Title Statement:

AI-Powered Drug Repurposing: Predicting Drug-Protein Interactions for Novel Therapeutic Applications

Introduction

Tired of textbook problems and theoretical scenarios? This hackathon is your ticket to the thrilling frontier of real-world drug discovery! Imagine yourself as a medical detective, armed with the power of AI. Your mission: decode the complex language of drug-protein interactions (**DPIs**) – the secret handshake that determines how drugs work within the human body.

This isn't just another coding challenge; it's your opportunity to tackle a problem with real-world implications. By predicting DPIs, you'll be uncovering hidden pathways for drug repurposing, potentially leading to groundbreaking new treatments for diseases that currently lack effective therapies. Think of it as a scavenger hunt through the vast landscape of biomedical data, where your analytical skills and innovative thinking could unlock the next medical breakthrough.

Are you ready to dive into the heart of modern medicine, where data science meets biology? Can you decipher the complex code of drug-protein interactions and rewrite the future of healthcare? If so, this hackathon is your chance to prove it!

The Challenge

Your mission is to develop the most accurate predictive model for drug-protein interactions using the provided dataset and any machine learning or data science techniques. The winning teams will be determined based on the performance of their models on a hidden test set.

Dataset

A curated dataset in tabular format will be provided, containing diverse features related to drugs and proteins. The dataset will be split into training, validation, and test sets.

Dataset Structure:

Column Name	Description	Data Type	Example
Drug_ID	Unique identifier for the drug (e.g., DrugBank ID, PubChem CID)	String	DB00532, CID2244
Drug_Name	Name of the drug	String	Aspirin

Chemical_Structure	SMILES string representing the chemical structure of the drug	String	CC(=O)Oc1ccccc1C(=O)O
Target_Protein_ID	Unique identifier for the target protein (e.g., UniProt ID)	String	P05067
Target_Protein_Name	Name of the target protein	String	Cyclooxygenase-1
Target_Gene	Gene symbol(s) associated with the target protein	String	PTGS1
Interaction	Binary label (1 for interaction, 0 for no interaction)	Integer	1
Drug_Molecular_Weight	Molecular weight of the drug	Float	180.1532
Protein_Sequence	Amino acid sequence of the protein	String	MSKGEELFTGVVPILVELDGDVNGHK...

Hackathon Goals

1. **Develop High-Performing Predictive Models:** Create machine learning models that accurately predict drug-protein interactions.
2. **Explore Innovative Approaches:** Experiment with various algorithms, feature engineering techniques, and model architectures.
3. **Foster Collaboration and Learning:** Share knowledge, exchange ideas, and learn from fellow participants.
4. **Identify Potential Drug Repurposing Opportunities**: Contribute to the discovery of new therapeutic applications for existing drugs.

Evaluation Criteria

- **Model Performance:** Accuracy, precision, recall, F1-score, and AUC-ROC on the hidden test set.
- **Innovation:** Novel approaches, feature engineering, or model architectures.
- **Clarity and Reproducibility:** Clear documentation and well-structured code.