

Predicting Happiness Index Using Machine Learning Based SVM-kernal

¹KUNCHAM RAMYA, ²THUMMALA MARUTHI RAO, ³MUDI REVANTH KUMAR,
⁴RAPURU SUNEEL, ⁵JATLA PAVAN TEJA

¹Assistant Professor, Dept. Of CSE, ABR College of Engineering and Technology, Kanigiri

^{2,3,4,5} BTech Student, Dept. Of CSE, ABR College of Engineering and Technology, Kanigiri

***Abstract:** For many years there has been a focus on individual welfare and societal advancement. In addition to the economic system, diverse experiences and the habitats of people are crucial factors that contribute to the well-being and progress of the nation. In recent years, device expertise has become an exceptionally popular topic in the technology sector. A variety of investigations are trying to implement this technology to improve the efficiency of government offerings. We use learning systems to investigate functions associated with the happiness index and make some predictions. Using a data set obtained from a survey that collects a random variety of Chinese people's happiness ratings and asking related questions, we analyse the relationship between people's happiness ratings and problem solving in different ways. Algorithms are presented. who answered. We can use the results to make predictions on new data sets that are nearly consistent with the original values. In our test, we took advantage of some key features like earnings, education and fitness. This document presents some new ideas for improving support for e-government.*

***Keywords:** Machine learning, happiness index, data mining, feature selection, classification.*

I. INTRODUCTION

People living around the world are always longing for happiness, which is an eternal truth. To better display the intangible happiness in data, researchers have created the term happiness index, which is used to measure people's feelings and experiences of their survival and development, also meaning people's sense of happiness. From a series of consecutive numbers, we can

measure people's happiness rationally and visually, making it more convenient to investigate people's assessment of their pleasure. That's what the happiness index can bring to us. So why is it essential to investigate the happiness index? What can it do for our society? According to some experts, the happiness index is a significant index to measure social harmony and an inevitable requirement of

the scientific outlook on development. Since happiness is a scorching topic in 1 people's daily lives, it is meaningful to have a deep insight into that for the government to make more beneficial policies and thoughtful improvements for people. When those changes from the research on happiness index become a reality, people will become happier, and the society will become more harmonious, forming a virtuous circle.

Many factors affect people's happiness, such as their income, living conditions, education level, health, and so on. And for different people, their understandings and interpretations of happiness are distinct. For example, an ordinary citizen A may say that happiness means a higher salary and less overtime at night. Another professional manager B might think the happiest thing is that he will make the company a world brand. The meaning of the happiness index is far more than happiness itself. It also includes the external living environment and self-development conditions of the people. For example, SARS happened in 2002 once made people's happiness index drop.

The successful launch of Shenzhou-5 in 2003 raised the happiness index again; people living in a city with low air pollution are relatively happier than those

living in urban areas with high air pollution. Those make the happiness index an intriguing topic to be investigated. In many past studies on the happiness index, researchers mainly focus on the factors influencing the happiness index and the significance of happiness index for the entire society. In our research, we not only analyse the relationship between the happiness index of some people and some factors but also predict the happiness index of other people through the analysis by utilizing machine learning, including using some models like Linear Regression, Decision Tree, Random Forest, Gradient Boosting. We first pre-process the data, then apply several models, and then analyze, and finally come to a conclusion.

II. REVIEW OF LITERATURE

This section briefly explains the related work in this project. Firstly, the national happiness analysis is described. It will discuss the importance of happiness analysis. Secondly, the used machine learning is introduced. Lastly, the proposed factor analysis is discussed.

A) World Happiness Analysis

The work mentioned that happiness could be a good indicator for how well a society is doing. This becomes important because Betham said that the best society is the one

where the citizens are happiest. Several researches have been conducted on positive aspects and the matters of happiness in policy making. As mentioned in the previous section, happiness can be determined based on various factors. Unfortunately, these factors were analysed manually. The complexity of the factor leads to the expensive cost of analysis. Therefore, the automatic analysis is needed.

B) Support Vector Machine

Due to its capability to learn from the past experiences, machine learning has been used in various areas. Support vector machine is one of the powerful machine learning algorithms. Support Vector Machine (SVM) is a learning technique which is used for classifying unseen data correctly. It is a learning technique which usually used for classifying the unseen data correctly. This technique has been used in various research field due to its remarkable performance. In order to perform the classification task, support vector machine builds a hyper plane which separates the data into different categories. One of the important advantages of support vector machine is its ability to handle the scarcity of the data. Moreover, support vector machine is able to learn about the complex decision boundaries in the high dimensional feature space

efficiently. Due to the complex features used to predict the national happiness, it is important to apply the technique with ability to handle the complex features.

C) Factor Analysis

As mentioned in the previous section, there are a large number of features used to predict the world happiness. However, some of the features may have no significant contribution to the prediction. Therefore, it is unwise to use the entire features to analyse the happiness. This work uses factor analysis to analyse the related features. Factor analysis aims to determine the contribution of a certain feature. This technique does not focus on dimension reduction. Therefore, there will be no features removed. The works in [10] and [11] have introduced the advantages of factor analysis. The first advantages mentioned is the ability to identify latent Dimensions or constructs that cannot be done using direct analysis. Moreover, this approach is easy to run and inexpensive in term of resources.

III. PROPOSED SYSTEM

The aim of this project is to predict the world happiness of a particular country using machine learning techniques. The proposed approach contains four main steps in the data mining process which are

data collection, data pre-processing, data analysis, and classification process as seen in Figure 1. The first process in the process approach is collecting the data. The data used in this project are gathered from the UN Human Development Project. The data contains of the human development index, GDI, healthy index of each country in the world. However, these data are quite dirty. It cannot be used directly as the input data for the learning process. Therefore, the second process is data pre-processing. Data pre-processing is used to increase the data quality. By increasing the quality of data results to the increasing number of prediction accuracy and consistency. The processes included in this process are data cleaning and data integration. The routine processes that should be done are filling the missing values, reduce the noise and identify the outliers



Fig.1 The proposed approach

The third process in the proposed approach is data analysis. This explanatory data analysis is used for finding the relationship among the attributes of the features. This

analysis is done by visualizing the data. Dimensional reduction also be done in this step. Using the information gain technique, the features can be reduced. Information gain technique is used because it can explore the interrelationships among a set of variables. The last part is this work is the classification process. In this classification process, SVM technique is used to predict the happiness of the data based on the important features. The validation process using k-fold cross validation technique is used to measure the performance of the data based on the accuracy, sensitivity and specificity values.

IV. RESULT AND ANALYSIS

This section discusses about the result of each step in the proposed approach. Moreover, this section also presents the analysis of significant factors to determine the happiness of a particular country.

A. Data Collection

As mentioned in the proposed approach section, the data used for this work are gathered from the UN Development Project. In total there are 187 countries listed in the data. Different types of factors are also mentioned in this data, such as human development index, education, environment, health care. The data consists of 105 types of features from 14 different

factors. The dataset contained no missing values.

B. Data Pre processing

In order to increase the data quality, the data preprocessing is needed. The collected data are scattered in various tables. Therefore, creating an integrated data is needed. The single integrated table consists of the entire features and sample that we are going to use in the next process. The dataset contained no missing values. The values within the dataset were present on a mixed scale, in the form of ratios percentage and average scores. The data was normalized using the `min.max()` normalization function in R. We used a realistic dataset so we did not have to consider outliers of the dataset and simulated our work using the original dataset [11].

C. Feature Selection

Selecting the related features is important in order to improve the performance of the classifier. In order to perform this process, WEKA package for attribute selection is used [12]. The evaluator used in this work is information gain. This technique is chosen due to its ability to measure the amount of information in bits about the class prediction [13]. Therefore, it

measures the expected reduction in entropy

D. Classification

In order to evaluate the selected attribute, this work also runs the classification using the entire gathered attributes. The same parameter used to compare both scenarios. The kernel for SVM is chosen based on cross validation. Table 1 shows the result for comparing different types of kernel. We can see that the normalized poly kernel gave an outperform result. Therefore, this kernel is chosen for this work

Table.1 Comparison table

| Kernel Type | Accuracy Rate |
|-----------------------|----------------------|
| Normalize Poly Kernel | 68.456 % |
| Poly Kernel | 60.402 % |
| RBF Kernel | 38.926 % |
| String Kernel | 43.624 % |

As mentioned before, this work also runs the classification using the entire attributes in order to validate the performance of selected attributes. Using the entire attributes we can see that the classification process result in 56.25% of accuracy. This result shows the improvement of accuracy rate and true positive rate. It also shows that using the selected attribute is able to reduce the mean square error. It means that the selected attributes have strong

correlation with the class attribute that can be used to predict the class which is the happiness of the country.

E. Analysis

Instead of classifying the data into happy and unhappy countries, this work classified the data into three categories which are happy, mid, and unhappy. Based on the classification results, it shows that most countries in the world are not in a happy state. As seen in Figure 4, 39 %, 38%, 23% of the countries are happy, mid-happy, and unhappy, respectively

In order to analyze the happiness factor, this work also shows the distribution of the happiness based on the country. A country with red shade is a country in an unhappy state such as Russia and Nigeria. Moreover, the country with yellow shade is mid-state country such as the United State and Australia. Lastly, the country with green shade has happy state such as Brazil and Indonesia. This figure showed one surprising fact that even though a country is developed, it does not mean that it has a happy state.

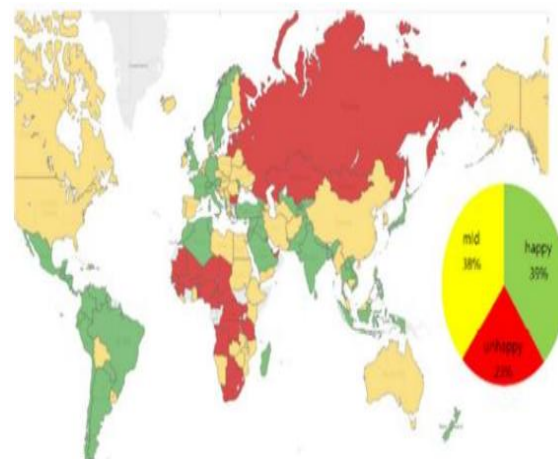


Fig.2 Distribution of Country Happiness

V. CONCLUSION

In the current work, a supervised two-tier ensemble approach for predicting a country's BLI score was proposed. The work presented a cost-effective method of BLI prediction with a high degree of efficiency. The dataset consisted of four different files for 2015 to 2019. The capability of the model to predict life satisfaction relied on the proper training features, chosen using a recursive elimination method with 10-fold cross-validation. The work combined three of the top four models, with simple averaging, to enhance the performance of the regression. The model was built using an ensemble approach and was evaluated using r , R , RMSE, and accuracy performance evaluators. The empirical relevance of diversity estimates was assessed with regard to combining the regression models by stacking. The model

was about 90% accurate for predicting the life satisfaction score of a country. The present study was the first step towards forecasting the BLI score using machine learning based regression model that can influence the survival of future generations and further aid the immigration process. The work can be extended by tuning the parameters of the base models using metaheuristic approaches to improve the prediction accuracy

REFERENCES

- [1] Helliwell, John F., Richard Layard, Jeffrey Sachs, and Jan-Emmanuel De Neve, eds. 2020. *World Happiness Report 2020*. New York: Sustainable Development Solutions Network Miao Yuanjiang.
- [2] Blanchflower, David G., and Andrew J. Oswald. *Happiness and the human development index: The paradox of Australia*. No. w11416. National Bureau of Economic Research, 2005. [3] Frey, Bruno S., and Alois Stutzer. "What can economists learn from happiness research?." *Journal of Economic literature* 40.2 (2002): 402- 435.
- [4] Frey, Bruno S., and Alois Stutzer. *Happiness and economics: How the economy and institutions affect human well-being*. Princeton University Press, 2010.
- [5] Silcock, Rachel. "What is e-government." *Parliamentary affairs* 54.1 (2001): 88-101
- [6] Diener, E., Lucas, R. E., Schimmack, U., & Helliwell, J. *Wellbeing for public policy*. New York: Oxford University Press, 2009.
- [7] Dolan, Paul, and Mathew P. White. "How can measures of subjective well-being be used to inform public policy?." *Perspectives on Psychological Science* 2, no.1 pp .71- 85.2007.
- [8] Viinamäki, H., Kontula, O., Niskanen, L., & Koskela, K. "The association between economic and social factors and mental health in Finland." *Acta Psychiatrica Scandinavica* 92, no. 3, pp.208-213, 1995.
- [9] Malhotra, R., & Jain, A. "Software Effort Prediction using Statistical and Machine Learning Methods." *International Journal of Advanced Computer Science and Applications* 2., pp. 1451-1521, 2011.
- [10] Garson, G. David, "Factor Analysis," from *Statnotes: Topics in Multivariate Analysis*. [11] Tucker, L. R., & MacCallum, R. C.. "Exploratory factor analysis." Unpublished manuscript, Ohio State University, Columbus, 1997
- [11] Prasadu Peddi (2015) "A machine learning method intended to predict a

student's academic achievement", ISSN:
2366-1313, Vol 1, issue 2, pp:23-37.