**Executive Summary:**

The project on Protein Function Prediction (Differentiating Kinases for Targeted Drug Discovery) addresses a critical challenge in drug development by distinguishing between kinases present in prokaryotes, precisely probiotic strains, and those in eukaryotes or pathogenic bacteria. Kinases, a protein essential for cellular signaling, are promising targets for drug discovery due to their involvement in various biological processes. However, the lack of differentiation between kinases across different organisms hinders the development of specific drugs for eukaryotic diseases and infections caused by pathogens.

The primary objective of this project is to develop robust models that can accurately predict the functions of proteins based on their sequences. We collected protein sequences from Kaggle.com (CAFA 5 Protein Function Prediction) to use as valuable training data for the models.

The methodology employed in this project involves several key steps.

* Firstly, the collected protein sequences are preprocessed and feature-engineered to extract relevant information from the data.
* Next, state-of-the-art machine learning techniques, such as deep learning algorithms, are employed to train the models using the protein sequence data.
* The models are fine-tuned and optimized to improve their performance in predicting protein functions.
* Cross-validation and benchmarking against existing protein function prediction datasets are conducted to assess the accuracy and generalizability of the models.

What we have for now:

**Future steps:**

To comprehensively understand the distinctive characteristics and kinase variations among prokaryotes, eukaryotes, and pathogenic bacteria.

**Comparative proteomic study:** To compare kinase sequences and regulatory elements, enabling the identification of distinctive patterns.

**Structural biology techniques:**  We will examine kinase protein structures, including active sites and binding pockets.

**Kinase profiling:** To analyze known kinases in prokaryotic strains, eukaryotes, and pathogenic bacteria to identify shared and unique features.

In conclusion, the Differentiating Kinases for Targeted Drug Discovery project represents a promising opportunity to advance drug development. By establishing strategies to differentiate kinases in prokaryotes from those in eukaryotes or pathogenic bacteria, the project opens doors for designing particular drugs for eukaryotic diseases and infections caused by pathogens. The resulting advancements in targeted drug discovery will have far-reaching implications, including personalized therapies, improved.