Application Of Machine Learning In Healthy Living Partner Mobile App

Machine Learning Engineering Project

Francis Odo

STATUS: First DRAFT

Overview

Costs of healthcare continue to rise at a significantly high rate in the United States with no limit in sight. More troubling is the cost of diagnosing chronic diseases. Healthcare providers, patients and insurance service providers are faced with this very challenging problem.

Healthcare Service Providers are exploring preventive care space to help their patient base increase physical activity and improve on healthy habits. While the overall goal is improved patient satisfaction and well being, the specific financial goal is reduced cost by emphasizing improved health prior to any adverse conditions developing. Overall, Healthcare Service Providers want to decrease spending on chronic diseases such as type 2 diabetes and related conditions.

This project provides the structure and analysis of a mobile app referred to as Healthy Living Partner that monitors, tracks and recommends healthy daily habits.

The approach adopted in this document is to propose a product solution and articulate the application of Machine Learning techniques using a classification model within a supervised learning architecture. The scope is limited to the prototype, but enough to demonstrate the concept and implementation.

Problem Statement

Chronic disease such as Type-2 Diabetes is affecting quality of lives significantly across age categories. It is a disease that causes damage to large blood vessels of the heart, brain and legs. Also, to small blood vessels, creating problems in the eyes, kidneys, feet and nerves (macrovascular complications). Due to the nature of the disease and how widely spread in the society, the cost of diagnosis and treatment is increasingly high. Healthcare and pharmaceutical industries are faced with uncontrollable costs, projecting 26% increase over a five year period.

The most viable and cost effective preventative approach to reducing and possibly eliminating this staggering cost problem is through controlled healthy habits. This can be achieved with a mobile app that monitors glucose level, tracks and recommend diets, along with appropriate exercise activities. The user gets to have total control of daily diet and exercise habits, which will enhance healthy lifestyle. Healthcare providers can save significantly on costs to diagnose and treat less severe cases.

Most available solutions in the market focus on boosting the will power and training human system on habits. The Healthy Living Partner(HLP) incorporates blood sugar-level monitoring, bundled with Activity Menu and API, thereby providing the user with a wide range of control of key components of a healthy living.

This solution offers a win-win environment for patients, healthcare providers, insurance carriers and the society in general. Healthcare can be truly affordable.

Metrics

1. Minimum Viable Product(MVP) that incorporates machine Learning algorithms for intelligent decision making. Decisions are transformed into recommendations for the user as means of achieving a healthy lifestyle.
2. User registration and subscription implementation with usage monitoring capabilities through account activity logs.
3. User frequency and performance data gathering and warehousing for machine learning performance tuning.
4. Quality - Efficiency/Ease of use, ratings, success stories, feedback, App Store ratings available.
5. Our evaluation will rely significantly on:
    - Accuracy of prediction
    - Precision
    - F1 Score
    - Recall

   All of the above will emerge from the results of:
    - False Positive Rate
    - False Negative Rate
    - True Negative Rate
    - False Negative Rate
6. Accuracy level of 75% and higher is ideal.

Analysis

Ideally, users of the app will consistently generate data over a period of time if committed to using the app and following the guidelines offered. The data is then used to train a model, which can be deployed and used to make predictions with respect to health status classification.

Since this proposal doesn't have its own generated data, I am using a sample data obtained from KAGGLE for this application. The data is CSV formatted (diabetes.csv). The difference will be that the data generated by the user reflects personal activity and performance within the app itself.

Data Exploration

The nature and characteristics of the dataset consists of 768 data points and 9 features. We are only interested in features that are contributing factors to the development of Diabetes disease in humans. Among them are:

- Glucose - Measures the type of sugar you get from the food you eat
- BloodPressure - The pressure of the blood within the Arteries
- Insulin - Scores the regulation of glucose level
- DiabetesPedigreeFunction - Measures likelihood of diabetes based on family history
- BMI - Blood Mass Index
- Outcome - Indicates Diabetic with (1) and Non-Diabetic with (0)

The source of data is the user and user activities. As the user participates, data is created and generated daily across the app and warehoused. The goal is to operate with each of the features at a reasonable level that will enhance healthy living.

It is important to note that this data is taken from pregnant women only. The dataset does not represent the general population. It is meant to demonstrate the application concept in Healthy Living Partner.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

Fig-1 DataFrame - Diabetic.csv

Data Visualization

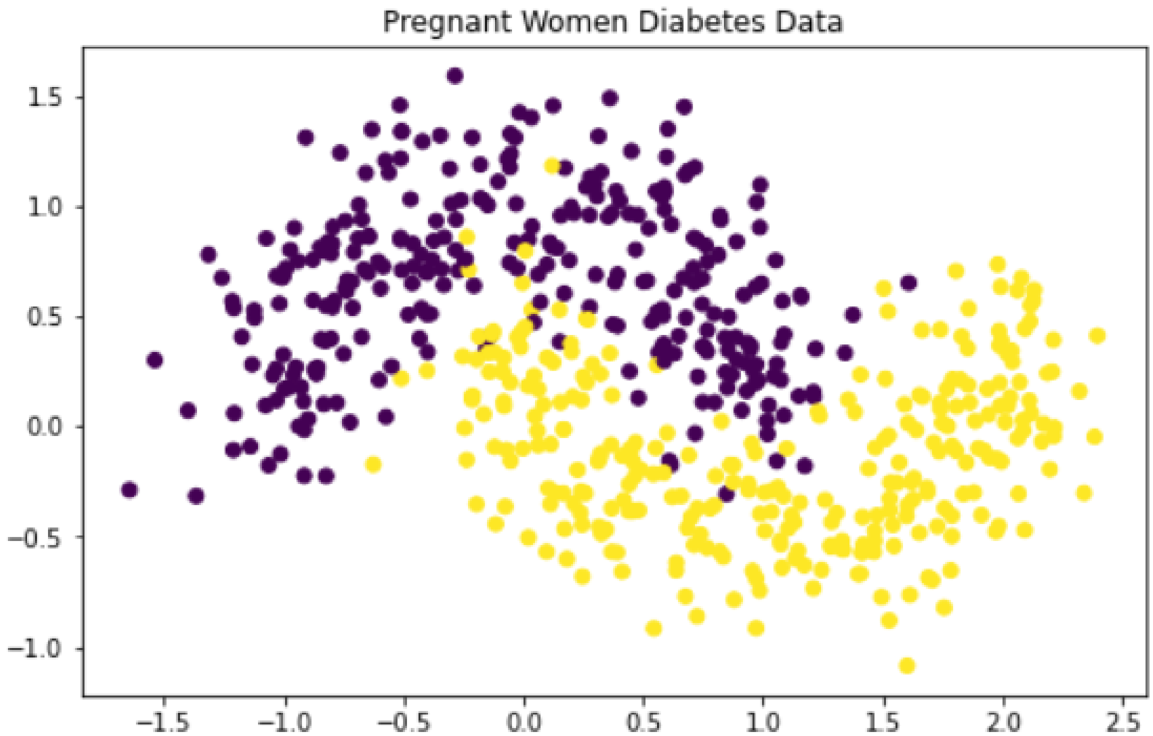

Fig-2 Showing Data point classes (diabetic.csv)

Methodology

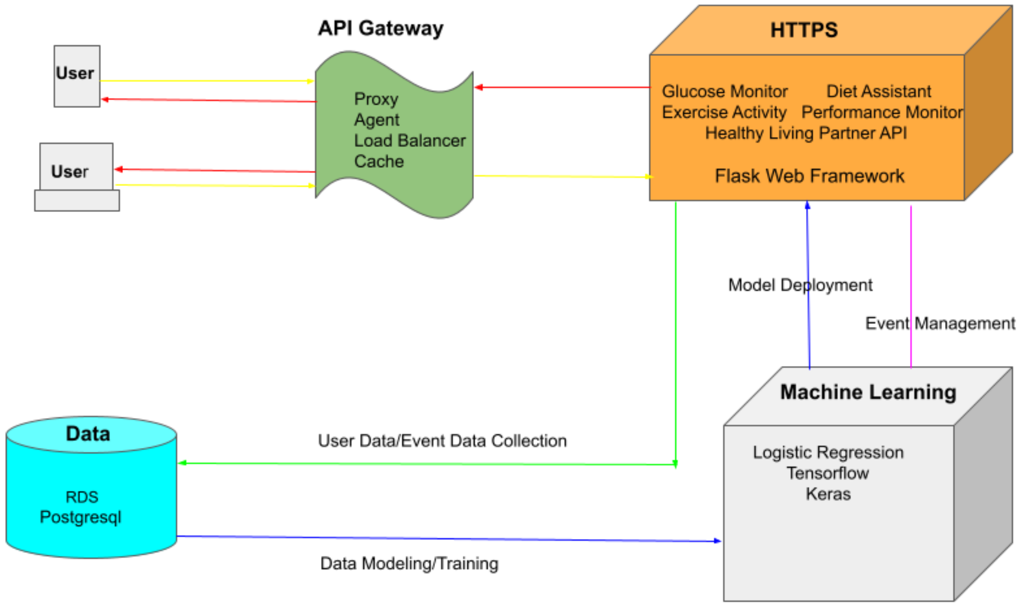

Figure-3 Healthy Living Partner System Diagram

Data Pre-processing

Environment & Tools

A. Scikit Learn
B. Python/Pandas
C. TensorFlow
D. MatplotLib
E. Jupyter Notebook

Implementation

This product solution is a web-based mobile application which requires hosting. The software development of the core features is not included in this proposal. The code sample provided along with this proposal is limited to the Machine Learning aspect only. However, the concept and key features are clearly articulated herein. The focus is to show how machine Learning is applied in the product solution.

Key Features & Scope

The proposed solution is to create an app that assists people to live a healthy life through monitored and controlled practice, thereby preventing chronic disease. Healthy Living Partner provides guides and recommendations to users as they approach and go through daily diet, exercise and health monitoring routine. It is the go-to-place for habit and lifestyle formation and modelling.

Product Features

1. Health Status Check
2. Diet Assistant
3. Exercise Activity Unit
4. Glucose Monitor
5. Performance Monitor
6. Application Programming Interface(API)

Health Status Check (Health Status)

Provides analytical report of user health condition

When the user signed on to the app, the user is taken to the Main Menu page, with a list of different options. The user selects Health Status, and will be required to input some information. The app will then analyze and return with a report on the user's health status. The goal is to give the user an insight to the level of urgency that will be required for his or her routine.

Diet Assistant (Diet List)

Provides a list of diet menu with recommended options for Breakfast, Lunch and Dinner

Entering the Main Menu page, Diet List button is one of the options. Users can select the button and will have access to a diet list. Users will also be presented with the options for Breakfast, Lunch or Dinner.

Exercise Activity Unit (Exercise List)

Provides the user with recommended exercise activity

From the Main Menu page, one of the option buttons is the Exercise List. Selecting this button will give access to a list of exercise activities from which the user can select. There are more options for recommended activities.

Glucose/Sugar-Level Monitor (Glucose Monitor)

Provides the user with measures of Blood Glucose level

The Glucose Monitor can be accessed from the Main Menu page by selecting the Glucose Monitor button. After supplying the required input, the Glucose Monitor will display the reading of the measure of blood sugar-level.

Performance Monitor (Performance Monitor)

Provides the user with an analysis and percentage rating of completion of tasks in a routine.

This is another option available on the Main Menu page. Select the Performance monitor button after completing an exercise routine. Users will get a display with analysis of the measure of completion, which transforms to level of achievement of and commitment to healthy lifestyle practices.

Application Programming Interface (API)

This feature is a technical development component purposely for integration with other complimentary products. API documentation is provided to developers upon request.

Data collection and usage in HLP

1. The Health Status Check feature checks the health status using the available data on the user’s health. It determines whether the user is healthy or unhealthy using a baseline threshold. Users are provided with diet and exercise recommendations.
2. The Performance Monitor gathers data relating to any aspect of the application that is used based on the events. For example, if the user adheres to diet and exercise routine recommended, activity is tracked, monitored and rated. This is to determine how well or closely the recommendation is being followed. It also serves as a measure of the user's commitment and accountability.

Data may be collected and utilized within the application to enhance the intelligence and feature capabilities as needed.

Data Storage

Data gathered and/or collected by the application is stored or warehoused in the cloud through cloud services. This is an essential part of the concept for scalability purposes.

Prototype

Proof of concept is provided with some pages of the apps as shown below along with a link.

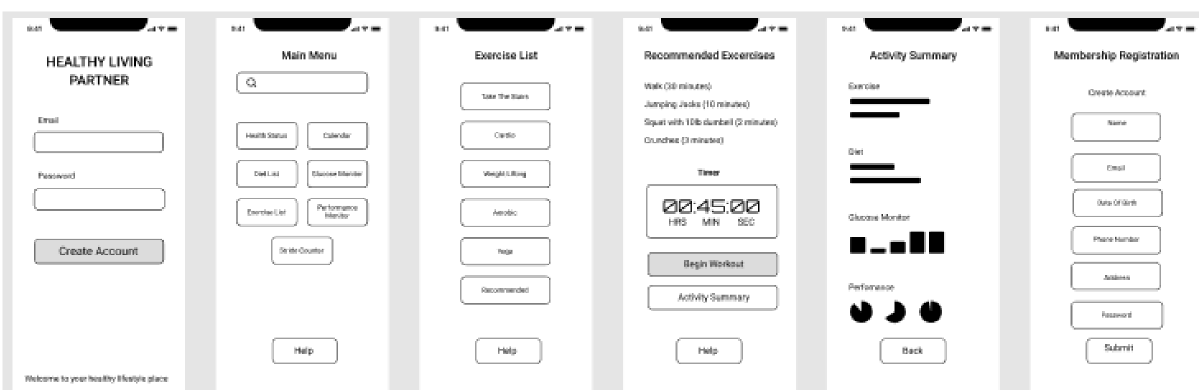

Fig-4 Pages (Healthy Living Partner)

[Healthy Living Partner(HLP)]

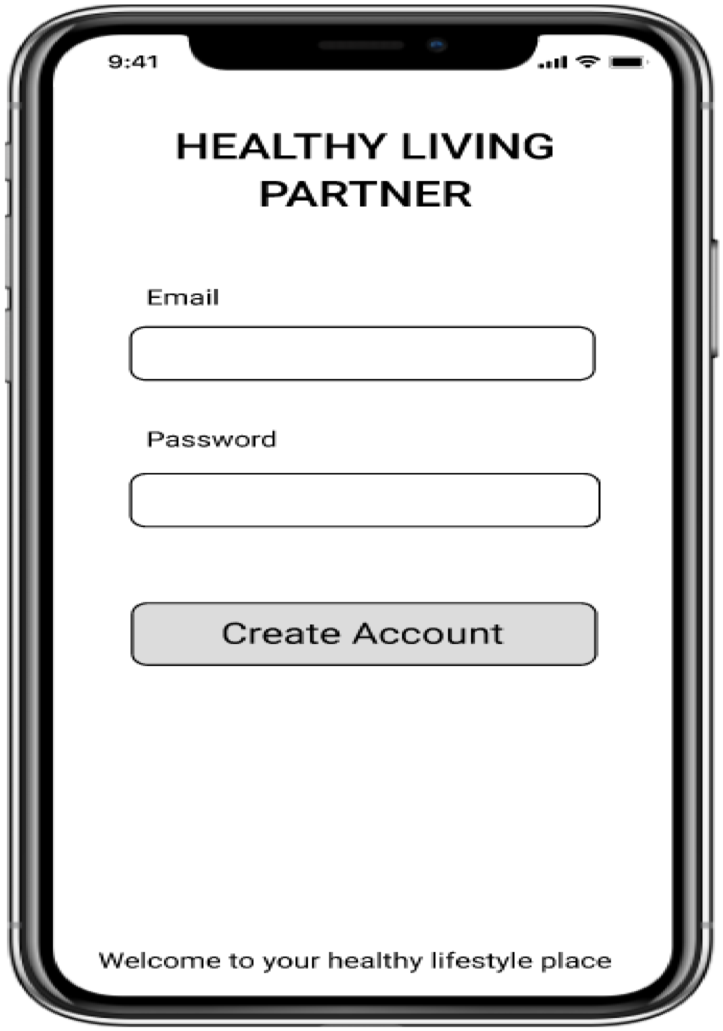

Figure-5 Healthy Living Partner

Applied Machine Learning Process

The solution presented herein applies machine learning techniques to solve a classification problem as a component within a product solution. We want to answer questions that have one of two possible outcomes as we attempt to determine and predict the user's health condition, which will lead to decisions on how to remedy and/or maintain good status habitually.

Logistic Regression binary classifier has the capability to predict outcomes. It is used to analyze the available data, and when presented with a new sample, mathematically determines its probability of belonging to a class. If the probability is above a certain cutoff point, the sample is assigned to that class. If the probability is less than the cutoff point, the sample is assigned to the other class.

Code Outline

1. Read in the csv formatted dataset as dataframe
2. Explore and understand the characteristics/nature of the dataset
3. Perform data cleaning as needed
4. Split training/test datasets
5. Preprocess numerical data using StandardScaler
6. Define the Logistic Regression model
7. Train the model
8. Evaluate the model(determine the accuracy)
9. Optimize, if and where necessary

The code can found here:
https://github.com/Francodo/Machine-Learning-Engineering/blob/main/healthy-living-partner.ipynb

Refinement

(a) Using a different classifier
The size of the available dataset in this proposal is fairly small for a typical Machine Learning model in production. A different classifier could produce better results depending on how the algorithm works. For example, SVM works by separating the two classes in a dataset with the widest possible margins. The margins, however, are soft and can make exceptions for outliers. This stands in contrast to the logistic regression model. In logistic regression, any data point whose probability of belonging to one class exceeds the cutoff point belongs to that class; all other data points belong to the other class.

(b) Adjusting Hype-rparameters
Model Hyperparameters can be adjusted for performance optimization.

Results

Model Evaluation and Validation

The process involved in establishing a Machine Learning model:

1. Create a model with LogisticRegression().
2. Train the model with model.fit().
3. Make predictions with model.predict().
4. Validate the model with accuracy_score().

Logistic Regression method is known for its accuracy and versatility. It compares the actual outcome (y) values from the test set against the model's predicted values. In other words, y_test are the outcomes, indicating whether or not a woman has diabetes from the original dataset that was set aside for testing. The model's predictions, y_pred, were compared with these actual values (y_test).

Logistic regression model accuracy: 0.729

```
Confusion Matrix
```

	Predicted 0	Predicted 1
Actual 0	105	20
Actual 1	32	35

```
Accuracy Score : <function accuracy_score at 0x0000017038E80948>
Classification Report
              precision    recall  f1-score   support

           0       0.77      0.84      0.80       125
           1       0.64      0.52      0.57        67

    accuracy                           0.73       192
   macro avg       0.70      0.68      0.69       192
weighted avg       0.72      0.73      0.72       192
```

The accuracy score is simply the percentage of predictions that are correct. In this case, the model's accuracy score was 0.729, meaning that the model was correct 72.9% of the time.
Justification

Data is expected to grow significantly as more users signed on and as users use the app more frequently. In anticipation of the Big Data scenario in a real production environment, a Neural Network(nn) model based on Keras/Tensorflow is also applied for training and testing the model. It offers flexibility in modifying hyper-parameters of the algorithm for better performance.

```python
# The keras model allow us to change hyper-parameters easily for optimization such as
# Relu and Sigmoid.

nn_model = tf.keras.models.Sequential()
nn_model.add(tf.keras.layers.Dense(units=16, activation="relu", input_dim=8))
nn_model.add(tf.keras.layers.Dense(units=1, activation="sigmoid"))

# Compile the Sequential model together and customize metrics
nn_model.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"])

# Train the model
fit_model = nn_model.fit(X_train_scaled, y_train, epochs=100)

# Evaluate the model using the test data
model_loss, model_accuracy = nn_model.evaluate(X_test_scaled,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")
```

```
Epoch 92/100
576/576 [==============================] - 0s 33us/sample - loss: 0.4168 - acc: 0.8003
Epoch 93/100
576/576 [==============================] - 0s 29us/sample - loss: 0.4165 - acc: 0.8021
Epoch 94/100
576/576 [==============================] - 0s 29us/sample - loss: 0.4164 - acc: 0.7986
Epoch 95/100
576/576 [==============================] - 0s 28us/sample - loss: 0.4159 - acc: 0.8003
Epoch 96/100
576/576 [==============================] - 0s 29us/sample - loss: 0.4158 - acc: 0.7969
Epoch 97/100
576/576 [==============================] - 0s 31us/sample - loss: 0.4157 - acc: 0.8003
Epoch 98/100
576/576 [==============================] - 0s 33us/sample - loss: 0.4152 - acc: 0.7969
Epoch 99/100
576/576 [==============================] - 0s 31us/sample - loss: 0.4150 - acc: 0.7951
Epoch 100/100
576/576 [==============================] - 0s 36us/sample - loss: 0.4145 - acc: 0.7986
192/192 - 0s - loss: 0.4877 - acc: 0.7500
```

Figure-6 Neural Network tf.keras output

Note that the Neural Network tf.keras algorithm produced a better accuracy with an epoch of 100. Changing some of the hyper- parameters for optimization will eventually result in better performance.

Conclusion

Reflection

The overall solution objective is improved health and wellness through controlled diet and exercise activities, paving way for reduced costs of diagnosing and treating chronic diseases. Healthy Living Partner(HLP) offers this solution with its core features namely,Health Status Check, Glucose Monitor, Diet Assistant, Exercise Activity Unit, Performance Monitor and API.

Machine Learning algorithm is the main driver behind decisions made in terms of what actions to take, as well as, recommend to the user. It collects and analyzes data generated by the user for guidance on maintaining a healthy lifestyle. For these reasons it is absolutely necessary for the Machine learning algorithm to continue to perfect itself in learning, training and making predictions in support of decisions.

As articulated and demonstrated in this proposal, Logistic Regression and Neural Network (TensorFlow/Keras) algorithms were used to demonstrate the capabilities of the application of machine learning in this solution to determine who's healthy and who's not healthy. Furthermore, output trends indicate improving or deteriorating health conditions.

Improvement

1. Develop the code for the core features of the product

2. Complete the deployment portion of the code

3. Expanding the API coverage
   There are other health related devices that monitor other conditions in humans, especially those with pre-existing conditions. Expanding the API coverage to other test, reporting and monitoring devices for critical data collection and analysis to support decisions.

4. Biases in the data
   It is highly recommended to make provision that will address any anticipated and occuring biases in the dataset.

5. Multiple decision criteria
   Investigate the use of algorithms that utilize multiple decision criterias for improved accuracy.

References

Pima Indians Diabetic Data
https://www.kaggle.com/uciml/pima-indians-diabetes-database

Economic Costs of Diabetes in the US 2017
https://care.diabetesjournals.org/content/early/2018/03/20/dci18-0007

Healthcare statistics for 2021
https://policyadvice.net/insurance/insights/healthcare-statistics