Delhi Technological University
Engineering Physics
Vth Semester

Project Proposal

EP307 : BioPhysics

Project Members

Member 1 : ARKAJYOTI CHAKRABORTY, 2K19/EP/022

Member 2 : ADITYA SINGH, 2K19/EP/005

Project Details

Subject : **BIOPHYSICS**

Topic: **DNA CLASSIFICATION**

Overview

1. Deoxyribonucleic acid (DNA) is a biological macromolecule. Its main function is information storage.
2. We will apply the Machine Learning Classification model which is a powerful technique for analyzing and predicting coding sequences of DNA.
3. We will use Biopython, a module in the python programming language to identify and perform some basic characterization of a coronavirus or any human body genome sequence.
4. The objective of this project is to introduce some of the Biopython modules and predict a gene's function based on its DNA sequence in an applied biological context. The use of a genome is merely illustrative, the analyses are generic, and could be applied to any small genome.

Analysis

- In this project we have done DNA sequencing classification over a dataset that is taken from Kaggle.
- Now DNA has four types of acronyms that are ATGC, so our target is to study the patterns of the bases and then classifying them into different classes of human genes.
- The method we use here is we first take the long biological sequence and break it down into k-mer length overlapping “words”.
- We will be using different types of machine learning classification models to derive a comparative report about the algorithms that performed the best among the algorithms that are being used.
- The study can be taken into consideration and applied in classifying DNA’s when needed in practical work.

Tools Used:

1. Machine Learning Techniques
2. Deep Learning Techniques
3. Python Programming Language
4. Google Collab
5. Classification
6. SnapGene for Results