

# Predicting Patients' Satisfaction With Doctors in Online Medical Communities: An Approach Based on XGBoost Algorithm

Yunhong Xu, Kunming University of Science and Technology, China

Guangyu Wu, Kunming University of Science and Technology, China

Yu Chen, Kunming University of Science and Technology, China

## ABSTRACT

Online medical communities have revolutionized the way patients obtain medical-related information and services. Investigating what factors might influence patients' satisfaction with doctors and predicting their satisfaction can help patients narrow their choices and increase their loyalty towards online medical communities. Considering the imbalanced feature of dataset collected from Good Doctor, the authors integrated XGBoost and SMOTE algorithms to examine what factors can be used to predict patient satisfaction. SMOTE algorithm addresses the imbalanced issue by oversampling imbalanced classification datasets. And XGBoost algorithm is an ensemble of decision trees algorithm where new trees fix errors of existing trees. The experimental results demonstrate that SMOTE and XGBoost algorithms can achieve better performance. The authors further analyzed the role of features played in satisfaction prediction from two levels: individual feature level and feature combination level.

## KEYWORDS

Machine Learning, Online Medical Community, Satisfaction Prediction, SMOTE, XGBoost

## INTRODUCTION

With the large-scale popularization of the Internet, the concept of "Internet +" has gradually penetrated into the medical industry, and the online medical communities have emerged (Li et al., 2019). Online medical communities have revolutionized the way patients obtain medical-related information and services. The online medical community enables patients to contact doctors without time and space constraints, thereby making it more convenient to receive medical treatment at a lower cost (Sims, 2016). In addition to convenience and cost, patients can also obtain diverse medical suggestions from different doctors, which allow them to make better medical decisions. Doctors can obtain social and economic benefits from engagement in online medical communities, e.g. social recognition of patients and financial returns.

Considering the benefits patients and doctors can obtain from online medical communities, more and more patients and doctors participate in online medical communities (Li et al., 2018). Under the traditional medical mode, due to lack of professional medical knowledge, patients habitually choose high-level hospitals or doctors with strong medical background (Liu et al., 2018). Medical treatment is a multi-dimensional and complex process. In addition to doctors' medical background,

DOI: 10.4018/JOEUC.287571

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

the way doctors communicate with their patients, how they respond to patients, and other patients' experiences can also influence patients' satisfaction with doctors. However, it is difficult to trace these variables under the traditional medical mode. Online medical community can record several features that occur during or after medical treatment, which provides a holistic framework to predict patient satisfaction with doctors.

Patients' satisfaction with doctors can be defined as the extent to which patients are content with the medical services which they receive from doctors. Understanding what factors that drive patient satisfaction and how these factors can be used to predict patient satisfaction can facilitate patients to make better decisions and promote doctors to provide better medical services. Previous studies have shown that patients' satisfaction with their medical service providers tends to be more positive, that is, patients are more willing to give doctors satisfactory evaluation (Marcinowicz et al., 2009). And data evaluated as dissatisfied may be more important than the satisfaction evaluation data when studying patients' satisfaction (Vaida & Osmo, 2003). Therefore, we should consider both satisfaction and dissatisfaction data when predicting patients' satisfaction. When both kinds of data are considered, there are 6933 satisfactory samples and 77 unsatisfactory samples in our dataset collected from Good Doctor. The data is imbalanced in that the samples classified as positive (which means patients who are not satisfied with their doctors) and negative (which means patients who are satisfied with their doctors) are not equally distributed. Particularly, positive samples account for about 1.1% of the total sample. Data imbalance would cause the performance of machine learning algorithms to degrade, because most machine learning algorithms will ignore or have poor performance on the minority class (Wang et al., 2020). There are two ways to deal with imbalanced dataset: data resampling and algorithms based on ensemble learning. For data resampling, the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) is used in this research to synthesize new samples from the minority class. XGBoost is a large-scale ensemble algorithm proposed by Chen et al. (2016), which performs a second-order Taylor expansion on the cost function, where both the first-order derivative and the second-order derivative are used. Previous research has demonstrated that SMOTE and XGBoost algorithm can achieve better results when confronted with imbalanced datasets (He et al., 2021; Meng et al., 2020; Wang et al., 2020). In this research, we integrate XGBoost and SMOTE to predict patients' satisfaction with doctors.

## **LITERATURE REVIEW**

Patient satisfaction with doctors is a multi-dimensional concept, which should be operationalized and considered under particular contexts (Naidu, 2009). In offline medical context, questionnaire was used to measure patient satisfaction with doctors. Hussain et al. (2019) found that doctor services and waiting time can influence patient satisfaction. Alharbi et al. (2018) reviewed relevant research and concluded that availability and accessibility, relational conduct, technical skills/knowledge, and personal qualities would affect patient satisfaction. Prado-Galbarro et al. (2020) found that a worse perception of health status and a longer time spent in the waiting room were negatively associated with patient satisfaction. Sl et al. (2020) found that while waiting time negatively influences patient satisfaction, positive communication experiences with health care providers could potentially alleviate negative effects of long waiting time. Boquiren et al. (2015) conducted a literature review and identified five dimensions that can determine patient satisfaction with doctors, including (a) communication attribute, (b) related conduct, (c) technical skill, (d) personal qualities, and (e) availability and accessibility. In comparison to research in offline medical context, the research on patient satisfaction with doctors in online medical community is still preliminary. The depth of an interaction and service content during a first consultation influence patients' subsequent consultation behaviors and satisfaction (Yang et al., 2019).

Since the research on patient satisfaction in online medical community is limited, we reviewed literature on doctor recommendation. Although doctor recommendation and patient satisfaction are

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/predicting-patients-satisfaction-with-doctors-in-online-medical-communities/287571](http://www.igi-global.com/article/predicting-patients-satisfaction-with-doctors-in-online-medical-communities/287571)

## Related Content

---

### Mining Potential Requirements by Calculation of User Operations

Yanqun Huang, Gaofeng Pan, Xu Li, Zhe Sun, Shinichi Koyama and Yanqun Yang (2021). *Journal of Organizational and End User Computing* (pp. 1-14).

[www.irma-international.org/article/mining-potential-requirements-by-calculation-of-user-operations/293289](http://www.irma-international.org/article/mining-potential-requirements-by-calculation-of-user-operations/293289)

### Supporting Large-Scale End User Specification of Workflows, Work Coordination and Tool Integration

John Grundy, John Hosking and Warwick Mugridge (1998). *Journal of End User Computing* (pp. 38-48).

[www.irma-international.org/article/supporting-large-scale-end-user/55753](http://www.irma-international.org/article/supporting-large-scale-end-user/55753)

### Characterizing Data Discovery and End-User Computing Needs in Clinical Translational Science

Parmit K. Chilana, Elishema Fishman, Estella M. Geraghty, Peter Tarczy-Hornoch, Fredric M. Wolf and Nick R. Anderson (2011). *Journal of Organizational and End User Computing* (pp. 17-30).

[www.irma-international.org/article/characterizing-data-discovery-end-user/58545](http://www.irma-international.org/article/characterizing-data-discovery-end-user/58545)

### VDT Health Hazards: A Guide for End Users and Managers

Carol Clark (2001). *Journal of Organizational and End User Computing* (pp. 34-39).

[www.irma-international.org/article/vdt-health-hazards/3733](http://www.irma-international.org/article/vdt-health-hazards/3733)

### Do Spreadsheet Errors Lead to Bad Decisions? Perspectives of Executives and Senior Managers

Jonathan P. Caulkins, Erica Layne Morrison and Timothy Weidemann (2009). *Evolutionary Concepts in End User Productivity and Performance: Applications for Organizational Progress* (pp. 44-62).

[www.irma-international.org/chapter/spreadsheet-errors-lead-bad-decisions/18644](http://www.irma-international.org/chapter/spreadsheet-errors-lead-bad-decisions/18644)