UQ-Bio^{23} Drug Discovery Challenge

Wednesday, May 31st 2023

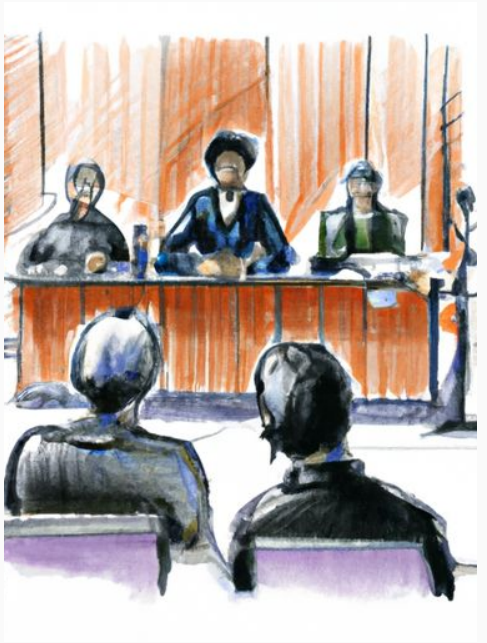
Disclaimer

The following exercise is provided solely for educational purposes and should not be considered real or implemented in any practical scenario. The instructions and information provided are intended for training and learning purposes only. *Any resemblance to real-world applications, scenarios, or outcomes is purely coincidental.* It is essential to exercise caution and seek appropriate professional advice or guidance when dealing with real-life situations or making decisions. The responsibility for any actions taken based on the content of this exercise lies solely with the individual. Any references to monetary transactions, rewards, or financial outcomes are purely hypothetical and should not be misconstrued as real or applicable to any actual financial situation.

ELM-THERAPEUTICS CEO INDICTED ON FRAUD CHARGES

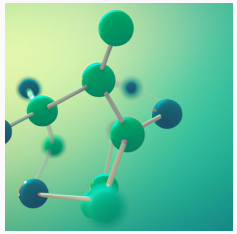
ELM-THERAPUTICS

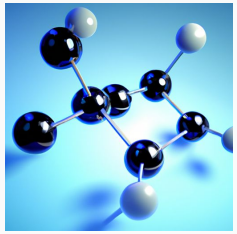
Pharmaceutical company Elm-Therapeutics finds itself in dire straits. The company's stock has been plummeting since the previous year, *following a series of failed drug candidates that incurred substantial financial losses*, and the CEO has been forced into retirement and faces prison after shoddy statistical analyses led to accusations that they had been falsifying data and hiding it from the public and shareholders.

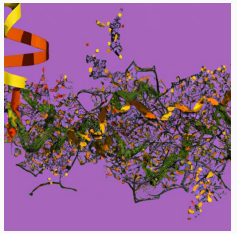


ELM-THERAPEUTICS THROWS HAT IN THE RING FOR THE HUNT TO CURE GIFFERSTRÖMM'S SYNDROME

Taking their last bit of funding, ELM-Therapeutics decides to attempt the creation of a novel drug to treat the newly described genetic disease, Gifferströmm's Syndrome. The disease is characterized by a *production of a mutated cytoplasmic protein that eventually builds to cytotoxic levels and causes damage across various tissues.*







ELM-THERAPEUTICS USES ML FOR DRUG DISCOVERY

Using advanced LLM-based machine learning for chemical discovery, the company has narrowed their search to *6 drug candidates*. The company performs human derived cell studies in house with these drug candidates and automated cell preparation and superresolution microscopy.

Compound	Tested efficiency	Mechanism of action
TDZ1098	Unknown	Unknown
ABT1564	Unknown	Unknown
ANX5511	Unknown	Unknown
KOS1273	Unknown	Unknown
OGX1900	Unknown	Unknown
TMCA187	Unknown	Unknown

ELM-THERAPEUTICS CONTRACTS PRESTIGIOUS UQ-BIO SUMMER SCHOOL IN FINAL DRUG GAMBIT

Oh no! ELM-Therapeutics cut their statistics modeling team in Q4 2022 to save money (thus causing the dramatic downfall of their reputation and CEO)! *As such, they have to contract out the statistics work on their data to determine which of the 6 drugs are the most promising.*

ELM-Theraputics CTO, Sam Jimmerson, was a previous UQ-Bio participant. Having build strong community ties through participation in the UQ-Bio Career Panel discussions, they reach out and ask if the *venerated summer school would help save their company*, and offer to hire the best performing group to replace the board of this struggling multi-trillion dollar company (dental is included).

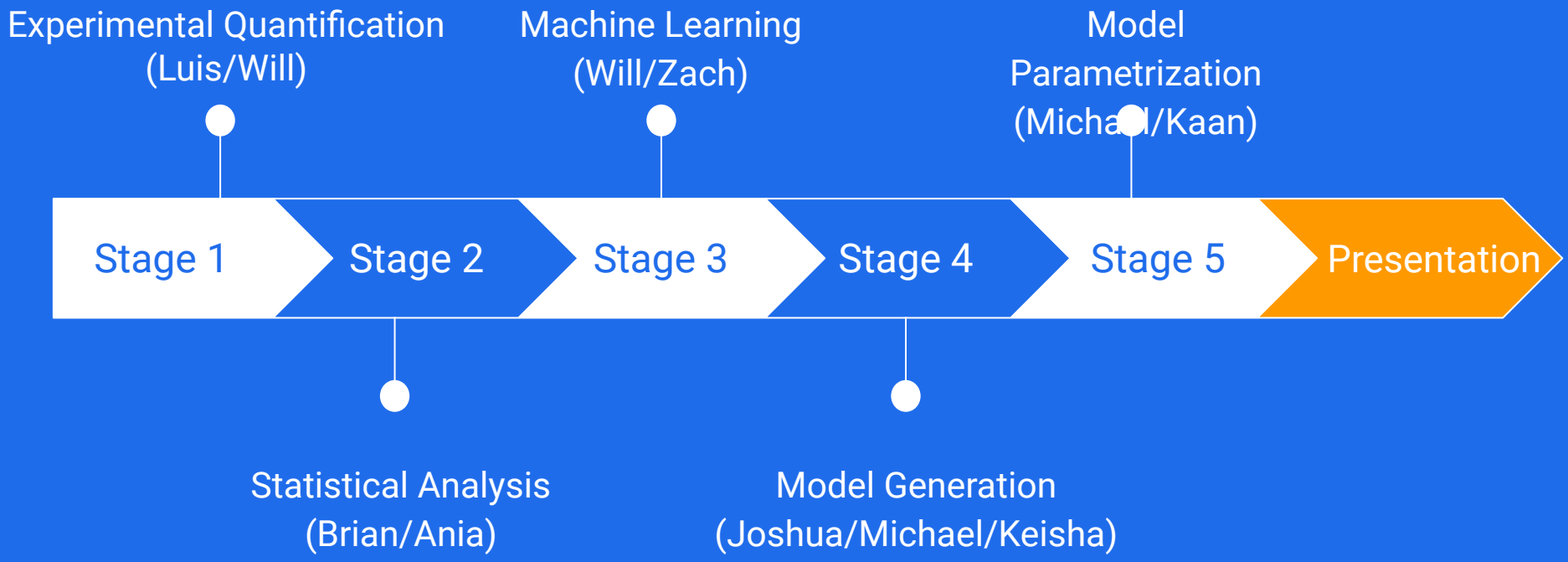
UQ-Bio^{23} Drug Discovery Challenge

The challenge requires the combined efforts of several diverse and skilled teams, each of which will draw upon expertise from various scientific disciplines. Armed with their collective knowledge and cutting-edge resources and their enthusiasm for collaboration, these research teams will embark on an intensive journey to unravel the mysteries surrounding each compound. *Their ultimate objective is to decipher the intricate mechanisms underlying the drugs' efficacy*, meticulously analyze the drug's impact on the genetic disease, and ensure the safety of the drug for potential use in patients.

Rules of the challenge

- ***Teams:*** 6 teams consisting of 5-6 scientists and one company mentor.
- ***Tasks:*** Each team will participate in five tasks: experiment quantification, statistical analysis, machine learning, model generation, and model parametrization.
- ***Time:*** The challenge last 2 weeks, with final presentations due Tuesday, June 13 at 2:00pm.
- ***Expected Results:*** Teams will present their final results for the analysis of microscopy data, their selection of the final drug, and their efforts to determine and model its mechanistic actions.
- ***Final Evaluation:*** The team that gives the clearest and most convincing presentation (including reproducible codes and clear statistical analyses) will be selected by a panel of the company mentors.

Challenge Stages



Stage 1: Experimental Quantification

Instructors: Luis/Will

Materials: Simulated microscopy images

Task: Create a streamlined image processing pipeline to segment cells/nuclei and quantify protein and rna expression in simulated videos.

Expected outputs: (1) CSV file with processed data and (2) streamlined function to generate these data from image sets. Data will include: cell size, nucleus size, protein concentration (AU), number of mRNA in nucleus/cytoplasm, number and intensity of transcription sites.

The team that produces the fastest and most accurate calculations of cell properties will be awarded the title of "Image Analysis Experts". Their codes will be made available to all other teams for subsequent tasks.

Stage 2: Statistical Analysis

Instructor: Brian/Ania

Materials: CSV dataset from image processing for all drug treatments and all time points

Tasks: (1) Visualize and analyze statistics for cell size, cell vitality, protein, and nuclear/cytoplasmic rna using barplots, histograms, and summary statistics. Plot trends of measured properties versus time. Calculate correlations between features. (2) Compare data sets to one another and compute statistical significance of drug treatment effects. (3) Create a streamlined computational pipeline to perform similar analyses on any appropriately formatted data set.

Expected output: (1) Presentation of data and comparison of analyzed genes. (2) Evaluation of assigned drugs in terms of their efficacy (fraction of repressed protein) and toxicity (fraction of cell death). Based on all teams analyses, the entire group will vote to select up to three drugs to continue testing in the next stage.

The team that produces the clearest and most statistically sound and complete analyses of cell properties will be awarded the title of "Communications Experts", and their codes will be made available to all other teams.

Stage 3: Machine Learning

Instructors: Will/Zach

Materials: Finalized image analysis pipeline (stage 1) that generates CSV data set (stage 2). New data sets (images and processed CSV files) for drug titration experiment (training data) and three held out data sets with unlabeled drug concentrations (testing data).

Task: (1) Perform regression analysis and machine learning to build a statistical model that estimates drug efficacy, cell vitality, and other cell response features as functions of drug concentrations. (2) Make statistical predictions for the held out data sets with unlabeled drug concentrations. (3) Use model to estimate the drug concentration that maximizes the reward metric for efficacy and minimizes toxicity: Reward = (cell survival fraction) / (protein expression).

Expected output: (1) Statistical model to relate cell features to drug concentrations. (2) Quantitative predictions for drug concentrations in the three held out samples. (3) Prediction for what drug concentration maximizes the value of Reward. (4) A streamlined notebook performing all regression analyses.

The team that produces the most accurate estimates of the drug concentrations in the three held out data sets will win the title of "Artificial Intelligence Experts", and their codes will be made available to all other teams

Stage 4: Model Generation

Instructors: Joshua/Michael/Keisha

Materials: Various hypotheses for the mechanism of the drug

Task: (1) Propose stoichiometry vectors and propensity functions for two or more hypotheses that could reproduce the observed statistics of the drug's (e.g., changing protein degradation rates). (2) Simulate the dynamics of these models before and after application of drug using ODE and SSA models. (3) Tune parameters by hand to reproduce observed statistics from stages 0-2.

Expected output: (1) Clear presentation of different candidate models, explanation of their various mechanisms, and direct comparison of statistics of simulated trajectories and actual cell responses for each models. (2) A streamlined notebook that generates and solves the proposed models.

The team that produces the most efficient and clearly presented simulation notebook will win the title "Cell Simulation Experts", and their codes and models will be made available to all other teams.

Stage 5: Model Parametrization

Instructors: Michael/Kaan

Materials: Chosen Model Hypothesis (Stage 3) + Dataset for the selected drug in various conditions (Stages 0-2).

Task: (1) Propose a simplified model to capture the statistics of your team's chosen hypothesis for gene expression and drug effect and solve this model using the FSP approach. (2) Compute the likelihood of data given the model and an arbitrary parameter set. (3) Manually or automatically adjust parameters to maximize the FSP MLE analyses to estimate model parameters given available data. (4-optional) Quantify parameter uncertainties given existing data using parameter sweeps or MCMC. (5) Use model to predict response under three new drug treatment regimens that have not previously been tested (provided by company CEO).

Expected output: (1) Parametrized model and FSP analysis to predict response over time. (2) Streamlined notebook to analyze model and search for MLE parameters. (3) Predictions of drug response and cell vitality versus time for required drug treatment regimen.

The Stage 5 Challenge will be evaluated alongside the entire project on Tuesday June 12.

Final Stage: Project Presentation

On Tuesday, June 13, 2023 at 2:00 pm, present your results, *pitch your drug selection*, and explain your team's approach. to a group of investors that will determine the winner team, based on the following factors:

- Presentation quality
- Reproducibility of approach
- Statistical rigour
- Team's ability to respond to technical questions.

end