# FIT N MODEL: A RECURRENT NEURAL NETWORK BASED CARDIO-ACTIVITY RECOMMENDER

**Chosen Realm G. Alfeche, Jovi E. Cataluna, Eduardo P. Mendoza Jr.**
[1]*University of San Carlos, Cebu City, 6000 Cebu, Philippines*
*For correspondence; E-mail: chosen.alfeche@gmail.com, jovicataluna@gmail.com, eduardopmendozajr@gmail.com

***ABSTRACT:*** *Many people want to achieve a physically fit body without the help of certain trainers or simply just want to do it alone. Unfortunately, due to the lack of knowledge on which exercises are most effective, they're unlikely to achieve that goal. One possible solution might be to create a recommendation model that can help people with better decision making from much complex information. Neural network is a computing system that is able to learn things by recognizing patterns and makes decisions similar to that of a human being. This research aims to build a neural network model that recommends what daily cardio-activities are needed to be fit. A lot of applications that records user's daily activities exist nowadays but none of which analyzes the data and give suggestions to the user using machine learning. This research focuses primarily on creating an LSTM neural network model that suggests common physical activities such as walking, running and cycling based on the current daily cardio-activities being performed by the user while also taking account of the calories being burned.*

**Keywords:** LSTM, Pandas, Square Mean Error, Symmetric Mean Absolute Percentage Error, Numpy, Tensorflow, Neural Recommender

## 1. INTRODUCTION

There exist a lot of activities that can help a person be physically fit. The only problem is due to many choices, one tend to be confused of what certain things to do [1]. A research shows that 50% of people want to get back in shape in 2017, but don't know where to start and almost three quarters (71 %) admitted to have never used a personal or group trainer and another 96 % aren't prepared to do a basic search online for quick tips and video workouts [2]. Although many trainers are available these days to guide people through becoming fit, many prefer not to since its time and money consuming. Also, based on a survey, people like time-efficient workouts and prefer to exercise alone [3]. It would really be helpful if a recommendation model or system exist to guide people on the way to fitness. This study will build a neural network model that takes users daily cardio-activities and calories burned as inputs and will output the activities expected to be done the next day to be fit.

## 2. OBJECTIVES

The general objective of this study aims to create a neural network model that recommends users what appropriate activities needed to be done to stay fit based on the daily activities acquired from the user. This research intends to: 1. Gather as much activity data from physically active people. 2. Preprocess collected data. 3. Build an LSTM neural network. 4. Train the neural network using data inputs. 5. Test and evaluate the model. incentive and the budget mechanisms. Farmer, education and research are often stressed because it is important to realize that there are many more actors in the food chain that directly influence decision making of farmers and their innovations.

## RELATED LITERATURE

There have been similar studies tackling about creating recommendation systems for health purposes.

A study made by Ponpisit et al., Sineerat et al. & Therdpong et al.[4] presented a mobile health application named "mHealth", created to perform recommendations for healthcare support with referred exercise on a mobile device. This application advices an activity or sport such as aerobics, cycling, jogging and etc. depending on the BMI, BMR and energy used. Special exercise advice are presented to patients with health issues. The system consists of several functions including the BMI calculation, BMR calculation, exercise caloric calculation, exercise recommendation, diary and profile for the user.

A study by Silva [5] developed an interesting mobile application called "SapoFitness" developed for dietary evaluation. The application accepts the user's height, weight, age and sex and will calculate both daily calories consumed and a weekly weight loss. The user can also input daily meal data to see the energy statistics. The goal of this application is to offer a motivation tool for weight loss and increase physical activity.

A research by Pramano et al. [6] proposes a recommendation system based on ontology that provides physical activity/exercise recommendations to diabetic patients. These recommendations

suggest the types of exercises including its intensity, frequency, and duration in relation to the condition of the patient.

A study done by Nassabi et al. [7] created research that aims to develop a new service model to prevent frailty focusing on the physical, nutritional and cognitive functioning. The focus of this research is to keep seniors physically active by developing an intelligent system that recommends exercises in mind with an individual's health status, goals and preferences for a personalized recommendation.

## TECHNICAL BACKGROUND

This section shows the definitions of the concept, tools and algorithms being used by the study.

**LSTM (Long short-term memory)**: is a type of recurrent neural network architecture suited for classification, process and predicting time series given time lags of durations from certain events. It is basically used as a solution for vanishing gradient problem.

**Gradient Descent Optimization:** an that algorithm basically implements a simple gradient descent optimizer.The current position is being updated at each iteration according to the equation:

$$p_{n+1} = p_n + \text{learningRate}\,\frac{\partial f(p_n)}{\partial p_n}$$

**MSE (Mean squared error)**: shows how close a regression line is to a certain set of points. The smaller the MSE, the closer it is to find the line of best fit. It is the sum of the variations of the individual data points squared, divided by the number of data points minus 2. When the data is displayed on a graph, you determine the MSE by summing the variations in the vertical axis data points. On an x-y graph, that would be the y-values.

**SMAPE (Symmetric Mean Absolute Percent Error)**: measures accuracy based on percentage or relative errors between 0% and 100%. It is defined by the following formula:

$$\text{SMAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|\text{Forcast}_t - \text{Actual}_t|}{|\text{Actual}_t| + |\text{Forcast}_t|}$$

However this formula does not treat over- and under-forcasts equally. So another formula of SMAPE was created and

$$\text{SMAPE} = \frac{\sum_{t=1}^{n}|\text{Forcast}_t - \text{Actual}_t|}{\sum_{t=1}^{n}(\text{Actual}_t + \text{Forcast}_t)}$$

defined as:

The other formula allows to measure the direction of the bias in the data by generating a positive and negative error. This formula is also better protected against outliers and the bias effect than the other formula.

**MDA (Mean Directional Accuracy)**: is a measure of prediction accuracy of a forecasting method. The forecast direction (upward or downward) is compared to the actual realized direction. It is defined as:

**Excel**: is a tool developed by Microsoft to organize, format, and calculate data with formulas using a spreadsheet system broken up by rows and columns. After preprocessing the data it is exported into an excel file to better visualize the type and owner. Excel is also an easy way to put a header on the data for easy categorizing of columns upon loading it for training the neural network.

**Tensorflow API:** is an open source software library for numerical computation using data flow graphs. It is developed for purposes of conducting machine learning and deep neural networks researches. Tensorflow has APIs available in several languages both for constructing and executing a TensorFlow graph. The Python API is at present the most complete and the easiest to use.

**Python**: is a popular object-oriented programming language greatly used in research for creating machine learning and natural language processing. It can be used for everything from web development to software development and scientific applications. 3.9 Numpy Numpy is the core library for scientific computing in Python. It provides a high-performance multidimensional array object, and tools for working with these arrays.Numpy array is a grid of values having the same data type indexed by a tuple of nonnegative integers.

**Pandas**: is an open-source,BSD-licensed software library written for Python for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is the most powerful and flexible open source data analysis and manipulation tool available in any language.

## METHODOLOGY

This section contains the systematic and theoretical analysis of the methods applied into this study. In this part, the steps needed are mentioned in order to finish the expected goal of the research which is to create a recommendation model for users aware of being fit.

**Concept:**

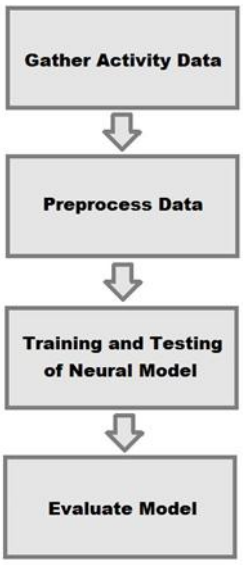


**Figure 1. Conceptual Framework**

Figure 1 shows the conceptual framework of the study. The researchers collected as much cardio activity data from physically active people in order to build the model. The datas are then filtered out to be split as training sets and testing sets for the model. After training, the model is then evaluated.

## ANALYSIS AND DESIGN:

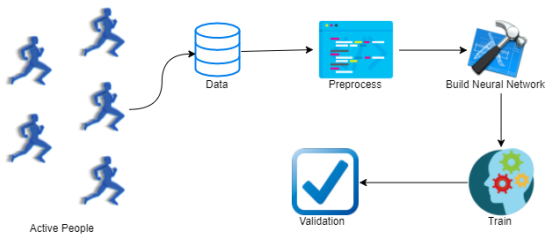


**Figure 2. Analysis and Design Activity**

Diagram Figure 2 shows that cardio-activity data are gathered from various people. The data is then preprocess to be prepared for training the neural network model. The model is built to cater this data and work in such a way to achieve

the objective. It is then trained, optimized and learned patterns on the preprocess data. The model is then evaluated with the use of three prediction accuracy measurements.

**DEVELOPMENT APPROACHES**:

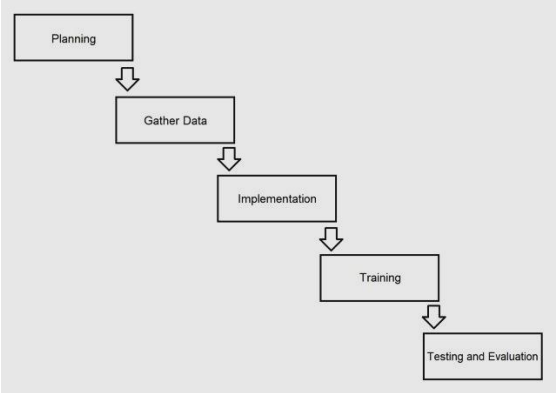


**Figure 4. Bottom up approach diagram**

Figure 4 shows the approach of this study followed in order to achieve the goals, representing it using the bottom up approach of software development. Gathering up important data and preprocessing it after to be able to feed it to the already built neural network as inputs. The trained neural network would then look for patterns within the given inputs and will output the suggested activities needed to be done by the user to be fit. After that model is then tested and evaluated.

**Software Development Tools:** In this research, Python was used as the programming language since it is a heavily supported language for machine learning with different kinds of libraries available. The libraries that are used in this research are Tensorflow, Numpy and Pandas. Pandas is used for the exporting and reading of the excel data file while Numpy provides mathematical functions for computation of non-tensor datas. Tensorflow is the main library used to create the model. It is similarly the same with numpy but the difference is that tensorflow offers more in terms of machine learning. With the use of a dataflow graph to represent the computations, tensorflow offers a low-level programming model that accelerates these computations.

**PRESENTATION, ANALYSIS AND DISCUSSION OF RESULTS**

This chapter aims to discuss the specific objectives in detail and the steps taken to achieve the specific objectives.

**Gather activity data**: In order to gather activity data from different physically activity people, the researchers mine the available Runkeeper data from OpenHumans.org in JSON format. These activity data are from people who are physically active and had gather their activities in an app called RunKeeper and synced it in to OpenHumans, an online open data source for researchers.

**Preprocessing the Data**: To preprocess the data, it was run on a python code and sorted out based on the year, day and owner. Similar types of data within the same day e.g. walking, running and cycling were added up to to have a single type per day. The data is then exported as an excel file with the columns of cycling, running walking, burnt calories, date. The Figure below shows the format of the data.

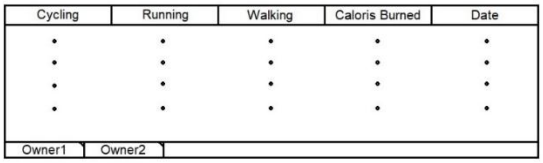


**Figure 5. Data format in excel file**

**Training and Testing the Neural Network:** The neural network was trained by importing the collected data from the excel file and was split into two separate sets namely training and testing sets. In order to produce the testing set, 30% of the data from each owner were pulled and appended to the testing set. The model is required to take in a sequence of data, so an array of zeroes with a shape of (7, 4) were created. With each iteration, the current activity data of the user was appended to the last position and the data at the first index was deleted to maintain the (7, 4) shape.

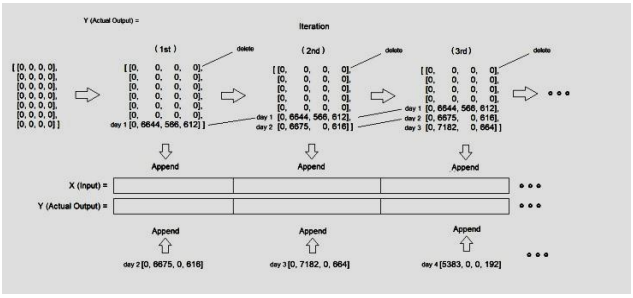


**Figure 6. Representation of data sets**

The neural network is an LSTM recurrent neural network model with a dropout regularization of 50% out keep probability implemented to help reduce overfitting. With the help of Tensorflow, the learning and optimization of the model was easily implemented. Gradient Descent with a learning rate of 0.01 was used of the as the optimization function. The loss function Mean Squared Error (MSE) was used to calculate the error between the actual and forecast values. To test the accuracy of the neural model, Mean Directional Accuracy and 2 variants of Symmetric Mean Absolute Percentage Error were used to calculate the accuracy. The researchers used three error algorithms because each one has function.

**Evaluation**: With each epoch, the test set and training set were fed into the neural network and the accuracy was calculated. Mean Directional Accuracy (MDA) was used to calculate the forecast direction to the actual output and the two SMAPE (Symmetric Mean Absolute Percentage Error) were used to calculate the percentage of relative error. On Table 1 and Table 2 to get the accuracy of the neural network the researchers subtracted the output of SMAPE (error percentage) from 100. SMAPE1 and SMAPE2 are the 1st and 2nd formula of SMAPE respectively.

**Table 1. Accuracy of training with each epoch**

| Epoch | SMAPE1 | SMAPE2 | MDA |
|---|---|---|---|
| 0 | 52.34 % | 30.29% | 43.72% |
| 1 | 58.19% | 28.49% | 49.51% |
| 10 | 62.80% | 34.73% | 54.83% |
| 20 | 62.73% | 33.33% | 53.36% |
| 24 | 63.82% | 34.16% | 54.68% |

**Table 2. Accuracy of testing with each epoch**

| Epoch | SMAPE1 | SMAPE2 | MDA |
|---|---|---|---|
| 0 | 49.97% | 29.24% | 42.42% |
| 1 | 58.98% | 27.85% | 50.32% |
| 10 | 64.45% | 33.3% | 55.52% |
| 20 | 63.58% | 34.14% | 54.21% |
| 24 | 66.73% | 34% | 56.53% |

The table 1 and 2 shows that with each epoch the accuracy of the model was able to go up and has reached 66.73% at the epoch 24. The model was able to converge both training and testing sets avoiding overfitting

## CONCLUSION

The results of the study produced a recurrent neural network that can look for the pattern and trend of the cardio-activity data from active people and can follow or produce an output similar to that data. The researchers successfully created a neural network with an accuracy of about 66% that can follow the trend and recommend an activity to be done based on what it has learned.

## FUTURE WORK

For future works, the researchers recommend to have access to more data and to train for the neural network this study was done with only about 1,200 data gathered from people who publicly shared their RunKeeper data. This kind of models can also be integrated to currently available systems to help and recommend the user on what to do. And with more data, features and algorithms implemented this model can be much improved.

## REFERENCES

[1] Schwartz B. (2014). The Paradox of Choice - Why More is Less

[2] Jabra (2017). Research Reveal People Prefer to Work out Alone

[3] TRUFITNESSTECH (2013). Survey reveals people prefer working out alone

[4] Ponpisit W., Sineerat R., & Therdpong D. (2014) mHealth: A Design of an Exercise Recommendation System for the Android Operating System

[5] Silva B. M., Lopez I. M. & Rodrigues J. J. P. C. (2011) SapoFitness: A mobile health application for dietary evaluation

[6] Djoko Pramono, Nanang Yudi Setiawan, Riyanarto Sarno & Mohamad Sidiq (2013) Physical Activity Recommendation For Diabetic Patients Based on Ontology

[7] Nassabi, M. H., Akker H. o.e., & Vollenbroek M. (2014). An ontology-based recommender system to promote physical activity for pre-frail elderly