Aim of the Project:Sequence alignment over large databases cannot yield results within a reasonable time,power, and cost.The aim of the project is to develop a machine learning model to accelerate the performance of sequence alignment.

Motivation:

The motivation behind solving the problem of sequence alignment over large databases is to identify s

Title Justification:

Sequence Alignment in Biological sequences using Machine Learning

Sequence alignment is a fundamental task in bioinformatics and computational biology that involves comparing the similarities and differences between biological sequences such as DNA, RNA, or protein. This task is crucial for understanding the function, structure, and evolution of genes and proteins, as well as for many practical applications such as drug discovery, genetic diagnosis, and environmental analysis. Machine learning is a powerful approach to solve sequence alignment problems over large databases. It involves training a model on a set of labeled sequences to learn the patterns and relationships in the data and then using the model to predict alignments for new, unseen sequences. Machine learning algorithms can efficiently handle large amounts of data and can learn to recognize complex patterns and relationships that may be difficult to identify using traditional alignment methods.

Moreover, machine learning approaches can be combined with other techniques, such as dynamic programming, to further improve their performance and accuracy. For example, some machine learning methods can use dynamic programming algorithms to align two sequences and then use the resulting alignments as training data for a classifier. This hybrid approach can achieve better accuracy than either method alone.

The scope of sequence alignment is broad and includes many different areas of biology and bioinformatics.some of them are Evolutionary studies, Functional annotation of genes, Drug discovery, Diagnosis of genetic disorders.

**Objectives**

The objectives of this project are listed as follows:

>Study on existing technologies to find out the draw backs and limitations.

>Collection of DNA/RNA Sequence alignment datasets.

>To develop a Machine learning model using ensemble methods, that increases the accuracy in sequence alignment.

>To train the model using a large dataset of biological sequences, and optimize it for enhancing the performance of the model.

>To evaluate the performance of the model on a test dataset, and compare it with the existing methods.