate 3D models of proteins directly from their amino acid sequences, providing insights into protein function, interactions, and stability. The predicted structures can aid in understanding diseases related to protein misfolding, such as Alzheimer's and Parkinson's, and contribute to drug design by identifying potential binding sites for therapeutic molecules. Additionally, the outcomes could support protein engineering efforts by enhancing the stability and solubility of proteins for industrial and therapeutic applications. Beyond practical benefits, the project will serve as an educational resource, demonstrating the integration of machine learning and bioinformatics in solving complex biological problems, ultimately bridging the gap between computational predictions and experimental validations in structural biology.