**Project Plan**

The goal of the project is to compare Bayesian deep learning with a deep neural network. We will accomplish this by using data from All of Us research to try and predict heart disease, arthritis, diabetes, and high blood pressure in patients. This will be accomplished by implementing both a basic deep neural network as well as a similarly-constructed Bayesian deep neural network and training them on the same dataset and comparing the results. Multiple experiments will allow for a demonstration of how the models compare at predicting real-world results and why one model might work more favorably than the other.

**Timeline**

Week February 14 - February 18:

This is the first week of the project. The goal of this week is to get an account ready from All of Us research as well as looking at how everything is set up on All of Us. This will take the full week as All of Us requires you to review their training, submit proof of identification, and take a quiz at the end. Since we are Ph.D. students at UNT we can apply and get access to All of Us research for free by using UNT credentials.

* Create accounts for All of Us.
* Proof of identity for All of Us.
* Finish All of Us training.
* Pass the test to get access to All of Us.

Week February 21 - February 25:

For the second week we will create a cohort on All of Us that will pull tables we might use for the project. This cohort will pull from patients that have taken the surveys (overall health, family history, lifestyle, and personal medical history) and patients that have a condition including heart disease, diabetes type 1 or type 2, arthritis (including Rheumatoid and ​​Osteo), or high blood pressure (hypertension). The cohort will also make sure the patients are adults (21+ years old). We will also take a look at the demographics surveys to use for visualization purposes mostly to get an idea on the type of patients and spread we have. Then we will take the cohort and create a workspace. In this workspace we will be able to write our code and analyze the data.

* Create Cohort to include patients with conditions and have surveys.
* Select Data from Cohort.
* Create Workspace to start writing code.

Week February 28 - March 4:

Reformat data from the default bit query to data frame format. Data will need to be combined from different tables and formatted so that one row is one patient. The end goal for this week is to have finished at least a preliminary analysis of the data (statistics, format, information covered, limits, etc.). During this week we will also write the first progress report.

* Selecting columns.
* Feature formatting if necessary.
* Data formatting (e.g. combining all of a patient’s data into one row).

Week March 7 - March 11:

The fourth week will be data selection and preprocessing. The data will be modified as needed in the form of feature selection and engineering. Further data selection will occur if the dataset’s statistics are skewed enough to be a concern or if there are any needed cutoff points in the dataset, such as a date cutoff. We will also remove any question or columns that do ont pertain to our diseases. If it is unfinished, we will finish the first progress report.

* Feature selection.
* Feature engineering.
* Data selection such as dropping patients to change statistics, eliminating data based off of cutoff points, training/validation/testing splits, etc.
* Finalize a plan of which models to test.
* Start of creation of neural network(s) if time permits.
* First report due March 11th.

Week March 14 - March 19:

This week is spring break. We will use this week if necessary to catch up on any work we got behind on. This week will also be used in case of any surprises or extra steps that need to be taken into account regarding data preparation and analysis. This week can be used to rinse and repeat week 2-4 if we need to include more or less data in our workspaces. We will have code written that we will be able to move to new workspaces with updated data and then update code as needed.

Week March 21 - March 25:

This week will be the start of the construction of a basic deep neural network. By the end of the week the model should be at least halfway constructed, preferably ¾ of the way constructed. We will start the writing of the background work for the final paper during this week.

* Start construction of the deep neural network.
* Work on parameters of the deep neural network.
* Begin writing the background of the final paper.

Week March 28 - April 1:

By the end of this week the basic deep neural network should be fully constructed and providing initial results to act as a baseline for the Bayesian model. This week should also include finishing the background work on the final paper.

* Complete construction of the deep neural network.
* Have the initial hyperparameters and model size selected.
* Generate initial results for the deep neural network.
* Have background research for the final paper completed.

Week April 4 - April 8:

During this week we will be writing the second progress report. We hope to have finished the basic deep neural network baseline and provide results for that model. We will be constructing the Bayesian Model this week and start working on hyperparameters for it. Construction of the Bayesian deep neural network will be started.

* Have initial hyperparameters for Bayesian model construction.
* Generate initial results for the Bayesian deep learning model

Week April 11 - April 15:

For this week, we will be finishing up the Bayesian model and have results for it. Finish testing hyperparameters for the model to find the best set. Start writing the methods portion of the final paper and use results from the deep neural network to start the results portion of the final paper. Finish writing the second progress report with detail from initial results from the Bayesian model.

* Finish construction of the Bayesian Model.
* Generate initial results for the Bayesian model.
* Have the final paper’s section on the deep neural network completed.
* Start writing methods portion of final paper.
* Start writing the results portion of the final paper.
* Second progress report due on April 20th.

Week April 18 - April 22:

During this week further results will be generated by experimenting with different feature selections, hyperparameter changes, changes to the model size, etc. This week will generate the final results of all experiments.

* Run further experiments on both models with changes to:
  + Feature selection
  + Hyperparameter changes
  + Different model sizes
* Analyze the generated results.
* Have at least half of the final paper’s section on the Bayesian deep neural network completed.

Week April 25 - April 29:

Any last minute tuning will be added in this week. Then we will be finishing the paper and submitting it is the goal of this week. We will be starting on the final presentation.

* Finish writing the final paper.
* Finish edits to the final paper.
* Start on the final presentation.

Week May 2 - May 4:

This is the final week of class and the presentation is due. We will add any last touches to the presentation. Practice the presentation and submit the presentation by the end of day on May 3. We will present the final presentation of the project in class on May 4. If it is unfinished, we will finish the final paper.

* Complete the final presentation.
* Present the final presentation
* Presentation is due on May 4th
* The final paper is due on May 9th.