Predicting and Forecasting Happiness

Amit Timalsina

Young Scientists Community (YSC), Kritipur, Nepal

E-mail: amittimalsina14@gmail.com

Category: Computer Science

Abstract

In this busy world, being happy is difficult. Developing an approach to predict and track happiness based on individual actions could enable us to select daily actions and behaviors for being happy and thereby enhancing the well-being of life. Also, being able to forecast happiness could enable us to take our later actions accordingly to be happy. Therefore, I purpose a novel approach of applying Machine Learning and Deep Learning-based Forecasting, for predicting, tracking, and forecasting happiness using just a very few lifelog items. I asked a good friend of mine to keep track of a few parameters like relationships, social interaction, a few atmospheric conditions, etc [and tracking continues till this day]. It's tedious to track a large number of such parameters so after several iterations of the model I finally came up with a Random Forest model to predict happiness using just 4 four parameters/behaviors namely, 'relationship', 'social', freshness', and atmospheric pressure with 88% accuracy. This suggests that our approach can predict the parameters/behaviors that increase individuals' happiness in their daily lives, thereby contributing to the improvement in their happiness. Furthermore, I developed a Time series forecasting model with only 0.26 Mean Absolute Error (MAE) to forecast happiness in the future by just using a week of data. If we know what's the degree of happiness that we are going to achieve at the end of the day before starting the day, then we could act accordingly throughout the day to be happy at the end of the day that too using just 1 week of data. Besides, the accuracy of the model keeps on increasing as we keep on accumulate more data.

Keywords: *Happiness prediction and forecasting; Machine learning; Time series forecasting;*

INTRODUCTION

Happiness has been considered an important factor for human lives as it impacts productivity, health, wealth, and many other social factors. Most people think they know when they are happy, or sad, but we humans are very poor at estimating the degree of happiness[1].

So, how do we know when we are happy? A system is required that predicts happiness based on individual conditions, individual actions, and atmospheric conditions. Also, as the person keeps on predicting happiness, the system could inform the person about selecting certain behaviors for enhancing well-being in life or informing about which of the personal actions matters the most for the happiness of that person.

Another important feature is being able to know our degree of happiness in the upcoming day based on the inputs from previous days. Let's say the system forecasts happiness of 4 [on a scale of 1 to 5; 1 being the worst and 5 being the best]for a day then a person would try his/her best to maintain that degree of happiness by tuning his behaviors and actions accordingly. But if the system forecasts the happiness of 1 then the person tries hard to not have a bad day by acting accordingly throughout the day. This means, If we can know what's the degree of happiness that we are going to achieve at the end of the day before starting the day using just 1 week of data, then we could act accordingly throughout the day to be happy at the end of the day.

To meet the above objectives, I have developed two models, one for predicting happiness [predictive model] and the other for forecasting happiness [forecasting model]. The predictive model is deployed to a simple web app to check its feasibility.

Figure 1. The input and output structure of the predictive model

METHODOLOGY

1. <u>**Predictive model:**</u>

Getting the data: I asked a good friend of mine to keep track of a few parameters like relationship, social interaction, a few atmospheric conditions, etc [and tracking continues till this day]. Most of the personal parameters/features like relationship, social, refresh, target happiness, etc were collected in a range of 1 to 5; 1 means the worst and 5 means the best.

Feature selection: Feature selection is crucial in machine learning, as it can have a significant impact on the outcome of the algorithms. During exploratory data analysis, In the first step, I looked at the quality and quantity of each feature like latitude, longitude, temperature, atmospheric pressure, relationship, refresh, etc. Then after looking at the correlation of each feature, I shortlisted the most correlated, qualitative features.
Then recursive feature elimination (RFE) [2] was used to create a feature set. In RFE, the classifier is first trained with all features. Next, the least important feature is removed. That procedure is repeated until a single feature is left. The best feature set is the one that achieves the highest accuracy. For the creation of the final feature set, a decision tree was used as an estimator. This algorithm can return the feature importance of each feature, which, in turn, can be used for determining the best features.

	timestamp	**Relationship**	**social**	**refresh**	**target**	**at. pressure**
0	5/12/21 0:36	5	2	3	5	1020
1	5/12/21 11:42	3	3	3	4	1017
2	5/12/21 14:47	4	2	3	4	1016
3	5/12/21 17:48	4	3	3	5	1014

Table 1: sample of data collected

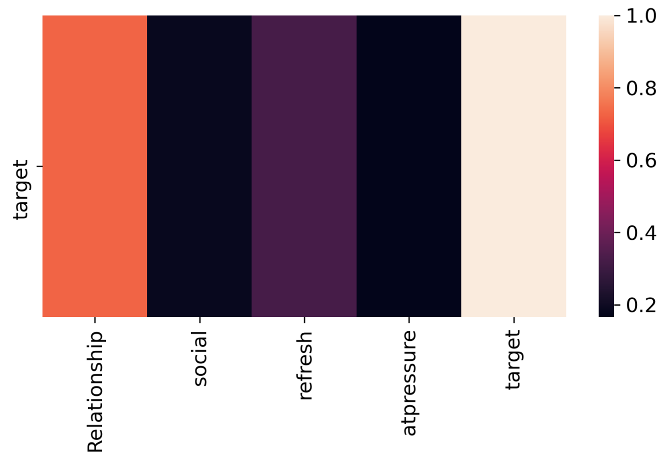

Figure 2. The heatmap of correlation of features with Target[target happiness of each data entry] of the final feature set

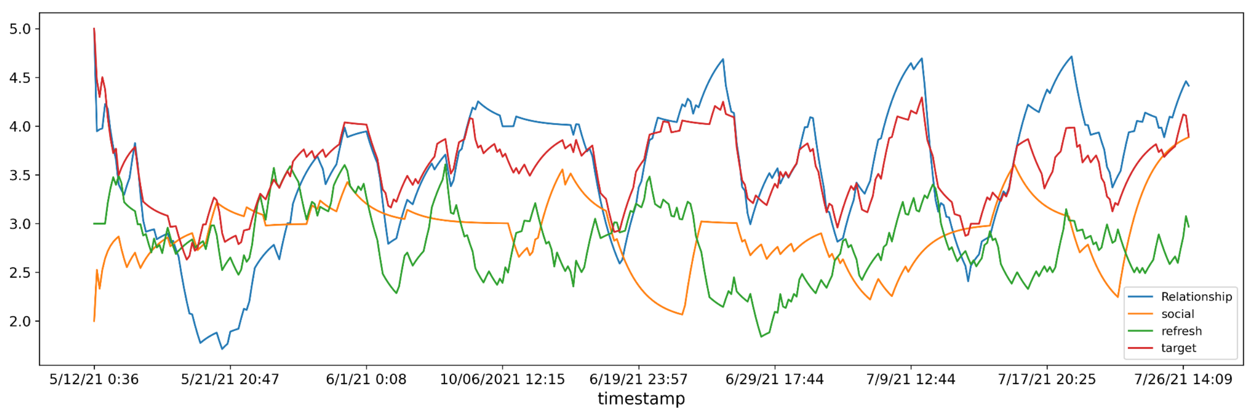

Figure 3. Exponentially weighted average of features vs time

Performance Measure: Accuracy and F1 score were used as the performance measures.

Evaluate Prediction: The dataset after data preprocessing contained 5 columns and 350 data entries in those columns. The dataset was split randomly into train and test sets of ratio 4:1. Then the train set was trained using algorithms. Table 2: shows the key performances of each algorithm for predicting happiness. The random forest was most accurate in the train set with an F1-score of 88.1% accuracy and in the test set with an F1-score of 67%.

Model	train set	test set

log_clf	0.685083	0.585366
decision_tree	0.762431	0.593943
SVM	0.726519	0.615860
forest_clf	0.881215	0.666994
bag_clf	0.845304	0.597448
ada_clf	0.881215	0.618809
grad_clf	0.881215	0.655398

Table 2: Performance metrics of each algorithm for predicting happiness

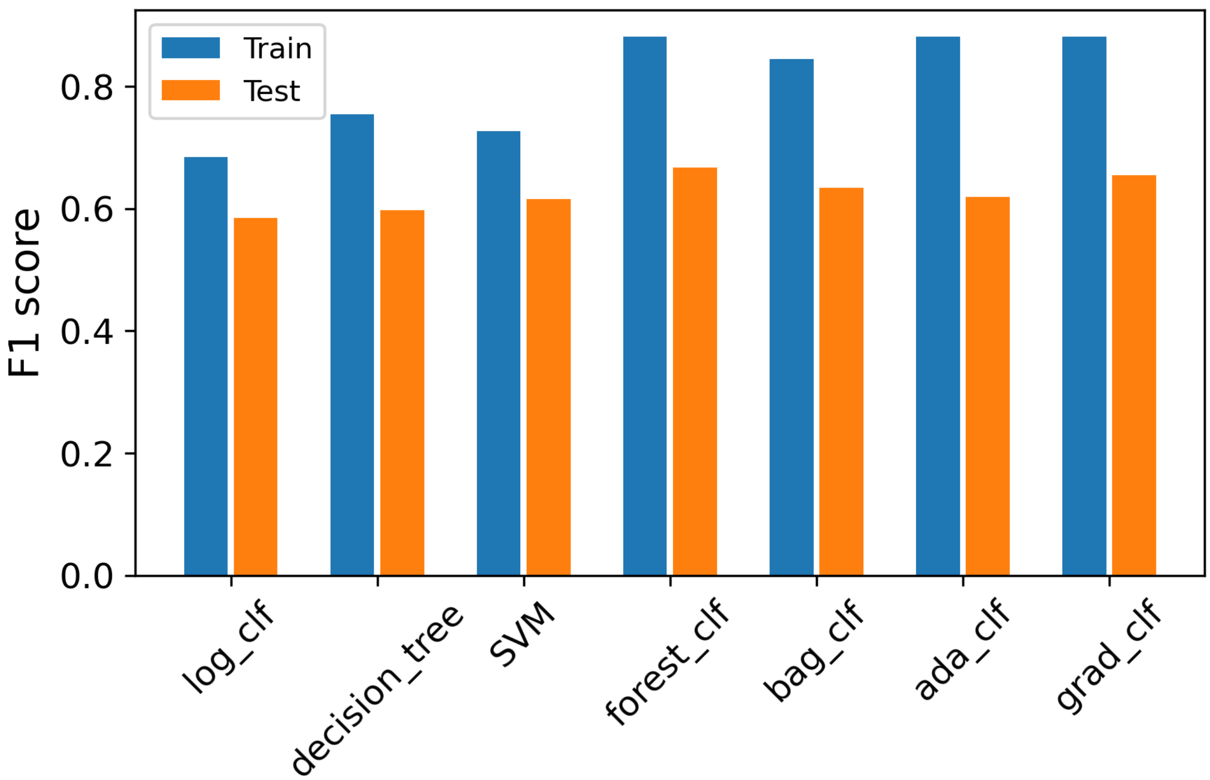

Figure 3. F1 scores of different models for train and test set; the random forest has the highest scores

Also, The second best model was Gradient Boosting with a score of 88.1% on the train set and 65.5% on the test set. While the F1-score of random forest and gradient boosting for the train set was the same, they differed in the test set score.

2. Forecasting model:

Getting the data: The data collected by my friend also contained timestamps in them for every entry. The data was collected numerous times each day for fast data collection. So, it was subsampled to 1-day intervals for dealing with daily predictions. I used the mean of data by day's for subsampling as it will also help to make data more reliable.

Preprocessing the data: Except for the timestamp column as shown in table 1, other columns didn't require much processing. So, the timestamp column was first converted into seconds. Time in seconds is also not a useful model for input. Being personal data collected each day it has some weekly, and yearly periodicity. But since the data had been collected for just two months, looking for only weekly periodicity makes sense. So, for dealing with periodicity and for converting it into a usable signal, sin and cos were used to convert the time to time of week signals.

Performance Measure: Mean Squared Error(MSE) and Mean Absolute Error(MAE)was used as the performance measure.

Evaluating Forecast: The developed model can forecast a single timestep in the future [1 timestep = 1 day] with a minimum of 1 week of the data window [i.e. 7 days]. Besides this, The accuracy of the model keeps on increasing as we keep on accumulating more data.

Table 2 shows the Mean Absolute Error(MAE) of each algorithm used for forecasting happiness. The Long Short Term Memory (LSTM) had the least error. It had an MAE of 0.44 on the validation set and only 0.227 on the test set which is 27.9% and 25.8% lower than the baseline MAE respectively.

	Validation set	**Test set**
Baseline	0.609524	0.314626
Linear	0.582443	0.430682
Dense	0.829249	1.034992
Multi-Step Dense	0.582268	0.425072

Conv	0.743726	0.602560
LSTM	**0.442987**	**0.227660**

Table 3: MAE of each algorithm for forecasting happiness

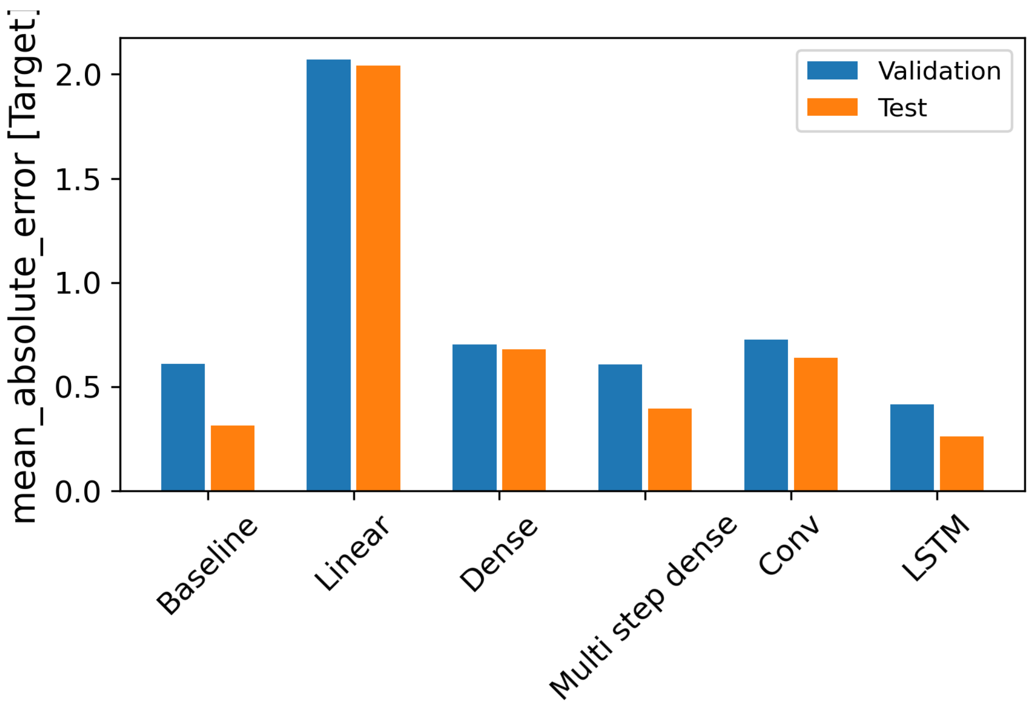

Figure 3.MAE of different models; LSTM got the least validation and Test error

RESULTS AND DISCUSSION

The participant recorded the lifelog items every day. Further, he reported increased awareness about his lifestyle through recording behaviors.

The highest accuracy of the predictor model is 88.1% and 67% on train and test set respectively. This level of accuracy is great with just only 5 input parameters and 300 data points. The predictor model is currently optimized for the participant as the input features are selected based on only his data. There are numerous personal behaviors and environmental factors that affect happiness. So, The next step is to further research the most appropriate input features for the preparation of a generic model. This suggests that our approach can predict the features that increase individual's happiness in their daily lives. Furthermore, after the preparation of the generic model, by looking at the importance of features on happiness based on the data entered we could fine-tune the model for that particular user.

Fig 3. Deployed working model

The developed model can forecast a single timestep in the future [1 timestep = 1 day] with a minimum of 1 week of the data window [i.e. 7 days] for fewer errors. Besides this, The accuracy of the model keeps on increasing as we keep on accumulating more data

The lowest MAE of the predictor model is 0.44 and 0.23 on validation and test set respectively. This suggests that our approach could be able to forecast any of the other variables, like relationships or any other. The forecasting model still has some fine-tuning steps left to perform and certainly, after doing so the errors would be significantly lower.

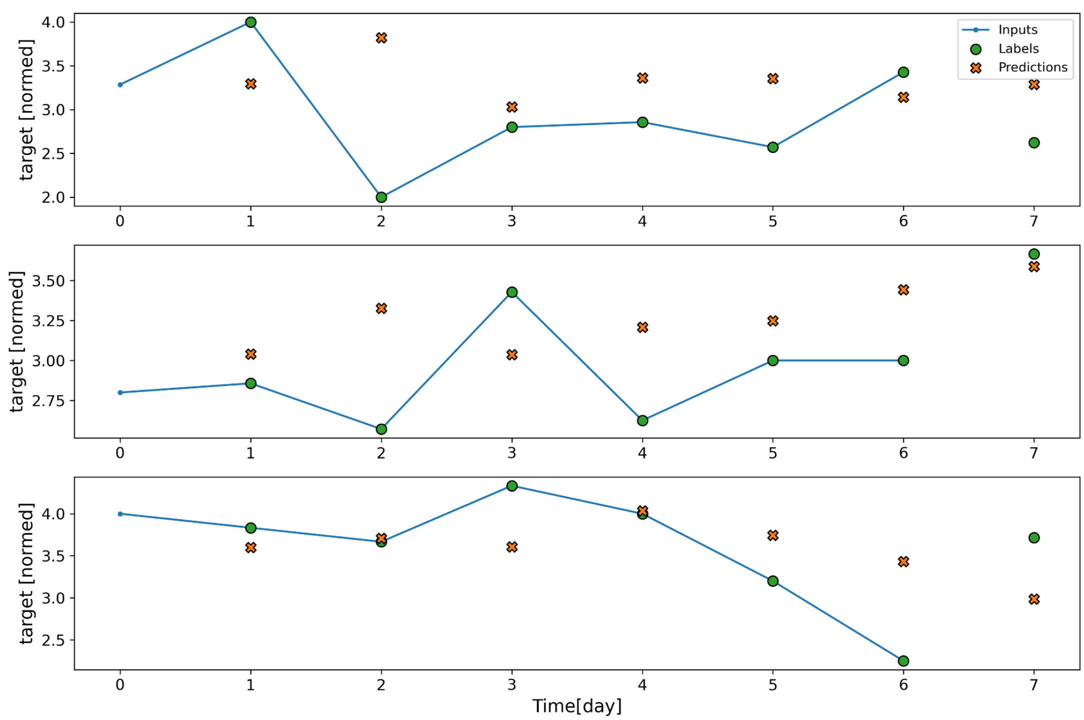

Figure 4.Forecasts of an LSTM algorithm in three different time windows

CONCLUSIONS

The findings of the predictive model indicate the application of Machine learning to individual actions and behaviors can predict daily happiness with high accuracy. It also opens up the possibility that the system developed could be made more personalized for the particular user through the data collected each day which could make the system able to suggest certain individual actions that affect user's happiness most.

The findings of the Forecasting model indicate the application of Deep Learning to individual actions and behaviors and forecast one time-step ahead based on a minimum of the 1-week window of data. Further development of features of the system and its use by more people will help us examine the usability and clinical applicability of this approach.

References

Reference to a journal:

1. Ekman, P. (Ed.). (2008). Emotional awareness: Overcoming the obstacles to psychological balance and compassion: A conversation between the Dalai Lama and Paul Ekman. New York: Times Books.
2. Roessler, J., & Gloor, P. A. (2020). Measuring happiness increases happiness. *Journal of Computational Social Science*, *4*(1), 123–146. https://doi.org/10.1007/s42001-020-00069-6

Reference to a book:

Géron, A. (n.d.). *Hands-On machine learning with SCIKIT-LEARN, Keras, and TensorFlow, 2nd edition*. O'Reilly Online Learning. https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/