Predictive Analytics Case Study: Fitness and Health

Ruben Brionez Jr

Bellevue University

DSC630: Predictive Analytics

Andrew Hua

07/07/2023

Predictive Analytics Case Study: Fitness Data

When thinking of a case study for predictive analytics, various topics come to mind. Predictive analytics can be seen in nearly every industry today, from professional sports to retail. Predictive analytics has many applications across industries, the focus for this case study will be the use cases of predictive analytics in fitness.

The case study was made possible by the increase in fitness tracking devices from various vendors. With the increase in fitness trackers, the available data also increased. With the increase in data came a desire to study and gain insight from it.

Problem to Be Solved

There are many different problems that predictive analytics can help solve when it comes to fitness. The issue studied for the case study was not a specific problem but more of a generic answer to the possibilities of using predictive analytics for health and fitness. The case study aimed "to apply predictive fitness analytics techniques, analyzing current and historical data and then promoting a plan to increase health and wellness." (Bennet. 2022).

Importance of Resolution

The importance of finding a resolution, or in this case, finding meaningful insights from the use of predictive analytics on fitness data, is the possibility of improving the health and wellbeing of people. Finding ways to improve the lives and health of others using the increase in available data and predictive analytics can benefit everyone.

Data Acquisition

The data acquired for the case study came from many different brands and types of wearable devices. Since the case study was focused on personal use devices, the devices would have been readily available at any retailer for purchase. Wearable devices such as the Samsung and Apple branded smart watches collect various types of health and fitness information throughout the day. They can record the

number of steps taken in a given period of time, measure heart rate, measure the amount of oxygen in the blood stream, and monitor sleep cycles when worn during sleep.

Many other devices exist for tracking different health and fitness parameters. From personal experience I own a blood pressure monitor, a Samsung watch, and a thermometer, all used to track different health and fitness metrics. All these different devices provide a plethora of data points that can be used in predictive analytics for monitoring health and fitness.

Methods and Results

Data Preparation

Data preparation is key in utilizing the data for predictive analytics. In this case study, data from various devices were gathered and compiled into a large data set. The specific health or fitness issue to be targeted would then dictate the methods of data preparation. This is due to the fact that various target conditions or fitness goals do not require the same metrics. Predicting a stroke or heart attack would require different parameters than tracking a fitness goal or predicting if a person will get diabetes. From the case study:

> "Depending on the illness we are trying to predict we have to reduce the number of variables
>
> and restrict the data by using the factors that contribute to the impending illness in question.
>
> For example, we are trying to predict if the person might have a stroke or an MI, then we
>
> restrict our data to heart rate, blood pressure and oxygen concentration." (Bennet. 2022).

This quote illustrates the importance of how the target will dictate the data preparation and preprocessing methods.

The data preprocessing for the case study involved several processes, the first of which was deciding the target. Once the target was decided, the next most important preprocessing step was the removal of outliers. Tracking devices available to the general public are not necessarily equivalent

medical grade equipment and thus may have more erroneous data and outliers. After the outliers are removed, the next step was to normalize the data:

> "The normalization in many learning algorithms is to assume that all features are zero mean and have variances in the same order. If the variance of the feature is several orders of magnitude larger than the variance of other features, it will dominate the learning algorithm, increasing the system weight of certain features and also skew the results."(Bennet. 2022).

Normalization of the data, as mentioned in the quote from the case study, reduces the weight of variance between the selected features. This preprocessing technique aims to reduce a single feature dominating other relevant features for the learning algorithms.

Problem Solution

The case study did not provide a definitive solution but more of an overall picture of how the data could be used for predictive analytics. The case study states that with the increase in wearable devices, health and fitness tracking data is more easily accessible and tracked. This data can and should be used to predict future conditions and illnesses that can be caught early and treated.

The case study goes on to mention that with this increase in available data and tracking more case studies are being created. "Several studies are going on to see if wearable data can provide early indication of viral illnesses of influenza and of even COVID in recent times." (Bennet. 2022).

Modeling

The case study stated that the Support Vector Machine algorithm would be used for predictive analytics of the health and fitness datasets. This type of model would be used for its superior processing of health and fitness related data. Per the case study "the SVM is based on the minimization of structural risk, rather than minimizing the empirical risk in traditional algorithms." (Bennet. 2022). The case study itself did not mention the scoring metrics used for the Support Vector Machine, but further

research found that Accuracy, F1 Score, Recall Score, and Precision Score can all be used. This was found from a different case study involving predictions on COVID-19 (ResearchGate. 2020).

Conclusion

Although the case study did not mention direct results or provide information on how the case study was resolved, it did provide insight into future possibilities. The conclusion of the case study referenced how the data being created by wearable devices could, and in fact has, spurred future case studies in predictive analytics for health and fitness.

The case study did not reference anything that the team may have learned, but it appears evident that the greatest insight from this case study is how health and fitness data can be used in the future to predict and treat various human ailments and diseases.

There was not a discussion on what would or could have been done differently. The most significant thing that could have been changed would be providing additional details and conclusions on the case study as opposed to a high-level overview of the case study. A detailed review of the data set used would have also been beneficial for readers and observers of the case study.

References

Bennet, Deepthi Tabitha. (2022). *Five Fascinating Case Studies in Predictive Analysis.* Medium. [URL](URL)

ResearchGate. (2020, July). *COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm.*

[URL](URL)