# An Intelligent Application for Determine Patient Satisfaction using Data Mining Techniques

**S.Kayalvizhi[1], T.Gunasekar[2], S.Thenmozhi[3]**

[1] *Research Scholar, Department of Sciences, Mahatma Gandhi University, Meghalaya, India*
[2] *Research Scholar, Department of Computer and Information Science, University Technology PETRONAS, Perak, Malaysia*
[3] *Assistant Professor, Department of Electronics & Communication Engineering, NPR College of Engineering & Technology, Tamil nadu, India*

## ABSTRACT

*The healthcare industry today generates large amounts of complex data data about patients, hospital resources, disease diagnosis, electronic patient records and medical devise. Data mining is a collection of algorithmetic ways to extract informative patterns from raw data. In this paper, the researcher describe a novel methodology which employs machine learning as an alternative means to explore hospital characteristics and client satisfaction, for decision making and improved quality of care. Applied well known feature selection and data mining algorithms such as forward selection and Naïve Bayes respectively, to determine patient satisfaction. The dataset comprised of three types of data which are patient perception about received care, Nurse perception about the working environment and Organizational attributes of the hospital. The experimental results exhibited high classification accuracy 91.3%, allowing valid conclusions to be reached about the organizational and workforce factors which attribute to patient satisfaction. The findings were validated using traditional statistical methods such as binomial correlation and linear regression.. The result of the study will be beneficial to current and future SMEs in India.*

***Keywords:*** Machine learning, Native Bayes, Healthcare, Patient Satisfaction.

## 1. INTRODUCTION

Customer satisfaction is considered as a very important metric to evaluate business services, since there are there are moderate-to-strong associations between satisfaction levels and financial and market performance [1]. In the healthcare sector, patient satisfaction is of paramount importance for the assessment of the quality of health services. However, there is no clear literature for a well defined set of patient satisfaction determinants due to the different definitions and reference context various studies investigate, neither at the hospital nor health system level. The patient satisfaction can be considered a decisive metric that can be used by therapists to increase the quality of services which they provide. In a broader sense it is accepted as an important indicator for the quality of healthcare, and may be used by hospitals for improving their services, defining their marketing directives and ultimately increasing patient satisfaction [2].

The patient satisfaction is being used as a quality indicator and is correlated with the quality of care, since it is proven that satisfied patients have higher adherence to the prescribed treatment and improved quality of life among others [3]. The feedback from patients is used by the hospital administration to assess and compare different departments

and institutions, so as to improve the provided services across the organization since regulation agencies use patient satisfaction for accreditation of hospitals [4]. The feedback is also used by clinicians to improve their interaction with the patients.

There are many other factors reported in the literature which may affect patient satisfaction, though without explicit implications about a direct relationship with the quality of provided care. These factors include the interaction with the clinician, number of tests and procedures and boarding time, the availability of services and accessibility of care [5]. Improved work environments and patient to nurse ratio have also been associated with better quality of care and patient satisfaction.

The majority of studies [6], investigate patient satisfaction using traditional statistical analysis methods. There is limited use of data mining techniques to explore healthcare services in relation to patients' perceptions. Comparing traditional methods with data mining though, indicates at least the potential of applying data mining methods to understand patient healthcare preferences.

Data mining can be used to extract meaningful knowledge from big data for making decisions in both the administrative and clinical fields. The challenges faced when using data mining techniques with this type of data is its diversity as well as missing information. In most cases the majority of data is not available in electronic form or may require extensive preprocessing to become usable [7], but this limitation is gradually overcome since most health organizations have already started to integrate electronic systems and digitize their clinical and administrative data.

In this paper the researcher utilize a multi-centric dataset which addresses the perceived patient care, the nurses' perception about their working environment and organizational attributes of the hospital. Apply data mining techniques in order to extract knowledge about the important dimensions of patient satisfaction. Further seek to test the potential of machine learning as an alternative method to explore organizational and employee features which affect the perspectives of customers, in our case, hospitalized patients. The extracted knowledge can be used for decision making and derivation of scientific hypotheses [8]. More specifically, data mining in healthcare can be used for both medical and administrative decision making which in turn improves the services provided.

## 2. PATIENT SATISFACTION AND DECISION MAKING

In this section explore the studies related to patient satisfaction and the interpretation of their outcomes to management strategies for a health-care institution.

### 2.1 Patient Satisfaction

Data concerning the patient treatment in healthcare is useful in understanding the factors that affect patient satisfaction and how personal experience of the patients is related to them. In a study from Hush et al [9] the researchers used least squares regression models to evaluate the correlation between various parameters, such as boarding time, facility issues and communication, and the perceived patient satisfaction, among patients who had received care during the preceding five years. The patient satisfaction proved to be highly correlated to personal experiences as well as expectations, health status and immunization coverage.

Though, various societal factors were hypothesized to play a major role too. In other studies [10], the psychological status and quality of life of cancer patients were used as metrics of satisfaction. Additional studies indicate that the interaction of the therapist with the patient as well as her/is overall experience is an equally determinant that is sometimes more important than the actual outcome of the treatment.

### 2.2 Management and Customer Satisfaction

By modeling the healthcare institution as a service provider organization and the patients as customers, management based strategies can be applied in order to maximize their satisfaction. Under this modeling assumption, different organizations are competing for customers, which in turn make their choice based on their satisfaction. Studies have shown that a more personalized interaction between the customers and the organization lead to a more satisfying experience. In addition, customers have become more demanding concerning the service traits they require, thus increasing the research interest from social scientists and government support [11]. The health care providers in particular use satisfaction related information to define their goals in both medical and administrative aspects. This information is also used to determine and track changes in quality of service and is utilized for accreditations through the use of surveys and customer support and problem solving, leading to improved marketing strategies [12]. As a sum up, information on patient satisfaction helps crystallize the needs of the patients and develop strategies to improve the perceived quality of services.

## 3. RESEARCH METHDOLOGY

In this study, used well known data mining techniques to determine patient satisfaction from the aforementioned multi-centric data. Data mining analyses large data sets to find relationships and summarize the information contained in a meaningful and useful manner. Data mining can be divided in unsupervised learning and supervised learning [13]. The former, clusters data by determining their similarity, helping patterns to emerge. The latter, defines classification rules in the form of models, which are then used to classify new, unknown data. Supervised learning includes classification, regression, time series analysis and prediction and was used in our experiments. The researcher finally used traditional statistical methods such as binomial correlation and linear regression for validation.

### 3.1 Classification

Classification is the process of labeling data using predetermined class labels. A class is an attribute of interest, and the classification process builds a model which assigns the available data to the different available classes. The two phases of this process are training, during which the classification model is built, and testing, where new unknown data are classified. In our system we used various classifiers, namely naïve Bayes, AdaBoost and J48 decision trees.

Naïve Bayes is a probabilistic classifier based on Bayes' theorem. It is called "naïve" due to its inherent assumption of conditional independence between different features. Its main advantages are its simplicity due to the need for only one probability multiplication, which results in low computational complexity. Additionally, it requires a relatively small training set since it only calculates the feature outcome frequency pairs. These characteristics make this classifier capable of handling large data sets very quickly and attain accurate results. Its

main disadvantage springs from the inherent assumption of feature independence. This assumption is often left unsatisfied with real data; however the effect on the classification accuracy may be minimal [14].

Another well known classifier is AdaBoost [15]. It uses the boosting technique, combining a number of weak classifiers with the weighted majority voting scheme. Classifiers with higher accuracy during the training phase acquire higher weights in the voting scheme of the final classifier. AdaBoost has been shown to outperform most classifiers (including SVMs) in the literature when applied on epiluminescence microscopy (ELM) and MRI images.

Finally, has been used the J48 decision tree classifier, which builds a tree structure in a top-down, divide-and-conquer manner [16]. The feature selection method used, divides data into classes, and the most informative features become nodes of the tree. The construction of the tree using "if-then" criteria constitutes the training phase. While its main advantage is the visualization of the decision making procedure, this visualization becomes of lesser value when too many features are used. In such a case, pruning discards the least important nodes. This makes the tree easier to interpret and reduces over-fitting, but also introduces minor inaccuracies.

### 3.2 Feature Selection

Feature selection is the process of determining the usefulness of available features for the current classification problem and discarding redundant and irrelevant features while retaining meaningful ones. Additional features which do not carry useful information for the current classification problem act as noise, reducing accuracy and hindering performance. In order for a subset of features to be selected, various methods may be used. In our implementation we used the subset evaluation method, according to which the optimal subset of features is specified using the greedy forward selection algorithm.

## 4. RESULTS AND DISCUSSION

The researcher has been conducted two sets of experiments. Initially, we wanted to explore the performance of various machine learning algorithms in the task of predicting patient satisfaction. Furthermore, examined the influence of different factors affecting customer satisfaction by applying feature selection in order to rank the attributes that can be available to a hospital's management. The statistical data which have been used were based on healthcare organization, patient and nurse surveys. More specifically, these questionnaires were distributed to nurses and their patients on the basis of a nursing workforce study conducted in hospitals during 2015.

For the classification, has been used the three well known classifiers described in section 3.1, namely naïve Bayes, AdaBoost and J48 decision trees. Followed by split the dataset randomly, using 90% of the data for training and 10% for testing. For determining whether a patient was satisfied or not, the attained accuracy was found to be 86.96% for naïve Bayes, 91.3% using decision trees and 87% for AdaBoost. Although AdaBoost and naïve Bayes outperformed decision trees classification, the results were very consistent. For discerning the most important hospital, employee and patient factors that influence patient satisfaction, employed the following feature selection scheme. The researcher has been used a wrapper in conjunction with the naïve Bayes classifier and the greedy stepwise search method. This

method starts in an arbitrary point in space with any number of features and searches forward or backward until the addition or deletion of features results in lower evaluation.

Using the framework with our data, has been concluded that important factors for patient satisfaction is the medical unit specialty, the patient's length of stay, the technological infrastructure of the hospital, the type of hospital in terms of educational status, the nurses' perception about the quality of provided care and the number of beds in the hospital. On the contrary, the perceived satisfaction of nurses and the burnout factor did not appear to affect the patients' satisfaction since they were ranked last in our model. This may be due to the professionalism and conscientiousness of nurses who are persistent to provide a high level of care regardless the circumstances. Finally, has been found out that the patients' level of education, their perceived care from the healthcare professionals, their views about the hospital environment as well as their opinions about the process followed on discharge may be determinants of patient satisfaction.

In order to validate the obtained results, has been applied traditional statistical methods. Specifically, the researcher performed bivariate correlation analysis using Pearson coefficient to find those variables which were statistically significant with patient satisfaction. As a follow-up performed linear regression analysis and then included in the model those candidate predictors that were found to be correlated with our dependant variable, during the bivariate analysis. As shown in table-2 the predictors have been found to be similar to those indicated by the previously used classification methods. The validation results indicate the potential of data mining techniques to predict organizational attributes in hospitals.

| Features | Average merit |
|---|---|
| Patients care from nurses | 0.015 |
| Patients opinion on hospital environment | 0.011 |
| Patients opinion during discharge | 0.010 |
| Patients level of education | 0.008 |
| Patients care from doctors | 0.029 |
| Hospital unit specializatioin | 0.011 |
| Patient length of stay | 0.010 |
| Technoligical infrastructure of hospital | 0.015 |
| Patient experience while hospitalized | 0.013 |
| Hopital connection with university | 0.015 |
| Nurses' opinion on quality of care | 0.022 |
| Number of beds | 0.013 |

Table 1: Selected features sorted by highest rank

| Predictor | Beta | Sig. |
|---|---|---|
| Patients level of education | 125 | 0.350 |
| Patients care from nurses | 391 | 0.000 |

| | | |
|---|---|---|
| Patients care from doctors | 391 | 0.046 |
| Patients opinion on hospital environment | 199 | 0.003 |
| Patients opinion during discharege | 147 | 0.013 |

Table 2: Predictors of patient satisfactioin in linear regressioin

## 5. CONCLUSIONS

In this paper has been presented various factors affecting patient satisfaction in a hospital setting, from the perspective of a customer service model. The researcher has been used data available from the staff and patients, in order to facilitate decision making based on the better understanding of the determinants of patient satisfaction. For this purpose has been applied of various classification methods which attained high classification accuracy. Furthermore, applied feature selection to deduce conclusions about the importance of various factors, and created a set of those that influence satisfaction the most. The researcher believe that the model will help towards more informed decisions which could focus on the education of patients, the training of healthcare professionals and also on building a supportive and safe hospital environment.

## REFERENCES

1. Williams P, Naumann E, Customer satisfaction and business performance: a firm-level analysis, Emerald, 2011.

2. Lis CG, Rodeghier M, Gupta D, Distribution and determinants of patient satisfaction in oncology: A review of the literature, 2009, Vol. 3, No.3, pp. 287-304.

3. Safran DG, Taira DA, Rogers WH et al, Linking primary care performance to outcomes of care, The Journal of Family Practice, 2008, Vol. 47, pp. 213–220.

4. Draper M, Hill S, The Role of Patient Satisfaction Surveys in a National Approach to Hospital Quality Management, Melbourne, Victoria, Australia: Faculty of Social Sciences and Communications, Royal Melbourne Institute of Technology, 2005.

5. Mpinga EK, Chastonay P, Satisfaction of patients: a right to health indicator? Health Policy. 2011, Vol. 100, No. 2-3, pp.144-50.

6. Aiken LH, Sermeus W, Van den Heede K et al, Patient safety, satisfaction, and quality of hospital care: cross sectional surveys of nurses and patients in 12 countries in Europe and the United States, BMJ, 2012, Vol. 344.

7. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, Hua L, Data mining in healthcare and biomedicine: a survey of the literature, Journal of Medical Systems, 2012, Vol. 36, No. 4, pp. 2431-48.

8. Yoo I, Song M, Biomedical ontologies and text mining for biomedicine and healthcare: a survey, Journal of Computing Science and Engineering, 2008, Vol.2, No.2, pp.109-136.

9. Hush JM, Cameron K, Mackey M, Patient satisfaction with musculoskeletal physical therapy care: a systematic review, Journal of Physical Theraphy, 2011, Vol. 91, No.1, pp.25-36.

10. Feyer P, Kleeberg UR, Steingraber M, Gunther W, Behrens M, Frequency of side effects in outpatient cancer care and their influence on patient satisfaction-a prospective survey using the PASQOC(R) questionnaire, Support Care Cancer, 2008, Vol. 16, pp. 567-575.

11. Cleary PD, McNeil BJ, Patient satisfaction as an indicator of quality care, Inquiry, 2008, Vol. 25, No. 1, pp.25-36.

12. Bergenmar M, Nylen U, Lidbrink E, Bergh J, Brandberg Y, Improvements in patient satisfaction at an outpatient clinic for patients with breast cancer, Acta Oncol, 2006, Vol. 45, pp. 550–558.

13. Han J, Kamber M, Data mining: concepts and techniques, 2nd ed. The Morgan Kaufmann Series, 2006.

14. Kononenko I, Machine learning for medical diagnosis: history, state of the art and perspective, Artifical Intelligence in Medicine, 2001, Vol. 23, pp.89–109.

15. Lopes R, Ayache A, Makni N, Puech P, Villers A, Mordon S et al., Prostate cancer characterization on MR images using fractal features, Medical Physics, 2011, Vol. 38, pp.83–95.

16. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, Hua L, Data mining in healthcare and biomedicine: a survey of the literature, Journal of Medical Systems, 2012, Vol. 36, No. 42, pp. 431-48.