E-Health: Type 2 Diabetes Management Analytical Engine
Progress Report

Members:	Lama Yassin Kassab (100969142) Leen Yassin Kassab (100969141) Pooria Shafia (100759037) Joe Samuel (100998314) Mohamed Hozayen (100997960)	Supervisor: Dr. Mohamed Ibnkahla Co-supervisor: Dr. Zied Bouida

Project Objective

The project objective is to develop a machine learning algorithm which helps Type 2 Diabetes Mellitus patient to manage their disease by giving them real-time recommendation based on factors such as their current glucose level, their sleep data, exercise, stress and diet. To this end, we aim at developing a data analytics engine that allows for short-term glucose level predictions. Furthermore, finding correlations between the factors discussed above will be used in order to give recommendations to the patients.

The following table summarizes our tasks as outlined in the project proposal and our progress to date:

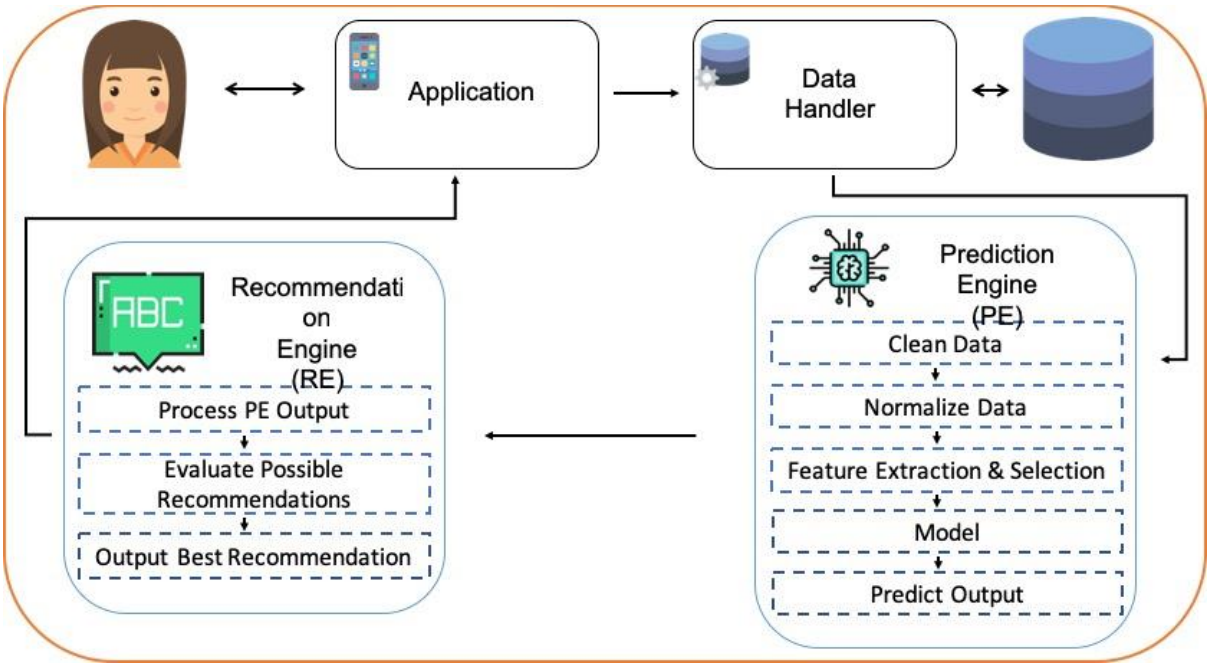
Milestone	**Proposed Due Date**	**Who's Responsible**	**Details**	**Progress**
Resource Acquisition	October 5	Leen, Lama, Pooria	Involves gathering/generation of datasets required to build machine learning models, acquiring necessary equipment.	
Initial Algorithm Development	October 13	Mohamed, Joe	Formulating the core algorithm logic for the data analytics engine.	
Development of System Architecture	December 3	Mohamed, Joe	Development of system architecture, detailing the infrastructure of the data analytics engine.	
Data Analytics Engine Development	December 3	Leen, Lama, Pooria, Mohamed, Joe	Developing a prototype for the data analytics engine based on the initial algorithm draft.	Work in Progress
Integration of Amazon Alexa with Data Analytics Engine	January 21	Leen, Lama, Pooria, Mohamed, Joe	Integrating the Data Analytics Engine with Amazon's Alexa voice assistant for voice-based data inputs.	Work in Progress
Testing – Verification and Validation	February 18	Leen, Lama, Pooria, Mohamed, Joe	Performing verification techniques and validating results through a variety of testing techniques.	Work in Progress

Detailed Progress Made to Date

Data Collection/ Generation

To address the first milestone of the project as outlined in the proposal that pertained to Diabetes Type 2 patients data acquisition, several institutions were contacted to request this data including Health Canada, Children's Hospital of Eastern Ontario (CHEO), The Ottawa Hospital Civic Campus, and Lastly Macadamian. Although requesting the data took up a long time of the project, none of the contacted institutions was able to provide any data. Additionally, when the Director of Emerging Technologies at Macadamian informed the team that his company has been working on a similar project and was unsuccessful in obtaining the desired data, it was decided that the data would have to be manually generated. For the data generation, literature reviews on glucose levels in healthy and diabetic patients were looked up and glucose data over a 24-hour period with 5-minute time interval for twelve cases was obtained from these reviews. After that, using MATLAB and the base cases created, 128 days of glucose data were generated while introducing random noise into the data and changing activity times. Lastly, SenseGen, an automated tool that generates a large dataset of data using a base dataset was used to generate glucose data for 1000 days using the 128 days of MATLAB-generated data. At this stage, the first milestone was achieved, and it was possible to start addressing the second milestone in the proposal which was the data analytics algorithm development.

System Architecture



The Data Analytics Engine ("System") is made up of five main components: Application through which the end-user will interface with the engine, a Data Handler to manage data transfer between the database and other components of the system, the database itself to store patient metrics and relevant prediction data, the Prediction Engine (PE) responsible for predicting glucose level based on given inputs, and the Recommendations Engine (RE) to produce a suitable recommendation based on output from the PE.

The Application will be responsible for getting input values and health metrics from the end-user for processing through the PE. The Application will be built using Node.js and Express.js for fluid user interaction and seamless user experience.

The Data Handler securely communicates with a NoSQL-based database built on MongoDB consisting of JavaScript Object Notation (JSON) files for each patient record. NoSQL was chosen thanks to its superior performance, scalable architecture, and its resolution concerning shortcomings of the relational model. Implementation of the Data Handler takes a secure approach with the separation of interfaces between the database and other components of the system to ensure no direct access is given to unauthorized software components.

The PE, responsible for processing input values to predict future glucose levels, harnesses machine learning libraries such as Python's Scikit to process the prediction outputs. The PE consists of three phases: Predictions, Correlations, Recommendations. The patient's glucose is input as a time series of glucose level gL(t) every 5 minutes. The goal of this phase is to predict glucose level at a point in time after t based on previous glucose level gL(t+tau), where tau is a multiple of 5. A simple linear regression prototype model is used to predict gL(t+1).

The RE processes the output produced from the PE and uses Triple Modular Redundancy (TMR) to retrieve the most suitable recommendation based on the data provided. This recommendation is relayed over to the end-user through the Application.

The architecture of the system was carefully crafted to ensure modularity and separation of concerns among its various components.

Appendix A – Revised Gantt Chart for the Project Progression