5-2018

Role of Decision Making in Predicting Health Behavior

Neetu Singh
University of Illinois at Springfield, nsing2@uis.edu

Apoorva Kanthwal
University of Illinois at Springfield, akant2@uis.edu

Prashant Bidhuri
University of Illinois at Springfield, pbidh2@uis.edu

Anusha Vijaykumar Munnolli
University of Illinois at Springfield, amunn3@uis.edu

Role of Decision Making in Predicting Health Behavior

Neetu Singh
University of Illinois at Springfield
nsing2@uis.edu

Apoorva Kanthwal
University of Illinois at Springfield
akant2@uis.edu

Prashant Bidhuri
University of Illinois at Springfield
pbidh2@uis.edu

Anusha Vijaykumar Munnolli
University of Illinois at Springfield
amunn3@uis.edu

ABSTRACT

The annual healthcare expenditure of United States is increasing at a tremendous rate. The decisions made by people concerning their health enforces the government's involvement in encouraging better health behavior through predefined strategies. The poor health behavior often leads to an unhealthy community which further leads to a poor health nation. The CDC administers Behavioral Risk Factor Surveillance System (BRFSS) survey annually. This paper aims to curb the poor decision making in health behavior utilizing 2016 SMART BRFSS data. Using data mining approach, predictive model is developed by analyzing the pattern of decision making of an individual and how it affects the health behavior. This model can aid future BRFSS surveys in better understanding the responses and provide an insight into the factors affecting these decisions. Further, the model will help in targeting poor decision makers to focus on their health behavior and prevent these decisions leading to major illnesses.

Keywords

Health behavior, decision making, public health, data mining.

INTRODUCTION

The annual healthcare expenditure of United States (US) is increasing at a tremendous rate. In 2014, the health expenditure increased by 5% as compared to 2013 and in 2015 the healthcare expenditure further increased by 17.8% in US (National Center for Health Statistics, 2016). The decisions made by people concerning their health is often questionable which enforces the government's involvement in encouraging improved health behavior for better public health through predefined strategies. The poor health behavior often leads to unhealthy community which further leads to a poor health nation. Health condition of every individual is mostly a direct consequence of behavior (Schwarzer, 2008). The decisions people make about their health, results in prevention or an outcome of a health issue, and are so critical that it directly affects morbidity and mortality rates (Sheeran, Klein, & Rothman, 2017). People recognize their health needs, are aware of them, but fail to act upon those needs (Rothman et al., 2015). Reports are generated regularly to derive new strategies for better public health, the most common strategy being interventions. Interventions are hopeful messages assuming people will be motivated by reading or hearing them. But the fear of terminal illnesses have proven to lead to poor health behaviors (McErlean & Fekete, 2017). Hence, it raises the question of what is the right approach to influence people's attitude towards positive health behavior. Some studies have even suggested experimental medicines as an approach as a form of intervention to influence a change in health behavior (Sheeran et al., 2017). Health behavior has been theorized to the point where the practical influence of such theories did not come to fruition. This is where we believe that prior studies in health behavior have lacked in developing defined models based on the vast data that is available.

The Centers for Disease Control and Prevention (CDC) administers a Behavioral Risk Factor Surveillance System (BRFSS) survey annually intending to gauge people's behaviors in the United States. The BRFSS survey is used for the Selected Metropolitan/Micropolitan Area Risk Trends (SMART) to "provide prevalence rates for selected conditions and behaviors" (SMART, 2017). Surveys such as BRFSS offer new direction into the research in health behavior and the decisions impacting people's health. Using the 2016 SMART BRFSS survey data, this paper aims to curb the poor decision making in health behavior by developing a predictive model using data mining approach. The model is developed by analyzing the pattern of decision making by an individual and how it affects the health of the person. This model can most likely aid future BRFSS surveys in better understanding the responses and provide an insight into the factors affecting these decisions. Further, the model will help in targeting poor decision makers to focus on their health behavior and prevent these decisions leading to major illnesses.

The paper is organized as follows. In the next section, we examine the existing studies in the field of health behavior and decision making. After conducting the literature review, data mining is performed to develop our predictive model. Finally, the conclusion section discusses the contribution, limitation and future research.

LITERATURE REVIEW

Health decisions made by individuals play a very crucial role in determining their health outcomes. The decision making is also affected by the extent to which patients are aware of their health condition, and what efforts they make to protect their health (Coulter, Parsons, & Askham, 2008). Individuals who have good health outcomes are the ones who are well aware of their health condition and actively take part in deciding ways to improve it (Levinson, Kao, Kuby, & Thisted, 2005). Health behavior has been well defined by Kasl & Cobb, (1966) as "any activity undertaken by a person who believes himself to be healthy for the purpose of preventing disease or detecting disease in an asymptomatic stage". In addition, there are many other theories for health behavior that have been developed, one of those consider health behavior as patterns and actions that might have good or adverse effect on maintenance, or improvement of health (Gochman, 1997).

While concepts of health have often been related with the study of illness and its management and care aspects (Millstein & Irwin, 1987), few studies have focused on decision making of individuals towards their health conditions. Most health related problems arise from poor behavior such as indulging in bad drinking habits, smoking, physical inactivity, and substance use (Jensen et al., 2011; Kahn et al., 2002; Prochaska & Velicer, 1997; Schwarzer, 2008).

The SMART BRFSS data, considered in our research, has been used to see the growing effect of social media activities such as Facebook Likes to predict the county wise mortality, diseases, or lifestyle habits in United States (Gittelman et al., 2015). Furthermore, the data has been used to study whether variation in local health led to health disparities depending on demographics or ethnicity (Shah, Whitman, & Silva, 2006). Similar studies have established cause-effect relation between health conditions and health behavior (Chunara, Bouton, Ayers, & Brownstein, 2013; McGuire, Ford, & Okoro, 2007; Pucher, Buehler, Bassett, & Dannenberg, 2010). This research focuses on the health decisions of individuals by understanding a pattern in their health conditions.

Considering the size of the BRFSS survey, the data has been used in multiple studies (Leslie, Frankenfeld, & Makara, 2012; Nandi, Charters, Strumpf, Heymann, & Harper, 2013) and has even come under the scanner for its reliability and validity (Pierannunzi, Hu, & Balluz, 2013). This is necessary as well as important as 1,387 articles were identified that have used the BRFSS survey data for research from 1984 to 2012, out of which, 84.2% articles were used within the last 10 years in that time period (Khalil & Gotway Crawford, 2015). Behavioral research has been a constant topic of study using the BRFSS dataset (Khalil & Gotway Crawford, 2015) but not many models have been developed to predict or classify behavior, and where models have been developed, they have focused on a single condition or a category of responses (Dwyer-Lindgren et al., 2015; Frazier, Harwell, Loveland, Zimmerman, & Helgerson, 2011; Michimi & Wimberly, 2015). In this paper we develop a predictive model that can provide a better understanding into the decision making of individuals affecting their health.

METHODOLOGY

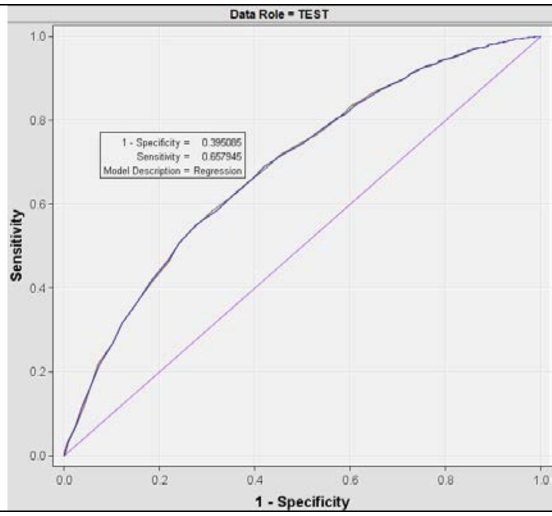
In this study we are using SMART BRFSS survey dataset for the year of 2016 (SMART, 2017) and develop a predictive model to understand the decision making of individuals regarding their health conditions. Data mining is performed using the SEMMA (Sample, Explore, Modify, Model, Assess) approach of SAS® Enterprise Miner to get preliminary results supporting the analysis using logistic regression as an initial model. The data set has been coded using conditional functions referring to the codebook of BRFSS data (Centers for Disease Control and Prevention, 2017) to subset and transform important information inclusive for analysis. We intend to understand an individual's decision making by determining the factors influencing their decision of going for a medical checkup. We used logistic regression on our data to analyze individual's physical health condition, general health condition, mental health condition, poor health condition, medical cost involved in checkups, education level attained, income, asthma, exercise, health plan, and smoking habits to determine if they would go for a medical checkup. Data is partitioned into training (70%) and test (30%) data set to develop the preliminary model and to evaluate the performance of the model. Exhaustive logistic regression in addition with forward, backward and stepwise regression are executed to see significance of all the conditions with respect to the target variable, CHECKUP2, which represents how long it has been since the last visit to a doctor for a routine checkup. The model comparison is preformed to determine the best model out of all four regressions that would be useful to conduct further extensive analysis. This will help to understand whether people throughout United States are making good health decisions.

The target variable (CHECKUP2) has two categories (0, 1) where "1" represents whether an individual had medical checkup within last year and "0" represents no medical checkup in last one year. The maximum likelihood estimates and model comparison including the significant variables for determining the individual's decision making patterns are shown in Figure 1. In addition, there was no overfitting of data. Exhaustive regression was chosen by SAS® Enterprise Miner as the best model with accuracy of 80.70% and misclassification rate 19.30% (Figure 1). The Type 3 Analysis of Effects were also analyzed to see the significance of input variables. It was observed that the input variables asthma, health plan, external (leisure) exercise, medical cost, poor health, income, physical health, general health, and smoking habits were significant as shown in Figure 1.

Parameter	**Label**	**β (Estimate)**	**S.E.**	**Wald**	**Sig.**	**Exp(β)**
ASTHNOW	Still has asthma	.1481***	.00261	32.12	<0.0001	1.160
EXERANY2	Leisure exercise	-.0747*	.0292	6.56	<0.0104	.928
HLTHPLN1	Healthcare coverage	.5720***	.0433	174.69	<0.0001	1.772
MEDCOST1	No visit to doctor because of cost	-.4373***	.0301	210.48	<0.0001	.646
POOR14D1	Poor health (0 days)	.1944***	.0488	15.90	<0.0001	1.215
POOR14D2	Poor health (1-13 days)	-.1619***	.0445	13.25	<0.0003	.851
_EDUCAG1	Education (attended high school)	-1.3525	3.2964	.17	<0.6816	.259
_EDUCAG2	Education (graduated high school)	-1.3203	3.2959	.16	<0.6887	.267
_EDUCAG3	Education (attended college)	-1.4513	3.2959	.19	<0.6597	.234
_EDUCAG4	Education (graduated college)	-1.4192	3.2959	.19	<0.6668	.242
_INCOMEG1	Income (<15k)	.1846**	.0659	7.84	<0.0051	1.203
_INCOMEG2	Income (15k-25k)	.1305*	.0570	5.24	<0.0220	1.139
_INCOMEG3	Income (25k-35k)	.00148	.0712	.00	<0.9834	1.001
_INCOMEG4	Income (35k-50k)	-.0380	.0651	.34	<0.5596	.963
_INCOMEG5	Income (>50k)	-.2593***	.0471	30.27	<0.0001	.772
_MENT14D1	Mental health (good)	.1564***	.0401	15.21	<0.0001	1.169
_MENT14D2	Mental health (fair)	-.0629	.0349	3.25	<0.0713	.939
_PHYS14D1	Physical health (good)	-.1493***	.0451	10.94	<0.0009	.861
_PHYS14D2	Physical health (fair)	.0197	.0361	.30	<0.5846	1.020
_RFHLTH1	Good or better health	-.1476***	.0324	20.80	<0.0001	.863
_RFSMOKE31	Does not smoke	.1785***	.0308	33.50	<0.0001	1.195
Model Comparison						
Selected Model	Model Description	Selection Criteria: Misclassification Rate			Train: ROC index	Test: ROC index
Y	Regression	0.1930			0.703	0.685
	Backward Regression	0.1934			0.702	0.686
	Forward Regression	0.1934			0.702	0.686
	Stepwise Regression	0.1934			0.702	0.686
Significance: *p <0.05, **p<.01, ***p<.001						

Figure 1. Maximum Likelihood Estimates and Model Comparison

The receiver operating characteristic (ROC) curve and confusion matrix, as shown in Table 1, are used for performance evaluation of models and to identify the best model (Barron, 1977). To further see the performance of the model, misclassification rate (19.30%) of the best model is compared with baseline misclassification rate (20.35%). The fit statistics for individual models were checked to see the misclassification rate for training and test data to ensure that there is no overfitting. The same was also checked by analyzing the cumulative lift for training and test data for each model. To further our analysis, we scored our model.



False Negative	True Negative	False Positive	True Positive
194	310	2158	9526

Table 1. ROC Curve and Confusion Matrix for best model

From this predictive model, we could infer that asthma, exercising, health plans, medical cost, poor health, income, mental health, physical health, general health, and smoking habits are the influencing factors of individuals in determining their health decision (to go for a medical checkup or not). In addition, the model helps to identify the patterns of decision making by considering these individual factors. We intend to extend this analysis by using additional data mining techniques (Neural Network, Decision Tree and *K*-Nearest Neighbor) to obtain the best classification model. This will assist in predicting the health outcomes depending on the health decisions made by individuals. It will further help to analyze whether people are making good decisions to have better health outcomes.

CONCLUSION

The purpose of this study is to generate health awareness through better understanding of health decisions. In this study we observed that 74% of people had asthma and 18% of which did not go for a checkup. Furthermore, it was found that 15% suffering from asthma and had health plans, did not go for a health checkup. This leads to the observation that people with health plans and poor health conditions do not prefer visiting a doctor indicating poor health decisions. Additionally, we observed 18% people with smoking habits did not go for health checkup and 4% of these expressed medical cost as a factor. The analysis also reflected bad decision making for 19% with poor physical health. Based on our analysis it is evident that there is a need for better health decision making in individuals. This necessitates enrichment of health awareness programs and development of applications that enhance decision making. We are aiming to use other data mining techniques to develop a model that can best validate the significant variables and provide a new direction to extend research on BRFSS data. The conclusion of this study will give a clear understanding into the factors affecting health decisions which in turn will help future surveys to have better questionnaires. The survey data has many variables that have not been used, but this study can encourage future replication of the developed model on other variables to identify more factors influencing health decisions.

REFERENCES

1. Barron, B. A. (1977). The Effects of Misclassification on the Estimation of Relative Risk. *Biometrics*, *33*(2), 414–418. https://doi.org/10.2307/2529795
2. Centers for Disease Control and Prevention. (2017). *LLCP 2016 Codebook Report*. Retrieved from https://www.cdc.gov/brfss/annual_data/2016/pdf/codebook16_llcp.pdf
3. Chunara, R., Bouton, L., Ayers, J. W., & Brownstein, J. S. (2013). Assessing the Online Social Environment for Surveillance of Obesity Prevalence. *PLOS ONE*, *8*(4), e61373. https://doi.org/10.1371/journal.pone.0061373
4. Coulter, A., Parsons, S., & Askham, J. (2008). *Where are the patients in decision-making about their own care?* (pp. 1–26). Europe: WHO. Retrieved from http://apps.who.int/iris/bitstream/10665/107980/1/E93419.pdf
5. Dwyer-Lindgren, L., Flaxman, A. D., Ng, M., Hansen, G. M., Murray, C. J. L., & Mokdad, A. H. (2015). Drinking Patterns in US Counties From 2002 to 2012. *American Journal of Public Health*, *105*(6), 1120–1127. https://doi.org/10.2105/AJPH.2014.302313
6. Frazier, J. C., Harwell, T. S., Loveland, K. M., Zimmerman, H. J., & Helgerson, S. D. (2011). Prevalence of Asthma Among Adults in Metropolitan Versus Nonmetropolitan Areas in Montana, 2008. *Preventing Chronic Disease*, *9*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3277405/
7. Gittelman, S., Lange, V., Gotway Crawford, C. A., Okoro, C. A., Lieb, E., Dhingra, S. S., & Trimarchi, E. (2015). A New Source of Data for Public Health Surveillance: Facebook Likes. *Journal of Medical Internet Research*, *17*(4). https://doi.org/10.2196/jmir.3970
8. Gochman, D. S. (Ed.). (1997). *Handbook of Health Behavior Research I: Personal and Social Determinants*. Springer US. Retrieved from //www.springer.com/us/book/9780306454431
9. Jensen, C. D., Cushing, C. C., Aylward, B. S., Craig, J. T., Sorell, D. M., & Steele, R. G. (2011). Effectiveness of motivational interviewing interventions for adolescent substance use behavior change: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, *79*(4), 433–440. https://doi.org/10.1037/a0023992
10. Kahn, E. B., Ramsey, L. T., Brownson, R. C., Heath, G. W., Howze, E. H., Powell, K. E., … Corso, P. (2002). The effectiveness of interventions to increase physical activity: A systematic review. *American Journal of Preventive Medicine*, *22*(4, Supplement 1), 73–107. https://doi.org/10.1016/S0749-3797(02)00434-8
11. Kasl, S. V., & Cobb, S. (1966). Health Behavior, Illness Behavior and Sick Role behavior. *Archives of Environmental Health: An International Journal*, *12*(2), 246–266. https://doi.org/10.1080/00039896.1966.10664365
12. Khalil, G. M., & Gotway Crawford, C. A. (2015). A Bibliometric Analysis of U.S.-Based Research on the Behavioral Risk Factor Surveillance System. *American Journal of Preventive Medicine*, *48*(1), 50–57. https://doi.org/10.1016/j.amepre.2014.08.021
13. Leslie, T. F., Frankenfeld, C. L., & Makara, M. A. (2012). The spatial food environment of the DC metropolitan area: Clustering, co-location, and categorical differentiation. *Applied Geography*, *35*(1), 300–307. https://doi.org/10.1016/j.apgeog.2012.07.008
14. Levinson, W., Kao, A., Kuby, A., & Thisted, R. A. (2005). Not All Patients Want to Participate in Decision Making. *Journal of General Internal Medicine*, *20*(6), 531–535. https://doi.org/10.1111/j.1525-1497.2005.04101.x
15. McErlean, A. R., & Fekete, E. M. (2017). Health-Related Social Control and Risk Perceptions in Emerging Adults. *Emerging Adulthood*, 2167696817729577. https://doi.org/10.1177/2167696817729577
16. McGuire, L. C., Ford, E. S., & Okoro, C. A. (2007). Natural disasters and older US adults with disabilities: implications for evacuation. *Disasters*, *31*(1), 49–56. https://doi.org/10.1111/j.1467-7717.2007.00339.x
17. Michimi, A., & Wimberly, M. C. (2015). The food environment and adult obesity in US metropolitan areas. *Geospatial Health*, *10*(2). https://doi.org/10.4081/gh.2015.368
18. Millstein, S. G., & Irwin, C. E. (1987). Concepts of health and illness: different constructs or variations on a theme? *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *6*(6), 515–524.
19. Nandi, A., Charters, T. J., Strumpf, E. C., Heymann, J., & Harper, S. (2013). Economic conditions and health behaviours during the 'Great Recession.' *J Epidemiol Community Health*, *67*(12), 1038–1046. https://doi.org/10.1136/jech-2012-202260
20. National Center for Health Statistics. (2016). *Health, United States, 2015: With Special Feature on Racial and Ethnic Health Disparities*. Hyattsville (MD): National Center for Health Statistics (US). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK367640/
21. Pierannunzi, C., Hu, S. S., & Balluz, L. (2013). A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Medical Research Methodology*, *13*, 49. https://doi.org/10.1186/1471-2288-13-49

22. Prochaska, J. O., & Velicer, W. F. (1997). The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion*, *12*(1), 38–48. https://doi.org/10.4278/0890-1171-12.1.38
23. Pucher, J., Buehler, R., Bassett, D. R., & Dannenberg, A. L. (2010). Walking and Cycling to Health: A Comparative Analysis of City, State, and International Data. *American Journal of Public Health*, *100*(10), 1986–1992. https://doi.org/10.2105/AJPH.2009.189324
24. Rothman, A. J., Gollwitzer, P. M., Grant, A. M., Neal, D. T., Sheeran, P., & Wood, W. (2015). Hale and Hearty Policies: How Psychological Science Can Create and Maintain Healthy Habits. *Perspectives on Psychological Science*, *10*(6), 701–705. https://doi.org/10.1177/1745691615598515
25. Schwarzer, R. (2008). Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. *Applied Psychology: An International Review*, *57*(1), 1–29. https://doi.org/10.1111/j.1464-0597.2007.00325.x
26. Shah, A. M., Whitman, S., & Silva, A. (2006). Variations in the Health Conditions of 6 Chicago Community Areas: A Case for Local-Level Data. *American Journal of Public Health*, *96*(8), 1485–1491. https://doi.org/10.2105/AJPH.2004.052076
27. Sheeran, P., Klein, W. M. P., & Rothman, A. J. (2017). Health Behavior Change: Moving from Observation to Intervention. *Annual Review of Psychology*, *68*(1), 573–600. https://doi.org/10.1146/annurev-psych-010416-044007
28. SMART: BRFSS City and County Data and Documentation. (2017, November 13). Retrieved March 6, 2018, from https://www.cdc.gov/brfss/smart/Smart_data.htm