2 Project Plan

2.1 Task Decomposition

1) Individual knowledge and understanding of ML models specifically Neural Networks
    a. Logistic Regressions, Linear algebra, Derivatives, Matrix calculus, probability, graph theory, and Python
2) Set up processing environment
    a. Setting up AWS and GCS
    b. Setting up necessary tools and libraries
    c. Importing data set
3) Possible Data preprocessing (might not be needed)
    a. Normalizing/standardizing data set
    b. Removing null values
    c. Determine vector or array for input layer
    d. Split data set into training and test sets
4) Actual training of the model
    a. Finding Neural Network to retrain
    b. Constant retraining till satisfactory
        i. Changing Depth
        ii. Changing Activation Function
        iii. Looking at Test MSE, Bias, Variance
    c. Final Model selection
5) Model interpretability
    a. Looking at heat map of weights
    b. Dimensional reduction
    c. Further tools/tests
6) Building user interfaces
    a. Login screen with authentication
    b. Easy to use or interpret displays or screens for results.

2.2 Project Management/Tracking Procedures

We are planning on using agile because members of the group have used that style in previous classes/internships, so we feel the most comfortable using this style. We think agile will work best with our project because of its cycles. It is not linear like the waterfall style. The waterfall style could make us feel constricted and feel like we are veering off-course if we must go back to

previous stages. Agile allows more flexibility and goes more for one feature at a time, which is less constricting and helps us not feel like we are behind if we need to go back to a previous step. We also feel sprints will help us achieve our goals easier and within a realistic timeframe. The daily scrums will also keep all members accountable for completing tasks.

We are planning on using Git and Github to track and share our progress with the group and professors. We all have experience with Git and Github, so we are the most comfortable using those tools. We also think it will be easier to work on the design/code of the project simultaneously using tools included with Github like push, pull, and merge. We can track our progress through issues and milestones.

2.3 Project Proposed Milestones, Metrics, and Evaluation Criteria

Milestone 1: Learn about neural-networks

Milestone 2: Set up Processing environment with 100% usability

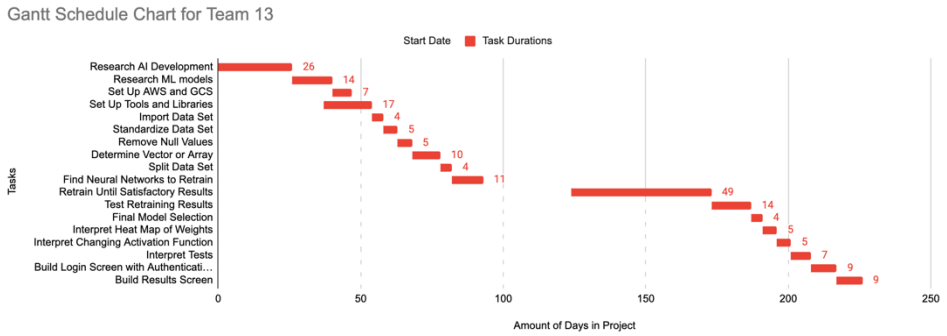
Milestone 3: Data Preprocessing – 95% of dataset usable and standardized (mean is at 0)

Milestone 4: Training of the Model till accuracy reaches 80%

Milestone 5: Model Interpretability – 70% interpretability of the model (Not only the weights and biases of the model)

Milestone 6: Building User Interface – 100% reliability

2.4 Project Timeline/Schedule



Research AI Development	9/13/2023	10/9/23
Research ML models for Neural Networks	10/9/2023	10/23/23
Set Up AWS and GCS	10/23/2023	10/30/2023
Set Up Tools and Libraries	10/20/2023	11/6/2023

Import Data Set	11/6/2023	11/10/2023
Standardize Data Set	11/10/2023	11/15/2023
Remove Null Values	11/15/2023	11/20/2023
Determine Vector or Array for Input Layer	11/20/2023	11/30/2023
Split Data Set Into Training and Test Sets	11/30/2023	12/4/2023
Find Neural Networks to Retrain	12/4/2023	12/15/2023
Retrain Until Satisfactory Results	1/15/2024	3/4/2024
Test Retraining Results	3/4/2024	3/18/2024
Final Model Selection	3/18/2024	3/22/2024
Interpret Heat Map of Weights	3/22/2024	3/27/2024
Interpret Changing Activation Function	3/27/2024	4/1/2024
Interpret Tests	4/1/2024	4/8/2024
Build Login Screen with Authentication	4/8/2024	4/17/2024
Build Easy to Use Screen that Displays Results	4/17/2024	4/26/2024

2.5 Risks And Risk Management/Mitigation

- Not able to get access to dataset | Minimal risk | 0.2
    - Would have to try and simulate a medical dataset either using examples and resources from online source.
- Not able to use AWS or GCP | No real risks | 0.0
    - Would have to get dedicated processing. We'd need to investigate purchasing GPUs for processing.
- Not able to use an existing Neural Network for retraining | Minimal risk | 0.2
    - We'd have to create our own Neural Network model.
    - Or switch to a different ML model.

2.6 Personnel Effort Requirements

Task	Average Time Per Person (hours)

Individual knowledge and understanding of ML models specifically Neural Networks	12 – This is based on the team being new to neural networks, along with background subjects/material. This includes Logistic Regressions, Linear algebra, Matrix calculus, and probability.
Set up processing environment.	2 – This will mostly be setting up AWS, GCP and the framework library for neural networks.
Possible Data preprocessing (might not be needed)	10 – This is a rough estimate. More analysis of the dataset would be needed to determine an actual timeline, but we haven't gotten the dataset yet.
Actual training of the model	14 – This will also be dependent on the data and how well it will translate into our model.
Model interpretability	10 – This will be dependent on the data and the model we end up using as more depth will make the model harder to interpret.
Building user interfaces	4- I think this is a good estimate for a barebones interface and more time can be added if available.

2.7 Other Resource Requirements

Medical dataset of allergens from mayo clinic

Neural Network for retraining

AWS and GCP access/accounts