International Journal of Commerce and Management Research
ISSN: 2455-1627, Impact Factor: RJIF 5.22
www.managejournal.com
Volume 3; Issue 3; March 2017; Page No. 08-13

Applying data mining model for analyzing the quality of healthcare services: the empirical study for the central region and highlands of Vietnam

Le Dan, Chau Ngoc Tuan

Faculty of Statistics, Informatics, University of Economics, The University of Danang, Danang city, Vietnam

Abstract
With the support from information technology, data analytic techniques and processes are significantly improved. Consequently, data mining model has been widely introduced and used in data analysis to support decision making. This paper introduces applications of data mining techniques to analyze the factors affecting patients' satisfaction to the quality of healthcare services. The analysis process is performed according to CRISP – DM (Cross - Industry Standard Process for Data Mining). It also uses the Gi-Du Kang & Jeffrey-James model (2004) to test the causal relationships between factors of healthcare service quality and patients' satisfaction. The sample data was collected from 1500 patients using healthcare services in hospitals in the Central region and Highlands of Vietnam in the year 2016. The reliability of the data source was assessed using EFA (Exploratory Factor Analysis); and the parameters of the model were estimated using SEM (Structural Equation Modeling), path analysis and Bootstrap analysis.

Keywords: data mining, CRISP-DM, satisfaction, service quality, healthcare, decision making, AMOS

1. Introduction
Nowadays, there are many unsolved problems in the quality of hospitals' healthcare services and the patients' satisfaction is still quite low in Vietnam. Therefore, The Vietnam Ministry of Health have promulgated many Circulars, Resolutions, Directives to encourage and orient the hospitals to improve the quality of healthcare services to satisfy the highest demands of patients and their families. The Ministry of Health also promulgated the standards to assess the quality of hospitals (The Ministry of Health, 2013a) [8], the circulars of guideline on the implementation of quality assurance on healthcare services at the hospitals (The Ministry of Health, 2013b) [9], the resolutions of guideline on examination processing at the outpatient departments of the hospitals (The Ministry of Health, 2013c) [10], the programs to improve the quality of examination and treatment of healthcare centers to satisfy the demands of patients with individual health insurance (The Ministry of Health, 2011) and the rules of behaviours of hospitals' staffs in healthcare services (The Ministry of Health, 2008).

There are more and more complains from patients and their families to the hospital services due to low quality. Those weaken the doctor-patient relationship and decrease the reputation of health services as well as increase the patient expenses and the costs of healthcare centers. Therefore, it is important for the service providers to improve the satisfaction of patients (Brown & Swartz, 1989) [2]. In analysis of healthcare services quality, it is proven quite complex in theory as well as in practice. The satisfaction depends on the feeling of clients which is affected by gender, age, social composition, experience, diseases, psychology, belief, behavior and the empathy of medical staffs. It is also difficult for the researchers to build up model of assessment and analysis methods. Therefore, it is very necessary and urgent to apply data mining techniques to measure the patients' satisfaction about hospitals' healthcare service. The project results will be the background to support the hospitals to launch resolutions to improve the quality of healthcare services to increase the satisfaction of patients and their families.

2. The CRISP-DM process
The CRISP – DM process was chosen to determine and measure the factors affecting healthcare services' quality. The CRISP – DM was conceived in the end of 1996 by the Daimler-Benz (now is the DaimlerChrysler), Integral Solutions Ltd. (ISL), NCR, and OHRA companies. One year later, the CRISP-DM was developed to become a standard data mining process model that commonly used by data mining experts. To the year 2000, the CRISP – DM was completed and became the leading methodology of the data mining process. The CRISP – DM includes six major phases.

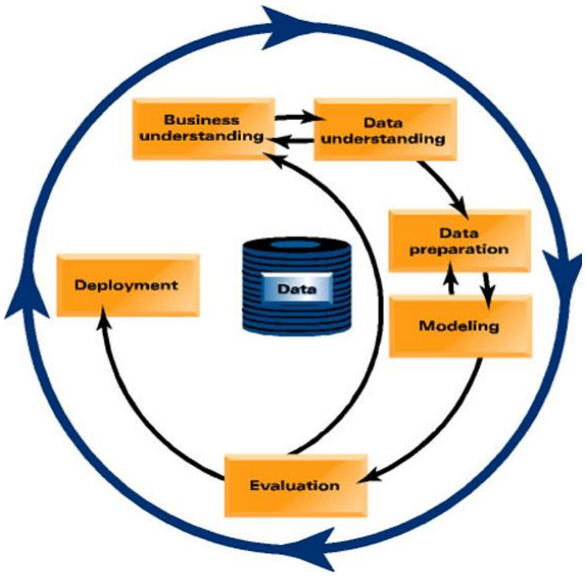

Fig 1: CRISP (Chapman, P. *et al* 2000).

2.1 Business Understanding

This phase focuses on determining the data mining objectives of the project, understanding and assessing the issue, and designing a preliminary plan to measure the factors affecting healthcare services' quality. Then the model was proposed and hypotheses were set up. A proposed model to assess the satisfaction of patients was built up based on the literature review about services, services' quality and the model of services' quality of Gronroos, Parasuraman and especially from Gi-Du Kang and Jeffrey James. According to this model, the satisfaction was affected by three factors including Functional quality, Technical quality and Image. These factors were measured by Likert scale with five levels. Each factor was measured by items.

- Satisfaction (SAT) was measured by six observed variables to reflect the satisfaction of patients: Satisfaction about physicians and doctors (SAT1); Satisfaction about nurses (SAT2); Satisfaction about procedure of examination and treatment of the hospital (SAT3); Satisfaction about the infrastructures of hospital (SAT4); Satisfaction about distributed medicine (SAT5); Satisfaction about the fee of charge to be examined and treated (SAT6).
- Functional quality (FUN) was the method that the hospital provides the healthcare service to the patients and was measured by five components:
  - Tangibles (TAN) including eight variables that express the exterior elements: Directional signs system in the hospital (TAN1); The name tag wearing of doctors, physicians and nurses (TAN2); The costume of doctors, physicians, nurses and hospital staffs (TAN3); The facilities at the hospital (TAN4); The airiness of the waiting rooms and treatment rooms (TAN5); The cleanness and the convenience of the rest rooms (TAN6); The cleanness of the treatment rooms (TAN7); The adequacy of hospital bed for patients (TAN8).
  - Reliability (REL) including two observed variables to reflect the capability to provide services accurately and punctually, prestige; respect the commitments and keep the promises to the patients: The reasonable costs and fees for examination and treatment (REL1); the accord between the fee of charge and the income of patients' families (REL2).
  - Empathy (EMP) including five observed variables measuring the care to the patients as behaving thoughtfully and conscientiously that brings the convenience and happiness to patients when using the services: The care of doctors and physicians to the situation, physiological and psychological issues of the patients (EMP1); The frequency and quality of doctors and physicians visiting patients (EMP2); The convenience for patients' families to take care patients (EMP3); The easiness for patients to contact with doctors, physicians and nurses when necessary (EMP4); The convenience and comfort that medical staffs make to the patients (EMP5).
  - Assurance (ASS) including eight observed variables measuring the factors affecting the belief and trust of the patients: The duration from being hospitalized to discharged from hospitals (ASS1); The expertness of the nurses (ASS2); The adequate and detailed explanations to all enquiries of patients (ASS3); The level of clarity and easiness of the explanations about disease situation to patients (ASS4); The fair when taking care patients (ASS5); The guideline of medicine usage (ASS6); The patients are required to do necessary medical tests (ASS7); The carefulness of doctors and physicians when diagnose for patients (ASS8).
  - Responsiveness (RES) including three observed variables measuring the capability to solve problems quickly and effectively as well as the willing to assist patients of hospital: The duration of waiting time for examination and treatment (RES1); The punctuality of giving back the test's results (RES2); The adequate publicity of all fees and costs for examination and treatment (RES3).
- Technical quality (TEC) including two observed variables measuring the results of examination and treatment process: The better health of patients when leaving hospitals (TEC1); the adequate guideline for taking rest and the regimen to patients when leaving hospitals (TEC2).
- Image (IMA) including three observed variables measuring the feeling of patients to examining and treating activities, the reputation and the prestige of the hospital: The public hospital improving the quality of life (IMA1); The hospital is the believable place for examination and treatment (IMA2); The medical ethics of doctors, nurses and physicians at the hospital (IMA3).

According to the previous researches and findings of experts, some hypotheses about the relationship between these factors were built up as following:

- *Hypothesis H1: The positive correlation between the functional quality and technical quality;*
- *Hypothesis H2: The positive correlation between the functional quality and image;*
- *Hypothesis H3: The positive correlation between the technical quality and image;*
- *Hypothesis H4: The positive correlation between the functional quality and satisfaction;*
- *Hypothesis H5: The positive correlation between the image and satisfaction;*
- *Hypothesis H6: The positive correlation between the techical quality and satisfaction;*

The proposed model is illustrated in the Fig. 2.

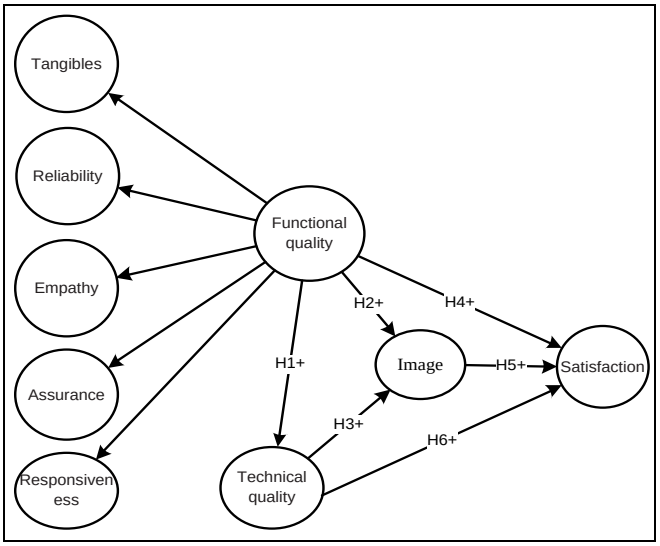

Fig 2: The proposed model

2.2 Data Understanding

The analysis methodology was conducted by the researchers, that identified the population are patients using the hopitals' healthcare services during the years 2012 and 2013 at DakNong, Quang Ngai, Da Nang provinces. The selected sample is 1500 random patients surveyed by questionaires and categorized by gender, insurance type, ethnic, career, the number of times using hospitals' healthcare services and service department.

2.3 Data Preparation

After the survey, all the questionnaires must be evaluated the response rate, the representative characteristic and the distribution of samples. According to the results, there were 1500 responses that make up 100% and well distributed as gender, age group, job and the quantity of examination and treatment. The data were inputted in SPSS 16.0 to be analyzed. Before analyzing, the data was cleaned to remove invalid and abnormal one and add valid data, the errors also was removed by the reliable statistic methods. In this research, there were three invalid questionnaires to be removed. The main object in this phase was the EFA (Exploratory Factor Analysis) to explore the factors affecting the satisfaction to the healthcare services. The Cronbach's Alpha coefficient was used to assess the reliability of scales and remove unsuitable items. In this project, the high coefficients of the factors were presented in Table 1.

Table 1: The Cronbach's Alpha coefficient of the factors

No	Factors	Cronbach's Alpha
1	Tangibles	.789
2	Reliability	.812
3	Empathy	.847
4	Assurance	.878
5	Responsiveness	.851
6	Technical quality	.811
7	Image	.817
8	Satisfaction	.846

Then the data were analyzed by the factor extraction method. The items with factor loading on any factor less than 0.4 were removed (Gerbing & Anderson, 1988) [4]. The factor extraction methods used in this analysis is the Principal Component analysis with Varimax rotation. The scale is accepted when the Total Variance Explained is larger than 0.5 (Gerbing & Anderson, 1988) [4], the Kaiser-Meyer- Olkin KMO index measuring of sample adequacy is larger than 0.5, the Bartlett's test of sphericity must have statistically significance (Sig. < 0.05). The results of this project are presented in Table 2.

Table 2: The result of rotation in factor analysis

		Functional quality	Technical quality	Image	Satisfaction
The Number of Items		25	2	3	6
The number of factors extracted		5	1	1	1
Total Variance Explained		59.556	84.137	73.409	56.559
KMO		0.929	0.501	0,701	0.823
Bartlett's Test of Sphericity	Approx. Chi-Square	17071.545	937.932	1620.019	3472.558872
	Df	300		3	15
	Sig.	0,000	0,00	0.000	0,000

2.4 The Modelling

The explored factors were used for Confirmatory Factor Analysis (CFA). When operating CFA, the SEM (Structural Equation Modeling) was chosen using AMOS 21. This is a statistical technique for testing and estimating causal relationship by using a combination of statistical data and qualitative causal assumptions. In the SEM, the following indices were used to choose the good model:

- The ratio of chi-square divided by its degrees of freedom (Chi –squared /df) is the index to measure the model fit. The value of CMIN/df ≤ 2 is considered as a sign of good model, the value of CMIN/df ≤ 3 is also acceptable in some cases (Carmines & McIver, 1981) [3];
- The Comparative Fit Index (CFI): The model has the CFI ≥ 0.9 was chosen (Bollen, 1989) [1];
- The Turkey & Lewis Index (TLI): The model has the TLI ≥ 0.9 was chosen (Bentler and Bonnet, 1980);
- The Root Mean Square Error (RMSE) to measure the differences between values predicted by a model and the values actually observed. According to Bollen (1989) [1], the model has the RMSE ≤ 0.07 should be chosen but Schumaker and Lomax (2004) [4] suggested the RMSE ≤ 0.05.

The model's parameters were estimated by the Maximum Likelihood Estimation method with AMOS. According to estimating results presented in the Table 3, all of the relationships have significant meaning. However, the indices to choose model was still not acceptable: Chi-quared/df= 5.767; CFI = 0.897<0.9 and TLI = 0.888< 0.9.

Table 3: The estimating parameters of the model

			Estimate	S.E.	C.R.	P
TEC	<---	FUN	.944	.050	18.962	***
IMA	<---	TEC	.137	.033	4.147	***
IMA	<---	FUN	.815	.054	15.025	***
TAN	<---	FUN	.384	.030	12.872	***
ASS	<---	FUN	.859	.045	19.214	***
RES	<---	FUN	1.068	.053	20.097	***
REL	<---	FUN	1.000			
SAT	<---	IMA	.325	.057	5.723	***
EMP	<---	FUN	1.098	.054	20.441	***
SAT	<---	FUN	.653	.071	9.208	***
SAT	<---	TEC	.171	.032	5.373	***

The proposed model did not meet the requirements so it was modified by using the Modification Indices. According to the estimating results after modification presented in Table 4, the modified model has met the standard with Chi–squared/df = 2.712<3; CFI = 0.964>0.9 and TLI =0.960> 0.9, RMSE = 0.034<0.05

Table 4: The estimating parameters of the modified model

			Regression Weights				Standardized Regression Weights
			Estimate	S.E.	C.R.	P	Estimate
TEC	<---	FUN	.949	.050	18.927	***	.733
IMA	<---	TEC	.152	.033	4.577	***	.168
IMA	<---	FUN	.838	.055	15.297	***	.714
TAN	<---	FUN	.307	.027	11.264	***	.620
ASS	<---	FUN	.822	.046	18.042	***	.871
RES	<---	FUN	1.070	.054	19.998	***	.758
REL	<---	FUN	1.000			***	.726
SAT	<---	IMA	.237	.047	5.055	***	.225
EMP	<---	FUN	1.163	.056	20.913	***	.887
SAT	<---	FUN	.725	.068	10.713	***	.585
SAT	<---	TEC	.179	.032	5.634	***	.187

According to the results presented in Table 4, the coefficients with standardized regression weights greater than 0.5 (Gerbring & Anderson, 1988) determine the convergent validity of the factors. Also, these factors have discriminant validity.

Table 5: The results of Bootstrap method with N=3000

Parameter			SE	SE-SE	Mean	Bias	SE-Bias
TEC	<---	FUN	.041	.001	.888	.001	.001
IMA	<---	TEC	.040	.001	.153	.000	.001
IMA	<---	FUN	.047	.001	.784	.001	.001
RES	<---	FUN	.041	.001	1.002	.001	.001
ASS	<---	FUN	.039	.001	.879	.000	.001
TAN	<---	FUN	.045	.001	.764	.000	.001
REL	<---	FUN	.045	.001	1.047	.001	.001
EMP	<---	FUN	.000	.000	1.000	.000	.000
SAT	<---	FUN	.071	.001	.680	.003	.001
SAT	<---	IMA	.057	.001	.237	-.001	.001
SAT	<---	TEC	.039	.001	.179	-.001	.001

The Bootstrap method was also used to assess the stability of the model. Bootstrap is a resampling method and the 'population' is in fact the sample. In this research, the repetition of taking a bootstrap sample was N=3000. According to the Table 5, the estimating results could be acceptable.

2.5 The Evaluation

After assessing the model, the hypotheses were tested. According to Table 6, all the hypotheses were accepted with significant level of 5%.

Table 6: The results of testing hypotheses about the relationship of factors

Hypothesis				Regression Weights				Testing result
				Estimate	S.E.	C.R.	P	
H1	TEC	<---	FUN	.949	.050	18.927	***	Acceptable
H2	IMA	<---	FUN	.838	.055	15.297	***	Acceptable
H3	IMA	<---	TEC	.152	.033	4.577	***	Acceptable
H4	SAT	<---	FUN	.725	.068	10.713	***	Acceptable
H5	SAT	<---	IMA	.237	.047	5.055	***	Acceptable
H6	SAT	<---	TEC	.179	.032	5.634	***	Acceptable

Intuitively, the final model is presented in Fig. 3.

The satisfaction of patients depends on gender, regions, the quantity of examination and treatment. The results of variance analysis are presented in Table 7.

Table 7: The results of variance analysis

Dummy variable	F-statistic	Sig.
Gender	20.707	.000
Region	144.135	.000
The quantity of examination and treatment	5.415	.001

The multiple comparison analysis also was used and the results were presented in Table 8

Table 8: Multiple Comparisons

(I) Local	(J) Local	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Da Nang	Dak Nong	.21563*	.03744	.000	.1422	.2891
	Quang Ngai	.60367*	.03632	.000	.5324	.6749
Dak Nong	Da Nang	-.21563*	.03744	.000	-.2891	-.1422
	Quang Ngai	.38804*	.03601	.000	.3174	.4587
Quang Ngai	Da Nang	-.60367*	.03632	.000	-.6749	-.5324
	Dak Nong	-.38804*	.03601	.000	-.4587	-.3174

*. The mean difference is significant at the 0.05 level.

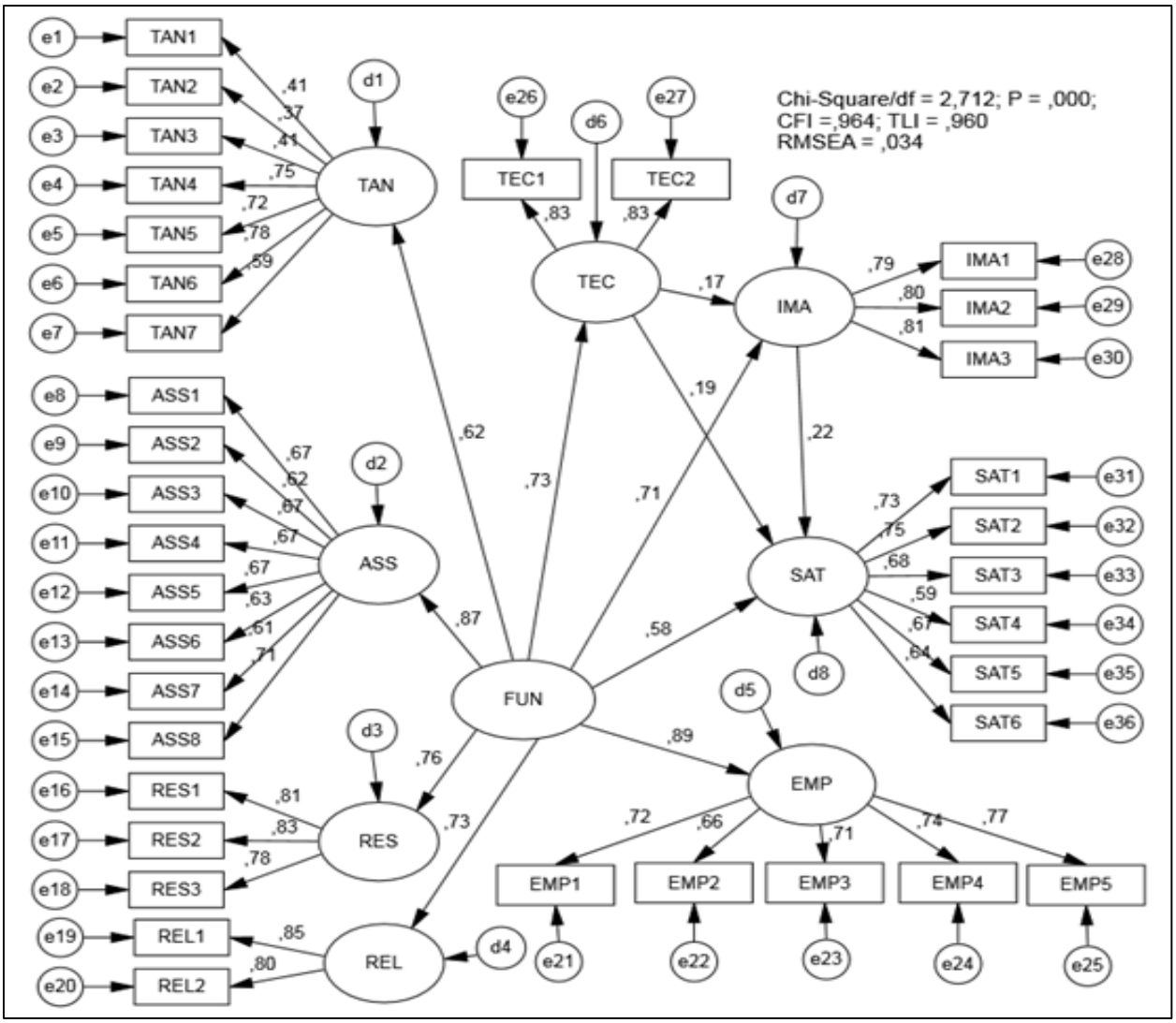

Fig 3: The standardized model

2.6 The Deployment

The model, after being assessed that meet the standards, will be used to analyze and raise ideas for the policy planners. According to the estimating results, there are some following conclusions.

- The satisfaction of patients is significantly affected by the functional quality factor (0.585) and slightly affected by technical quality factor (0.187). The functional quality impacts not only the satisfaction but also the technical quality factor (0.733) and image factor (0,714).
- To improve the functional quality, the hospital need to take more attention on the serving process including the items of tangible, reliability, empathy, assurance and responsiveness. The elements of Empathy item should be mostly considered (0.887), it means that the hospitals need to take care about the physiological and psychological issues of patients as well as understand the patients' family situation and take care frequently the situation of patients. The elements of Assurance item including the fair in treatment, the expertness, the communication skills and behavior of serving also need to be considered (0,871). The healthcare centers and facilities need to be cleaned frequently and safe to build up the patient satisfaction as well as increase the usage time of facilities and hospital

infrastructure. The hospital need to improve the responsiveness items (0.758) to provide services as well as to assist patients quickly and effectively.

- The satisfaction level depends not only on functional quality but also on image factor that increases the reputation and the brand of the hospitals.
- According to the results, the level of patient satisfaction to the healthcare services' quality was different depending on gender, the quantity of examination and treatment and regions. The satisfaction level at Da Nang was highest whereas this one is lowest at Quang Ngai.

3. Conclusion

The satisfaction to healthcare services' quality depends on many factors and there is a large difference between regions, the quantity of examination and treatment, ethnic and so on. The leaders of the hospitals have searched for solutions to improve the healthcare services' quality. The fact that there are more and more data need to be analyzed for decision making. Moreover, the data is increasing and more various, and in some cases data is abnormal and inadequate to give information. Therefore, the analysis methods need to be more modern, accurate and the analysis process must be more completed. Besides, there have been database system softwares that satisfy the demands of massive data storage. Especially, there are many data mining softwares as SPSS, SPSS MODELER, AMOS, STATISTICA, WEKA…there in the MODELER and STATISTICA already incorporates the CRISP – DM process.

4. References

1. Bollen KA. Structural Equations with Latent Variables, Wiley, New York, 1989.
2. Brown SW, Swartz TA. A Gap Analysis of Professional Service Quality. Journal of Marketing, 1989; 53(2): 92-98.
3. Carmines EG, McIver JP. Analyzing Models with Unobserved Variables. In G.W. Bohrnstedt & E.F. Borgatta (Eds.), Social Measurement: Current Issues, Beverly Hills, Sage, CA, 1981.
4. Gerbing DW, Anderson JC. An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment. Journal of Marketing Research, 1988; 25: 186-192.
5. Gi-Du Kang, Jeffrey James. Service quality dimensions: an examination of Gronroos's service quality model, Emerald Group Publishing Limited, managing service quality, 2004; 14(14):266-277.
6. The Vietnam Ministry of Health. Decision No. 29/2008/QĐ-BYT dated on August 18, 2008. On The rules of behaviours of hospitals' staffs in healthcare centers.
7. The Vietnam Ministry of Health. Decision No. 527/EMPr-BYT dated on June 18, 2009. On The programs to improve the quality of examination and treatment of healthcare centers to satisfy the demands of patients with individual health insurance.
8. The Vietnam Ministry of Health. Decision No. 4858/QD-BYT dated on Dec 03, 2013 (a). On The pilot promulgation of Standard Set for assessing the hospital's quality.
9. The Vietnam Ministry of Health. Circular No. 19/2013/TT-BYT dated on July 12, 2013(b). On The guidelines for implementing quality assurance of examination and treatment services at the hospitals.
10. The Vietnam Ministry of Health. Decision No. 1313/QD-BYT dated on April 22, 2013(c). On The promulgation of guidelines for examination processing at the outpatient departments of the hospitals.
11. Tran Xuan Hien, Le Dan. Assessing the satisfaction level of patients to healthcare services at the public health center in Dak Nong. The Provincial-level Research Project, 2012.
12. Randall Schumacker E, Richard Lomax G. A Beginner's Guide to Structural Equation Modeling, Lawrence Erlbaum associates Publisher London, 2004.
13. Rex Kline B. Principles and practice of structural equation modeling, The Guilford Press, New York, 2005.
14. Robert Ho. Handbook of Univariate and multivariate data analysis and interpretation with SPSS, 2006.
15. Chapman P *et al*. CRISP-DM 1.0 - Step-by-step data mining guide. Accessed from http://www.staff.it.uts.edu.au/~paulk/teaching/dmkdd/ass2/readings/methodology/CRISPWP-0800.pdf. 18 Dec 2016