Capstone Project Proposal

Project Name: Insulin Calculator

Student: Maggie Dube

Area of Concentration: Data Science

Final Product: Machine Learning Algorithm for calculating Insulin intake

1 PROJECT BACKGROUND

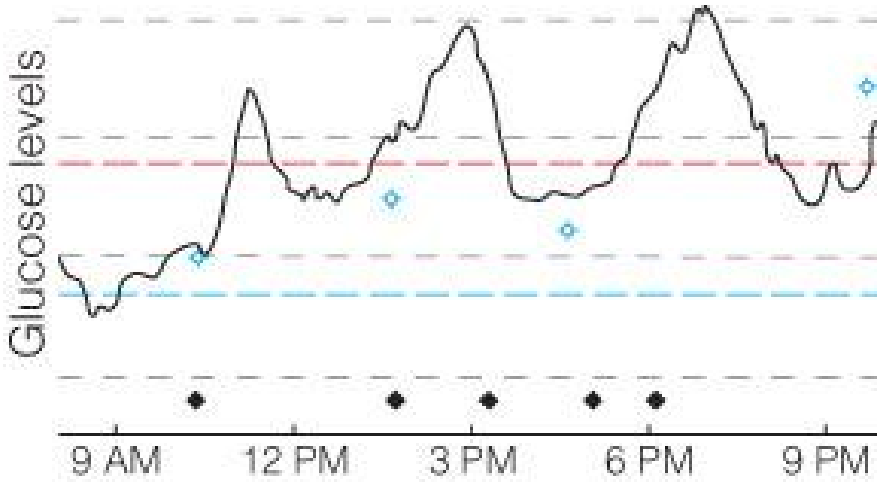
Ever since I wrote my essay for my Center for Information Technology application in eighth grade– which included a description of how I want to use technology to help improve my Diabetic mother's health - I've wanted to improve my mother's Diabetic regimen. For almost 10 years, my mother has been struggling with varying levels of glucose that aren't predictable. My mother, during the course of one day, will have multiple hyperglycemic or hypoglycemic episodes; this means that she has either very low glucose or very high glucose. Both hypoglycemia and hyperglycemia are deadly, so making her glucose levels more stable is the most important problem that needs to be solved.

This project will entail a machine learning algorithm that predicts the amount of insulin a person should inject given their HbA1c, glucose level, carb intake, and BMI. This project is important for me not only because it could help balance my mother's glucose levels, but it will also help other diabetics balance their glucose levels as well. Undertaking this problem, given that I am in the Data Science area of Capstone, will help me explore and learn about Machine Learning as well as more detailed areas of Data Science, including Data Extraction and Data Visualization.

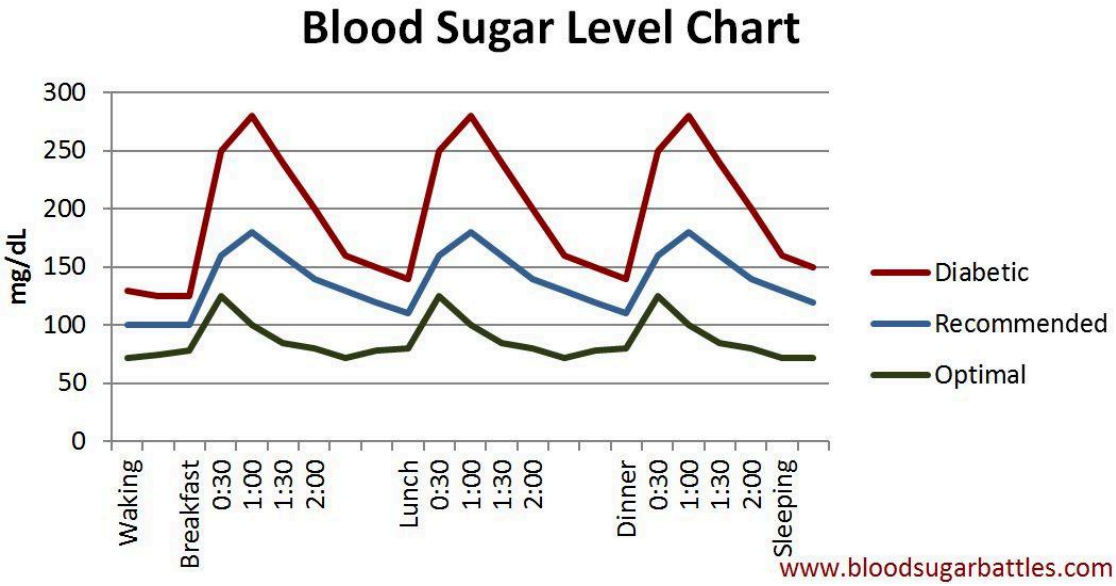
2 PROJECT SCOPE

At the start of the project, I will begin research on Machine Learning and how it works. The subtopics of the research will include Machine Learning types, applications, history, steps, and tools. The research will allow me to begin collecting specific data for my Machine Learning algorithm; show me what tools I can use to define, train, and test the algorithm; and create the insulin-calculating algorithm. The end product of this product should include a machine learning algorithm and a functioning web app.

The algorithm should attempt to assume the amount of insulin a diabetic should inject so that their glucose (Figure 1) levels out, like the glucose of a nondiabetic (Figure 2).



(Figure 1: A chart of a diabetic's average glucose during the day. The red line shows the highest a normal glucose should be and the green line shows the lowest a normal glucose should be. Source: https://www.diabetesselfmanagement.com/diabetes-resources/tools-tech/continuous-glucose-monitoring-making-sense-of-your-numbers/)



(Figure 2: This is a chart mapping the average diabetic glucose versus the recommended and optimal glucose during the day. Source: http://www.bloodsugarbattles.com/blood-sugar-chart.html)

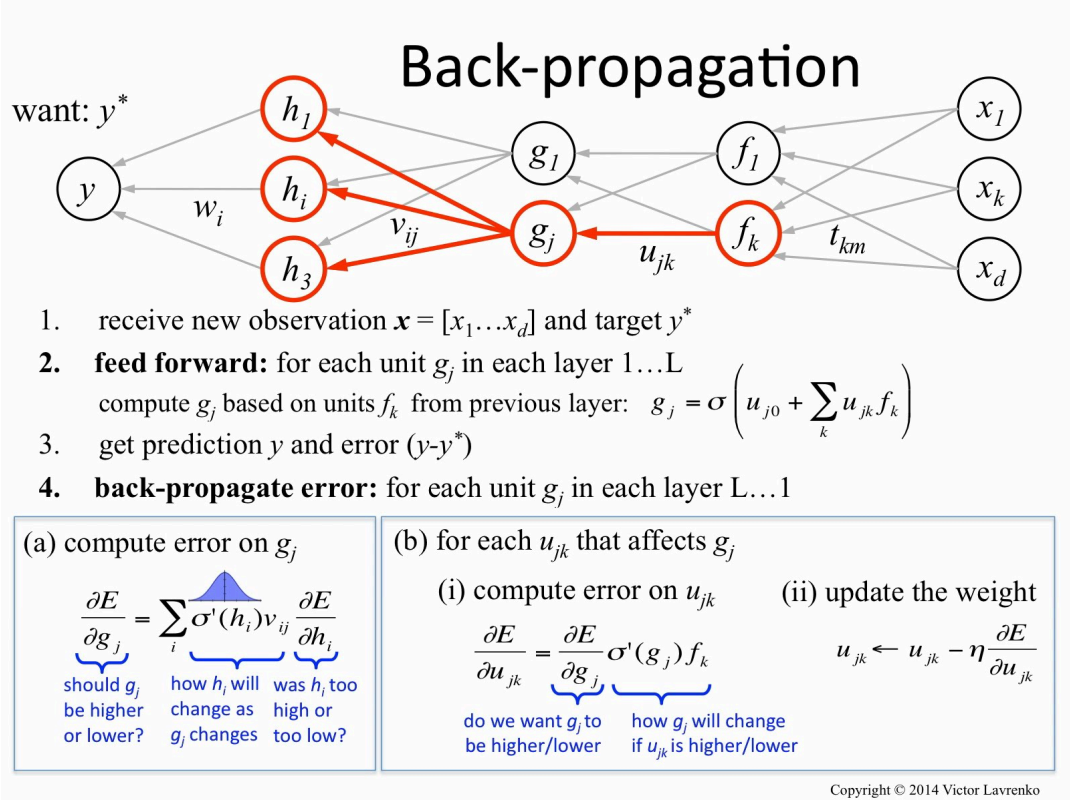
After the research, I will move into collecting and validating the diabetic data my mother has stored in applications and notebooks. The data will include three entire months of my mother's diabetic data from October 2018 to December 2018. The diabetic-related data will include her HbA1cs and BMI for each month as well as her glucose level, insulin intake, and carbohydrate intake for each day – morning, afternoon, and night. They will all be inputted onto an Excel spreadsheet. They will be validated by checking for correctness (ex. If the data is supposed to be a whole number and it is not, a SQL query can fix it), meaningfulness (ex. Will it benefit the project as a whole and provide the algorithm with sufficient data to make an accurate

assumption), and security (ex. Making sure the data cannot be tracked to the specific person – i.e. my mother.) All the data used for this algorithm (HbA1c, BMI, glucose, carbohydrate intake, and insulin intake) help a diabetic calculate their insulin.

The machine learning algorithm code and the web app code will be stored on GitHub so that it is easy for both my mentor, Mr. Neville, and I to share and access the code. The machine learning algorithm code will be mostly in Python, and the Web App code will be in Python, HTML, CSS, and other web design languages if needed. GitHub tracks changes and repairs, and they're called commits. These commits will determine how much effort and time I spend working on and testing the machine learning algorithm and the web app. On average, I expect to have an average of 3-4 commits each week, five average commits being my goal for each week.

While defining the machine learning algorithm for calculating the amount of insulin, I hope to confine the amount of inputs an end user makes. By doing this, it will be less of a hassle on the user. I will have to train the algorithm to output correctly in order to limit the amount of inputs a user must have.

When training the algorithm, I have to determine how much different the desired result is to the output. In order to fix the output to make it more accurate (like the desired result), I must do back propagation by gradient descent (Figure 3.) Back propagation is a process of taking the original output of the algorithm and going to each layer, starting with the last, and fixing the equations to adjust the output slightly until the output reaches the desired result. Once I have a number closer to the desired result, I have to calculate how much the error has been minimized, rather the root mean squared error (Figure 4), then make the calculation a percentage.



(Figure 3: This graphic shows how back propagation by gradient descent is computed.. Source: https://www.youtube.com/watch?v=An5z8lR8asY)

5. *RMSRE*: root mean squared relative error [48,49]

$$\mathrm{RMSRE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{H_d^{i,m} - H_d^{i,c}}{H_d^{i,m}}\right)^2} \quad (59)$$

6. *RRMSE*: relative root mean square error. This indicator is calculated by dividing RMSE with average value of measured data. According to [50], model accuracy is considered excellent when RRMSE $< 10\%$, good if $10\% <$ RRMSE $< 20\%$, fair if $20\% <$ RRMSE $< 30\%$, and poor if RRMSE $> 30\%$.

$$\mathrm{RRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(H_d^{i,m} - H_d^{i,c}\right)^2}}{\sum_{i=1}^{n} H_d^{i,m}} \times 100 \quad (60)$$

(Figure 4: This image shows what RRMSE is, how it is calculated, and what the percentages mean. Source: https://stats.stackexchange.com/questions/260615/what-is-the-difference-between-rrmse-and-rmsre)

After the machine learning algorithm is done, I will start making the web app. My goal is to have a web app on the internet that you can input into, which means the user, when going to the website, can just input into something and receive some type of output. If I am able to add more to the web app, I will allow the user to access the machine learning algorithm, input into it, and receive output.

The end product will be of use to my mother and other diabetics. It should be used as an unofficial way of calculating the amount of insulin a diabetic should intake. The code will include a message to the users stating that the program was not made by a medical professional and shouldn't be the only thing used to calculate the amount of insulin one should give themselves. Hopefully, in the future, this could be tested by actual professionals, like those who do research at the American Diabetes Association and used as a tool to help diabetics accurately calculate how much insulin they should inject themselves with.

American Diabetes Association®

(Source: http://main.diabetes.org/site/Calendar?id=34721&view=Detail)

2.1 Goals and Objectives

Goals	Objectives
Relearn Python and start course on Machine Learning at Codecademy	1. Completely finish the Python module on Codecademy 2. Complete Machine Learning module on Codecademy to start learning about the frameworks and libraries
Research	1. Complete project based on Research 2. Research pertains to Machine Learning - how it works, its history, tools for it, and its applications 3. 30 or more sources used to research 4. Research will help me learn about Machine Learning
Worked 100 hours on project	1. Work 100 or more hours on research, data collection and validation, the machine learning algorithm, and the web application.
Worked with Mentor on the Project	1. Communicate with mentor at least once a week or biweekly 2. One in-person meeting every month with mentor 3. Mentor helps when needed
Collected and Validated Diabetic Data	1. Collect three months of mother's diabetic data 2. Data from the last three months of research (October, November, and December of 2018) 3. Data includes HBa1c and BMI of each month 4. Data includes Glucose level, Insulin intake, and Carbohydrate intake) 5. All data in an Excel spreadsheet 6. All data has been validated
Minimized Error	1. Calculate minimized error of output while training machine learning algorithm 2. Find the root mean squared error of the output 3. Determine the percentage of the root mean squared error
Number of inputs into the machine learning model	1. Start with six inputs by the user into the algorithm, including HbA1c, glucose, BMI, carb intake, average insulin intake, and time of day) 2. Go to three inputs - glucose, carb intake, and the time of day
Make web app	1. The web app is running on the internet 2. User can input data and receive output
Use GitHub	1. 3-4 commits on average 2. Use GitHub to code machine learning algorithm and web app

2.2 Organizational Impacts

Organization	Impact to and Participation of Organization
American Diabetes Association	Project will aid in their research to making glucose levels stable.
My Family	My mother is a diabetic, so the product would help her calculate a guess for her insulin intake.

2.3 Project Deliverables

Project Milestone	Date Estimate	Deliverable(s) Included
Research Paper Outline	10/26/18	- Synthesis matrix - Outline of paper - 15 or more sources
Research Paper Progress Check	11/12/18	- Half of the paper is done - Research includes sources used in synthesis matrix
Research Paper Due	12/3/18	- All parts of the paper have been completed - 30+ sources cited - 12-15 pages
Mentor Evaluation Checkpoint 1	12/10/18	- Mentor must complete evaluation to give to Mrs. Norris
Data Collected and Validated	1/1/19	- Mother's diabetic data is put into an excel spreadsheet - Data has been validated as specified in the project scope
Portfolio Check	2/15/19	- Most portfolio requirements are met
Progress Report PowerPoint Presentation Due	3/1/19	- PowerPoint presentation with all requirements met is completed
Mentor Evaluation Checkpoint 2	3/4/19	- Mentor must complete second evaluation and give it to Mrs. Norris
Machine Learning Algorithm Training is Completed	3/15/19	- Machine Learning algorithm has been defined and tested in GitHub with mentor - Minimized Error has been calculated - average of 3-4 commits have been

		made on GitHub
Preliminary Project Portfolio Due	4/15/19	- Project Portfolio with all the required elements in it plus the code of the Machine Learning Algorithm
Proof of completed project and final portfolio due to Advisors	4/26/19 by 4PM	- All project parts are completed along with proof of communication from mentor and code from the machine learning algorithm and other parts that are coded.
Reflective Essay	4/29/19	- Reflective Essay is completed and turned into Mrs. Norris
Web App	4/30/19	- Web app is running on the internet so that users can input data into and receive output on.
Presentations to the Project Review Committee	5/14/19-5/24/19	- Completed presentation with all required elements
Mentor Evaluation Checkpoint 3	5/24/19	- Mentor must complete third evaluation and give it to Mrs. Norris
Sr. Capstone Project Extravaganza Night	5/29/19 from 6:30PM to 8:30PM	- Present at the extravaganza

3 PROJECT CONDITIONS

3.1 Project Assumptions

- The data will be valid for the project.
- The data is sufficiently consistent that it is modelable and predictable.
- The data that we are getting is free of errors and tracked consistently

3.2 Project Risks

#	Risk	Likelihood	Project Impact-Mitigation Plan
1	End Users take the product too seriously	Medium	There will be a message to the users that use the program stating that the program was made by someone who isn't a medical professional and shouldn't be used as an accurate insulin calculator.

3.3 Project Constraints

- Size of the data set
- Available computing power for model training

4 Project Structure Approach

This project will be done in Agile Project Management. It will include strict deadlines, communicating with my mentor often for tips and help. The project is very depended on learning about Machine Learning. The project will be planned by myself with the help of my mentor, Chris Neville.

- Research/Learning
- Validate data
- Look through sample datasets
- Practice making Machine Learning
- Start the Machine Learning Algorithm
- Project Presentation

5 MENTOR ASSISTANCE

- Programming help
- Resources for research, Machine Learning, and Data Science
- Proofreading and reviewing work
- Enforcing accountability
- Staying on Task
- Selecting and Identifying correct tools
- Managing Scope

6 APPROVALS

Prepared by _____ _____
Maggie Dube Date

Approved by _____ _____
Mr. Christopher Neville Date

_____ _____
Mrs. Lynne Norris Date