Acıbadem Univ. Sağlık Bilim. Derg. 17 (April, May, June) 2026
https://doi.org/10.31067/acusaglik.1866468

ORIGINAL ARTICLE / ARAŞTIRMA YAZISI Health Care Administration/Sağlık Hizmetleri Yönetimi

Modeling the Determinants of Satisfaction with Health Care Services in Türkiye Using Machine Learning

Cuma Çakmak[1]

[1] Dicle University, Faculty of Economics and Administrative Sciences, Department of Health Management, Diyarbakır, Türkiye

Cuma ÇAKMAK
0000-0002-4409-9669

Correspondence: Cuma Çakmak
Dicle University, Faculty of Economics and Administrative Sciences, Department of Health Management, Diyarbakır, Türkiye
Phone: +90 412 241 10 01
E-mail: cuma.cakmak@dicle.edu.tr

Received: 18.01.2026
Accepted: 28.03.2026

ABSTRACT

Purpose: Patient satisfaction is a fundamental indicator for evaluating the quality of healthcare services and is of great importance for the sustainability of healthcare systems. In this study, machine learning methods were used to determine the factors affecting individuals' satisfaction levels with healthcare services and to predict satisfaction levels.

Methods: In this study, machine learning methods were used to identify factors affecting individuals' levels of satisfaction with healthcare services and to predict their satisfaction levels. The Turkish Statistical Institute (TUIK) life satisfaction dataset was used; the data were split into training (80%) and test (20%) sets, and the SMOTE method was applied to address class imbalance.

Results: Logistic Regression, SVM, Decision Trees, Random Forest, and Naive Bayes models were tested and Logistic Regression was determined to be the most suitable model. According to the findings, the variables of age, gender, health satisfaction scale, frequency of hospital visits, examination problems, hygiene problems, examination satisfaction, insufficient healthcare personnel, medication cost problems, and waiting time problems were found to be statistically significant.

Conclusion: Ultimately, machine learning approaches are practical for analyzing healthcare satisfaction.

Keywords: Patient Satisfaction, Machine Learning, Healthcare Service Quality, Classification, Data Analytics

ÖZET

Amaç: Hasta memnuniyeti, sağlık hizmetlerinin kalitesini değerlendirmede temel bir göstergedir ve sağlık sistemlerinin sürdürülebilirliği açısından büyük önem taşır. Bu çalışmada, bireylerin sağlık hizmetlerinden memnuniyet düzeyini etkileyen faktörleri belirlemek ve memnuniyet düzeyini tahmin etmek amacıyla makine öğrenmesi yöntemleri kullanılmıştır.

Yöntem: Türkiye İstatistik Kurumu'nun (TÜİK) yaşam memnuniyeti veri seti kullanılmış; veriler eğitim (%80) ve test (%20) olarak ikiye ayrılmış, sınıf dengesizliği durumunda SMOTE yöntemi uygulanmıştır.

Bulgular: Lojistik Regresyon, SVM, Karar Ağaçları, Rastgele Orman ve Naive Bayes modelleri denenmiş, en uygun modelin Lojistik Regresyon olduğu belirlenmiştir. Bulgulara göre, yaş, cinsiyet, sağlık memnuniyet ölçeği, hastaneye başvuru sıklığı, muayene sorunları, hijyen sorunları, muayene memnuniyeti, sağlık personeli yetersizliği, ilaç maliyeti sorunları ve bekleme süresi sorunları değişkenleri istatistiksel olarak anlamlı bulunmuştur.

Sonuç: Sonuç olarak, makine öğrenmesi yaklaşımlarının sağlık memnuniyeti analizinde etkili olduğu görülmektedir.

Anahtar Kelimeler: Hasta Memnuniyeti, Makine Öğrenmesi, Sağlık Hizmeti Kalitesi, Sınıflandırma, Veri Analitiği

The quality and effectiveness of healthcare services are becoming increasingly important, and patient satisfaction is gaining prominence in this context (1). Patient satisfaction is considered one of the most important indicators evaluating the quality of healthcare services, and satisfaction with healthcare services not only reflects the individual experience but also the effectiveness of healthcare institutions, the efficiency of service delivery, and the success of health policies (2). The level of satisfaction is affected by numerous factors such as patients' interactions with service providers, waiting times, information acquisition processes, and the communication skills of healthcare personnel. Accurate identification and analysis of these factors are fundamental to improving the quality of healthcare services (3). Traditional statistical methods have been used for many years to analyze factors affecting patient satisfaction. However, these methods may be limited in their ability to identify complex relationships, especially in large datasets (4). At this point, machine learning (ML) techniques stand out as an important tool in the healthcare sector. ML is a subfield of artificial intelligence technologies. By leveraging a data-driven, algorithmic approach, ML can identify relationships among multiple variables, detect patterns, and make predictions (5,6).

The proliferation of advanced algorithms and large datasets has enabled the application of complex data analysis across sectors, with a focus on solving problems in healthcare. Machine learning applications in healthcare services mainly focus on predicting diseases and health problems (7). Machine learning (ML) is applied through various algorithms in modeling factors affecting satisfaction with healthcare services. Classification algorithms categorize satisfaction levels into high, medium, or low, while regression algorithms predict satisfaction scores as continuous values. The variety of methods, such as Artificial Neural Networks (ANN), Bayesian Networks (BN), Support Vector Machines (SVM), and Decision Trees (DT), enables the development of effective, accurate predictive models. In particular, Random Forest, SVM, Gradient Boosting, and Neural Network algorithms show strong predictive performance in variables such as patient characteristics, service duration, appointment system, quality of communication with healthcare personnel, and institutional factors. The applicability of machine learning methods in different fields and their ability to extract basic features from complex datasets show that these techniques can also be used in healthcare satisfaction analyses (8). The application of machine learning, particularly in healthcare, is not limited to measuring current satisfaction levels; it can also guide improvements in service quality. Big data and machine learning methods have the potential to improve service delivery by supporting decision-making processes in healthcare (9). In addition, the factors with the most significant impact on patient satisfaction can be identified using machine learning models, allowing institutions to focus their resources and training programs on them (10). Machine learning techniques analyze data from patient-doctor interactions to identify psychological and emotional factors that affect patient experiences, thereby contributing to the development of patient-centered service strategies (11).

In recent years, measuring satisfaction with healthcare services has become an important research area for health management and policy development. Traditional studies conducted in Türkiye have frequently used surveys and regression analyses to identify factors affecting individuals' satisfaction with healthcare services. For example, in a study conducted by Uğurluoğlu et al. (12), appointment scheduling processes, examination and analysis costs, insufficient number of physicians and healthcare personnel, drug prices, waiting times, cleanliness/hygiene, and the attitudes of healthcare personnel were identified as the main factors affecting the level of satisfaction with healthcare services. In the study, hierarchical regression was used to analyze satisfaction levels among 9,397 individuals. In contrast, today, machine learning techniques enable more comprehensive and accurate modeling of the factors affecting satisfaction with healthcare services. These methods not only measure the current level of satisfaction but also enable the prioritization of factors affecting patient experience and support healthcare managers' strategic decision-making (11,13). Algorithms such as Random Forest, Decision Tree, and Gaussian Naive Bayes, in particular, enable accurate prediction of satisfaction levels by considering variables such as patient characteristics, service duration, appointment system, and quality of communication with healthcare personnel. This approach offers a more flexible, data-driven modeling opportunity than traditional regression analyses.

This study aims to model the factors affecting satisfaction with healthcare services in Türkiye using machine learning methods. The findings are expected to guide healthcare institutions and policymakers and contribute to the development of strategies to increase patient satisfaction. Furthermore, the study aims to demonstrate the applicability of machine learning techniques in healthcare and their importance in healthcare quality management.

Material and Method

In this study, satisfaction with general health services was included as the dependent variable, while the other variables were included as independent variables. Table 1 presents the variables, descriptions of the variables, and variable abbreviations.

Table 1. Description and Abbreviation of Variables

Variables	Variable Description	Variable Abbreviation
Overall, General Health Satisfaction Scale	Overall satisfaction with healthcare services	**OHS**
Age	Age of the individual (years)	**AGE**
Gender	Gender of the individual	**GEN**
Marital Status	Marital status of the individual	**MS**
Education Level	Educational level of the individual	**EDU**
Income Group	Household income group	**INC**
Employment Status	Employment status of the individual	**EMP**
Health Satisfaction Scale	Satisfaction with the individual's general health status	**HSS**
Hospital Visit Frequency	Number of hospital visits in the last 12 months	**HVF**
Examination Problem	Problems caused by insufficient consultation time with the physician	**EXAM_P**
Hygiene Problem	Problems caused by inadequate hygiene of the healthcare environment	**HYG_P**
Examination Satisfaction	Satisfaction with medical examination services	**EXAM_S**
Doctor-related Problem	Problems caused by the physician's lack of attention	**DOC_P**
Nurse-related Problem	Problems caused by nurses	**NUR_P**
Insufficient Health Staff	Problems caused by an insufficient number of healthcare personnel	**STAFF_P**
Examination Fee Problem	Problems caused by high examination fees	**FEE_P**
Drug Cost Problem	Problems caused by high medication prices	**DRUG_P**
Waiting Time Problem	Problems caused by long waiting times	**WAIT_P**
Health Insurance Cost	The cost of health insurance	**INS_COST**
Health Problem	Presence of health problems or chronic conditions	**HLTH_P**
Copayment Problem	Problems caused by co-payment charges	**COPAY_P**

Sample

The dataset used is the 2024 Life Satisfaction Micro Dataset of the Turkish Statistical Institute, with the institution's permission. The dataset was modified, and the dependent variable was converted into a binary classification. As a result of this conversion, the dataset contains only two classes: "not satisfied (1)" and "satisfied (0)." The new dataset consists of 2811 observations and 21 variables. When the class distribution is examined, the dissatisfied group (minority class) comprises 827 people (29.4%), and the satisfied group (majority class) comprises 1984 people (70.6%). The dataset was then divided into 80% for training and 20% for testing. The training set consists of 2248 observations (661 dissatisfied, 1587 satisfied), and the test set consists of 563 observations (166 dissatisfied, 397 satisfied).

Data Preprocessing

Before proceeding to the analysis phase, a series of preparatory and preliminary testing procedures were performed on the dataset. First, descriptive statistics for the variables were calculated, and the risk of multicollinearity among independent variables was tested using VIF (Variance Inflation Factor) values. A consistent coding scheme was applied for categorical variables; for "problem"-focused variables, the assignment was "1=Yes, 0=No," and for 'satisfaction'-focused variables, the assignment was "1=Yes, 0=No." The dependent variable, OHS, was converted into a binary format where "1=Not Satisfied" (the focus of the risk analysis) and "0=Satisfied." To address the class imbalance, the dataset was first split into 80% training and 20% testing sets to prevent data leakage. The SMOTE (Synthetic Minority Oversampling Technique) method was applied exclusively to the training set to balance the minority class ("Not Satisfied"). Consequently, the number of observations for the "Not Satisfied" group in the training set was increased from 661 to 1587, achieving a 1:1 balance. The test set (n=563) was preserved in its original distribution to ensure an unbiased evaluation of the model's real-world predictive performance.

Model Selection and Evaluation

During the model setup phase, the performance of the Logistic Regression, Support Vector Machines (SVM), Random Forest, XGBoost, LightGBM, Naive Bayes, and Multilayer Perceptron (MLP) algorithms was compared. Grid search and cross-validation techniques were used for hyperparameter optimization; specifically, SVM and Logistic Regression models were tested under different parameter values (kernel type, C coefficient, solver options). The performance of the models was evaluated based on Accuracy, F1 Score, and ROC-AUC criteria.

To clarify the complex decision structure (black-box) of the SVM model with the highest predictive power, Logistic Regression was used as a complementary tool. This allowed us to leverage the high predictive capacity of machine learning algorithms while also reporting the statistical significance (p-values) and coefficient directions of the variables through Logistic Regression. The variable importance levels of the most successful models were analyzed using both Permutation Importance and SHAP (Shapley Additive exPlanations) methods; thus, the relative contributions and effect directions of the independent variables to the model outputs were revealed from a holistic perspective.

To clarify the complex decision structure (black-box) of the SVM model, which demonstrated the highest predictive power, Logistic Regression was employed as an interpretable complementary tool. In this study, the "Not Satisfied" (1) class was designated as the target variable to specifically investigate the risk factors negatively impacting health service quality. This approach enabled the alignment of high-capacity machine learning outputs with the statistical rigor of Logistic Regression, providing coefficient directions and p-values for each predictor. Furthermore, variable importance was cross-validated using SHAP (Shapley Additive exPlanations). By ensuring that the direction of Logistic Regression coefficients remains consistent with SHAP values, the model's internal validity was strengthened, and the specific impact of each health-related problem on dissatisfaction was revealed from a holistic perspective.

Hyperparameter Optimization

To enhance the generalization ability of the models and prevent overfitting, a comprehensive hyperparameter optimization was performed. During this process, a wide parameter space was explored using Grid Search and Randomized Search methods. For example, in the Logistic Regression model, the regularization parameter (C) and penalty type (L1, L2) were optimized; in the SVM model, the kernel type (RBF, Linear), C coefficient, and gamma parameters were focused on.

To measure the stability and reliability of model performance, a 5-fold cross-validation method was applied throughout the entire process. This optimization process ensured that the obtained variable importance levels and coefficients were derived from optimized models rather than randomly, by enabling the comparison of the best capabilities of different algorithms (especially SVM and Logistic Regression).

Data Analysis

In this research, data analyses were conducted on the Google Colab platform using Python. Google Colab, a cloud-based analysis environment, enables online data processing, statistical modeling, and visualization. The analyses primarily utilize the pandas, numpy, scipy, statsmodels, and matplotlib libraries. This method was used to clean the data, calculate descriptive statistics, and apply regression analyses.

Ethical Considerations

This research utilizes data from the 2024 Life Satisfaction Survey provided by the Turkish Statistical Institute (TÜİK). Permission to use the dataset was obtained from TÜİK. This study does not require ethical committee approval, as it used secondary data. Permission to use the data was obtained from the Turkish Statistical Institute with application number 588938. Only anonymized data were used in the study; participants' identifying information or personal data were not included in the analysis process. In this respect, the study was conducted in accordance with the principles of research and publication ethics.

Methodological Limitations

This study identifies factors determining health satisfaction through machine learning algorithms, yet it has certain methodological limitations. Firstly, the TÜİK 2024 Life Satisfaction Micro Data Set used in the analyses is a secondary data source and is limited to the variables included in the survey; the absence of individuals' clinical health history or more detailed socio-psychological indicators in the data set narrows the scope of the analysis. Secondly, while the SMOTE technique, used to address class imbalance in the dataset, enhances model performance by augmenting the minority class with synthetic data, it should be noted that these synthetic observations may differ structurally from real-world data and could cause a low level of bias in the results. Finally, the cross-sectional nature of the data used prevents the observation of changes in the relationships between variables over time.

Results

Table 2. Descriptive Statistics and VIF Value for Variables				
Variable	**Value**	**n**	**%**	**VIF**
Overall, General Health Satisfaction Scale (OHS)	Satisfied (0)	1984	0.70	NA
	Not Satisfied (1)	827	0.29	
Age (AGE)	(Mean: 45.33; Std: 15.24; Min: 18; Max: 93)			9.024
Gender (GEN)	Male (1)	1493	0.53	2.414
	Female (0)	1318	0.46	
Marital Status (MS)	Married (1)	2041	0.72	3.864
	Single (0)	770	0.27	
Education Level (EDU)	Collage and Lower (1)	1759	0.62	7.162
	Higher Education (2)	908	0.32	
	Postgraduate (3)	144	0.05	
Income Group (INC)	25.501+ TL	803	0.28	5.611
	8.501–12.500 TL	561	0.20	
	17.501-25.500 TL	509	0.18	
	0–8.500 TL	473	0.16	
	12.501- 17.500 TL	465	0.16	
Employment Status (EMP)	Employed (1)	1452	0.51	2.857
	Not Employed (0)	1359	0.48	
Health Satisfaction Scale (HSS)	Satisfied	1984	0.70	6.674
	Mid	560	0.19	
	Not Satisfied	267	0.09	
Hospital Visit Frequency (HVF)	Public Hospital	1216	0.43	7.853
	Family Health Center	1152	0.41	
	Private Hospital	443	0.15	
Examination Problem (EXAM_P)	No (0)	1422	0.50	2.865
	Yes (1)	1389	0.494	
Hygiene Problem (HYG_P)	No (0)	1997	0.71	1.956
	Yes (1)	814	0.29	
Examination Satisfaction (EXAM_S)	Yes (1)	1713	0.60	2.778
	No (0)	1098	0.39	
Doctor-related Problem (DOC_P)	No (0)	2132	0.75	2.634
	Yes (1)	679	0.24	
Nurse-related Problem (NUR_P)	No (0)	2247	0.79	2.482
	Yes (1)	564	0.20	
Insufficient Health Staff (STAFF_P)	No (0)	1824	0.64	1.710
	Yes (1)	987	0.35	
Examination Fee Problem (FEE_P)	Yes (1)	1508	0.53	3.617
	No (0)	1303	0.46	
Drug Cost Problem (DRUG_P)	Yes (1)	1713	0.60	4.410
	No (0)	1098	0.39	
Waiting Time Problem (WAIT_P)	No (0)	1482	0.52	3.263
	Yes (1)	1329	0.473	
Health Insurance Cost (INS_COST)	Public Insurance (SGK)	2665	0.94	9.883
	Private Insurance	95	0.03	
	Out-of-pocket insurance	51	0.01	
Health Problem (HLTH_P)	No (0)	2437	0.86	1.337
	Yes (1)	374	0.13	
Copayment Problem (COPAY_P)	No (0)	1525	0.54	2.418
	Yes (1)	1286	0.45	

Note: This data was obtained from the Turkish Statistical Institute's 2024 Life Satisfaction Micro Data Set.

Table 2 presents descriptive statistics for variables obtained from the Turkish Statistical Institute's 2024 Life Satisfaction Micro Data Set and VIF (Variance Inflation Factor) values measuring the risk of multicollinearity. When examining participants' OHS, it is seen that a large majority, 70%, are satisfied. The demographic structure of the sample consists of individuals with an average age of 45.33; the gender distribution is balanced between men (53%) and women (46%). It is noteworthy that the college and below group accounts for 62% of the education level, while the employment rate stands at 51%. Looking at healthcare service usage habits, it is understood that the most frequently visited institutions are Public Hospitals (43%) and Family Health Centers (41%).

Based on the problems encountered during the service process, approximately half of the participants reported issues related to EXAM_P (49.4%) and WAIT_P (47.3%). Financially, DRUG_P (60%) and FEE_P (53%) stand out as the most prominent complaints. On the other hand, hygiene problems (29%) and doctor/nurse-related issues (20-24%) occur at lower rates. When examining VIF values, which are critical for statistical reliability, it is seen that all variables remain below the accepted threshold of 10, but variables such as INS_COST (9.883), AGE (9.024), and HVF (7.853) have high values. A sensitivity analysis was performed to assess the impact of this situation on model stability. In the tests performed, it was observed that removing the variables in question from the model did not significantly improve prediction performance (ROC-AUC and F1-score) and did not significantly improve the model's AIC value (3477.02 vs. 3478.54, respectively). In light of these findings, it was decided to include all variables in the analysis to prevent data loss and preserve the interaction structure between variables.These findings prove that there is no multicollinearity problem in the model but that some variables have a high potential for correlation with each other, supporting the validity of the analyses (14).

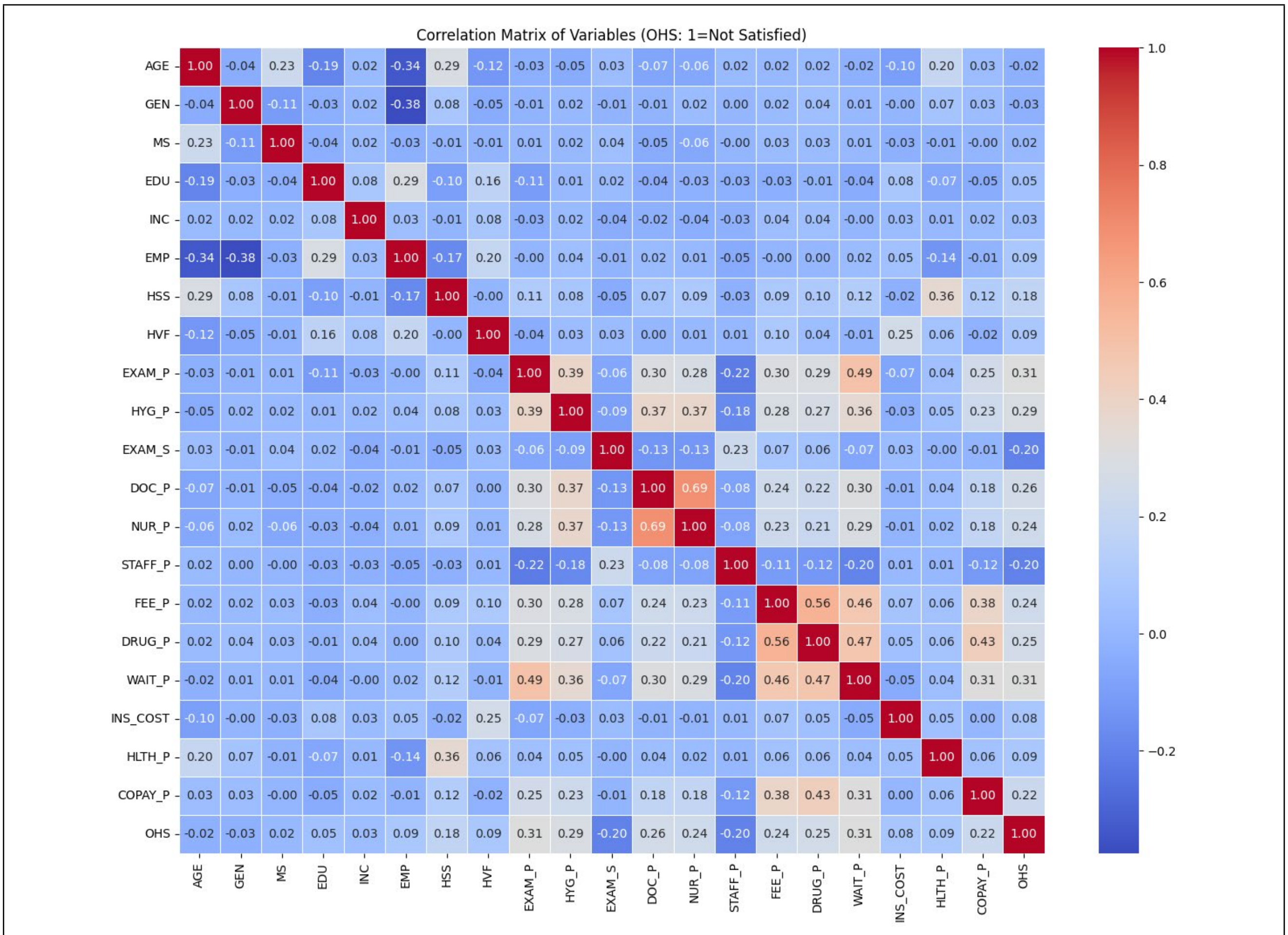

Figure 1: Correlation Matrix of Variables

The correlation analysis conducted as part of the research reveals that disruptions in healthcare services and problems perceived by patients are highly interrelated. The most striking finding in terms of institutional perception is the strong positive relationship between Doctor-Related Problems (DOC_P) and Nurse-Related Problems (NUR_P) (r=0.69). This high correlation indicates that patients do not differentiate between disruptions in the service provided by medical personnel; rather, inadequacies in one unit directly and negatively affect perceptions of another unit. Similarly, the significant relationships between Examination Process Problems (EXAM_P) and Waiting Time Problems (WAIT_P) (r = 0.49) and Hygiene Problems (HYG_P) (r = 0.39) indicate that a bottleneck in operational processes creates a chain reaction of negative effects on the overall quality perception of the facility.

When examining the key "problem" areas that determine Overall Health Satisfaction (OHS), waiting times and the flow of examinations were identified as critically important. According to the findings, Waiting Time Issues (WAIT_P) and Examination Process Issues (EXAM_P) are the variables with the highest correlation (r=0.31) with overall dissatisfaction (OHS). This proves that the greatest disappointment in the patients' overall experience stems from temporal and procedural difficulties in accessing medical intervention rather than the intervention itself. Furthermore, while Hygiene Issues (HYG_P, 0.29) emerge as a significant factor contributing to overall dissatisfaction, the negative relationship between Staff-Related Issues (STAFF_P) and OHS (-0.20) confirms that reducing issues with staff plays a buffering role in alleviating overall system dissatisfaction.

When examining the cost dimension of the service, it is understood that the economic burden parallels other operational issues. Fee Issues (FEE_P) are highly correlated with Drug Cost/Access Issues (DRUG_P, 0.56) and Waiting Time Issues (0.46). This situation shows that when patients' basic expectations, such as "fast service" and "easy access to medication," are not met in exchange for the fees they pay, they perceive the amount paid as a greater problem area (disruption of the price-benefit balance). Finally, the negative correlation between Age (AGE) and Employment (EMP) in the demographic data (-0.34) reminds us that changes in employment-related social security or income structure as age increases may differentiate the perspective on healthcare issues on a socio-economic level.

Table 3. Model Performance Table

Model	Accuracy	F1 Score	ROC-AUC
Logistic Regression	0.753108	0.515679	0.792950
Random Forest	0.756661	0.498168	0.779794
SVM	0.761989	0.521429	0.785932
Naive Bayes	0.719361	0.561111	0.778186
MLP (Neural Net)	0.730018	0.509677	0.692786
XGBoost	0.730018	0.500000	0.756108
LightGBM	0.747780	0.520270	0.776152

When examining the model performance data presented in Table 3, significant differences are observed among the success criteria of the seven different machine learning algorithms. In terms of Overall Accuracy, SVM (0.7619) exhibits the highest performance, followed by Random Forest (0.7566) and Logistic Regression (0.7531). When considering the F1 Score, which represents class imbalance and model sensitivity, the Naive Bayes model (0.5611), which lags behind in accuracy, is seen to provide the best balance among all algorithms. Looking at the ROC-AUC values, which measure the models' ability to distinguish between classes, Logistic Regression (0.7929) stands out as the strongest discriminating model, achieving the highest score. In contrast, MLP (Neural Networks) lags behind other models in terms of both ROC-AUC (0.6927) and accuracy rates, performing below expectations on the dataset. In general, it is understood that the stability of the SVM model and the discriminatory capacity of Logistic Regression are critical for the prediction success of the study. In light of these findings, a strategic role distribution was made among the models in the study. SVM was determined to be the primary prediction engine of the study because it offered the highest overall accuracy rate. On the other hand, Logistic Regression has been positioned as the primary interpretative model, explaining the effects of independent variables on the dependent variable, thanks to its high discrimination capacity (ROC-AUC) and the coefficients it provides (Table 4). Thus, the study aimed for high prediction success while also preventing the model from becoming a 'black box' and revealing the relationships between variables with statistical transparency.

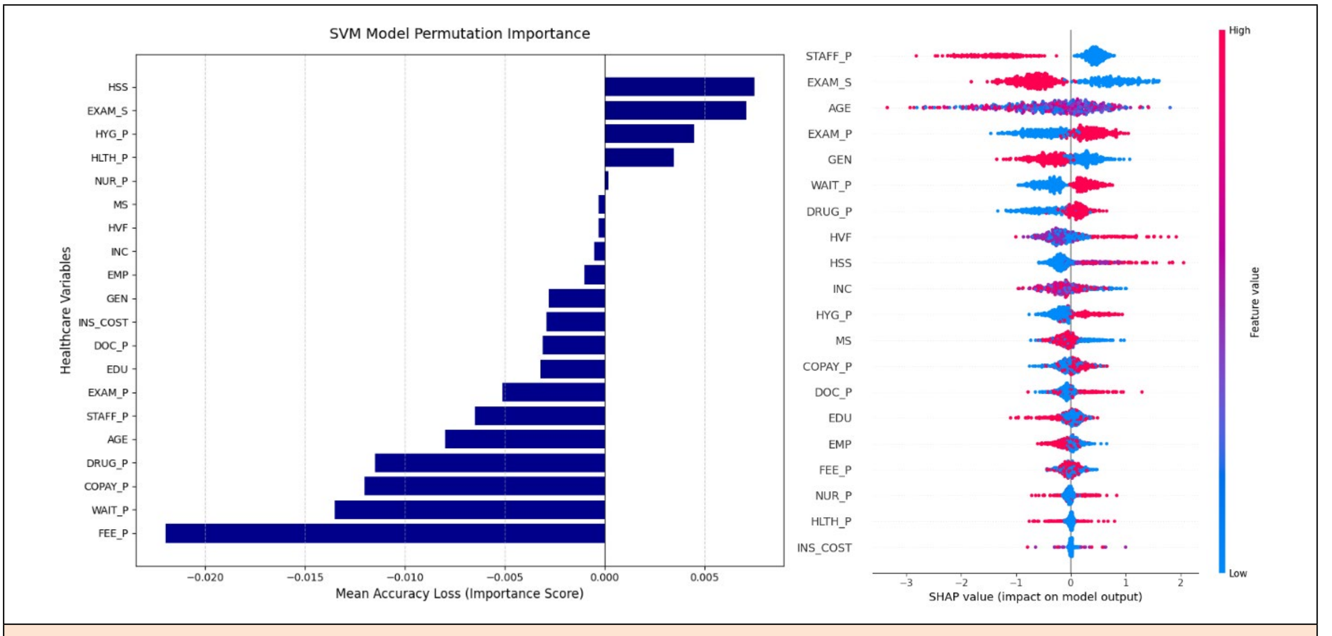

Figure 2. Permutation and Shaping for SVM.

When examining the permutation importance level graph of the SVM model presented in Figure 2, it can be seen that the fundamental parameters determining the model's prediction performance are ranked in a hierarchical structure. Based on the evaluation using mean accuracy loss, it was determined that the EXAM_S and EXAM_P variables have the most dominant weight on the model's discriminatory power. This proves that the model uses operational and subjective patient assessments related to the examination process as the primary reference point when generating output. On the other hand, the fact that variables such as WAIT_P (Waiting Time Issues) and FEE_P (Fee Issues) are located in the negative accuracy loss region indicates that if these factors are removed from the model, the margin of error will increase significantly, thus pointing to these "problem"-focused variables being indispensable negative predictors for the model's classification success.

The SHAP analysis results, which explain the prediction direction and individual effects of the variables, provide more micro-level insights into the model's decision-making mechanism while visually validating the logistic regression findings (Table 4). The fact that the STAFF_P (Personnel Issues) variable ranks highest in the SHAP summary graph and that high values (red dots) are concentrated on the positive SHAP axis confirms that personnel shortages directly and strongly trigger the model's final "dissatisfaction" (OHS=1) output. This finding is fully consistent with the high significance level ($p < 0.001$) and coefficient direction in Table 4.

Similarly, the clustering of high satisfaction values (blue dots) in the negative SHAP region in the EXAM_S variable proves that clinical interaction quality is the strongest protective factor minimizing the risk of dissatisfaction. In the DRUG_P and WAIT_P variables, which represent financial and time barriers, the positioning of red dots on the positive axis shows that an increase in these problems directly leads to patient dissatisfaction. In particular, the narrower and more negatively inclined range exhibited by demographic factors such as education (EDU) and employment (EMP) in the SHAP distribution supports the inference that high socioeconomic status increases the likelihood of satisfaction through health literacy. Consequently, this integrated alignment between the importance hierarchy in permutation analysis and the effect orientations in SHAP analysis reinforces the model's inferential reliability.

Table 4. Logistic Regression Results

Variable	Coef. (B)	Std. Err.	z	P>|z|	[95% Conf. Interval]
Constant	-0.0618	0.300	-0.206	0.837	[-0.651, 0.527]
Age (AGE)	-0.0064	0.003	-1.974	**0.048**	[-0.013, 0.000]
Gender (GEN)	-0.6491	0.092	-7.030	**0.000**	[-0.830, -0.468]
Marital Status (MS)	-0.0934	0.095	-0.981	0.327	[-0.280, 0.093]
Education Level (EDU)	-0.0605	0.081	-0.751	0.452	[-0.218, 0.097]
Income Group (INC)	-0.0420	0.029	-1.454	0.146	[-0.099, 0.015]
Employment Status (EMP)	-0.1892	0.101	-1.882	0.060	[-0.386, 0.008]
Health Satisfaction Scale (HSS)	0.3339	0.074	4.505	**0.000**	[0.189, 0.479]
Hospital Visit Frequency (HVF)	0.1277	0.063	2.012	**0.044**	[0.003, 0.252]
Examination Problem (EXAM_P)	0.5033	0.098	5.160	**0.000**	[0.312, 0.694]
Hygiene Problem (HYG_P)	0.3121	0.103	3.025	**0.002**	[0.110, 0.514]
Examination Satisfaction (EXAM_S)	-1.1409	0.087	-13.154	**0.000**	[-1.311, -0.971]
Doctor-related Problem (DOC_P)	0.1759	0.135	1.305	0.192	[-0.088, 0.440]
Nurse-related Problem (NUR_P)	-0.0027	0.143	-0.019	0.985	[-0.282, 0.277]
Insufficient Health Staff (STAFF_P)	-1.1761	0.103	-11.376	**0.000**	[-1.379, -0.973]
Examination Fee Problem (FEE_P)	0.1657	0.103	1.602	0.109	[-0.037, 0.368]
Drug Cost Problem (DRUG_P)	0.5757	0.111	5.168	**0.000**	[0.357, 0.794]
Waiting Time Problem (WAIT_P)	0.3582	0.104	3.444	**0.001**	[0.154, 0.562]
Health Insurance Cost (INS_COST)	0.1712	0.153	1.121	0.262	[-0.128, 0.471]
Health Problem (HLTH_P)	-0.2138	0.144	-1.484	0.138	[-0.496, 0.068]
Copayment Problem (COPAY_P)	-0.0080	0.093	-0.086	0.932	[-0.190, 0.174]

Note: Dependent Variable: OHS (1 = Dissatisfied, 0 = Satisfied).

Among the machine learning algorithms used in the previous stage of the study, SVM stood out as the model with the strongest prediction capacity. To overcome the "black-box" nature of SVM and validate its results within a robust theoretical framework, Logistic Regression was employed as a parametric benchmark.

According to the results in Table 4, where the dependent variable (OHS) is coded as 1=Dissatisfied and 0=Satisfied, the model provides high inferential power. The variables with the most dominant impact on the probability of dissatisfaction were Examination Satisfaction (EXAM_S) and Insufficient Health Staff (STAFF_P), both exhibiting strong negative coefficients (B=-1.141 and B=-1.176, respectively; $p<0.001$). These results indicate that higher satisfaction with the examination process and an adequate number of healthcare staff are the most critical factors in reducing the risk of overall dissatisfaction.

Conversely, Examination Problems (EXAM_P) (B=0.503, $p<0.001$) and Drug Cost Problems (DRUG_P) (B=0.576, $p<0.001$) yielded the highest positive coefficients. This confirms that as these process-oriented and financial barriers increase, the likelihood of a patient being in the "Dissatisfied" category rises significantly. Furthermore, the significance of Waiting Time Problems (WAIT_P) (B=0.358, p=0.001) reinforces the conclusion that operational speed is a non-negligible determinant of patient judgment.

The alignment between these logistic regression coefficients and the SHAP value distribution (Figure 2) ensures that the model is not only predictive but also inferentially consistent. For instance, the negative impact of Gender (B=-0.649, $p<0.001$) and the positive effect of Hospital Visit Frequency (B=0.128, p=0.044) provide a more nuanced understanding of the demographic drivers of satisfaction. In summary, the consistent directions of both the problem-focused and satisfaction-focused variables across different models confirm the reliability of the study's findings regarding the primary pain points in healthcare delivery.

Discussion and Conclusion

Machine learning is a significant tool that enables computers to acquire the ability to learn without explicit programming (15). This provides a major advantage in complex problem areas where predefined rules are insufficient (16). Machine learning models automatically discover hidden patterns from large samples, outperforming classical statistics in processing high-dimensional data such as medical records and repeatedly improving their performance as they adapt to new data (17). In this study, factors affecting dissatisfaction with healthcare services in Turkey were comprehensively analysed using machine learning and logistic regression methods. The findings reveal that demographic characteristics, socioeconomic status, and service delivery variables form a multidimensional and complex network of interactions that influence patient satisfaction. The capacity of machine learning models to uncover hidden patterns in high-dimensional data has significantly increased the interpretability and reliability of results obtained using classical statistical methods.

In the logistic regression analysis conducted within the scope of machine learning, dissatisfaction with healthcare services (OHS = 1) was considered as the dependent variable. According to the findings, the variables of age, gender, health satisfaction scale, frequency of hospital visits, examination problems, hygiene problems, examination satisfaction, insufficient healthcare personnel, medication cost problems, and waiting time problems were found to be statistically significant.

When demographic variables are evaluated, it is revealed that the variables of gender (GEN) and age (AGE) are statistically significant. The negative coefficient of the age (AGE) variable indicates that the likelihood of dissatisfaction decreases as age increases. This situation can be explained by the relatively lower expectations of individuals in older age groups regarding health services or their higher level of adaptation to the health system. The negative and highly significant level of the gender (GEN) variable indicates that women have a lower probability of dissatisfaction compared to men. This finding may be related to women developing a more harmonious interaction with the system due to their more frequent use of healthcare services or to differences in their expectations. Some studies in the literature prove the relationship between age and gender variables and satisfaction with healthcare services and support our findings (18–21).

The fact that the coefficient directions obtained from the logistic regression model show complete alignment with SVM-based SHAP analyses confirms that the model has both high predictive performance and high inferential validity. In particular, the fact that the FEE_P, WAIT_P and EXAM_S variables have the highest importance and impact levels in both analytical approaches clearly demonstrates the decisive role of financial burden, waiting times and the examination process in the healthcare experience. The fact that problem-focused variables have positive SHAP values in the SHAP summary graphs strongly supports the dissatisfaction-increasing effect of these variables.

According to the model results, variables related to the examination process are the strongest determinants of dissatisfaction. The negative and high coefficient of the Examination Satisfaction (EXAM_S) variable indicates that the quality of doctor-patient interaction, communication level, and the effectiveness of the clinical process play a critical role in patient perception. The negative coefficient of the Satisfaction (EXAM_S) variable, which falls under Staff and Relationship Quality, indicates that examination satisfaction significantly reduces the likelihood of general dissatisfaction with healthcare services. This confirms that positive clinical interactions are a primary factor in preventing negative patient outcomes. It is stated that if healthcare providers wish to minimise patient dissatisfaction, they should encourage their doctors to spend more time with patients and show genuine interest in their problems (28). Conversely, the positive coefficients of the Examination Issues (EXAM_P) and Waiting Issues (WAIT_P) variables reveal that operational process disruptions directly trigger patient dissatisfaction. Waiting times are considered one of the most critical factors driving negative patient experiences (25). For example, in Peru, both waiting and examination times were found to be key determinants of overall patient satisfaction (26). These findings suggest that process management and effective planning of patient flow in service delivery are fundamental elements in improving healthcare quality. Another factor affecting overall dissatisfaction with healthcare services is Hygiene Issues (HYG_P). Hygiene issues (HYG_P) also received a positive coefficient, indicating that deficiencies in physical facilities and cleanliness significantly increase the risk of 'dissatisfaction'. This is consistent with findings that patients prioritise the cleanliness of hospital facilities, such as toilets and common areas, as key indicators of quality (27).

The Health Satisfaction Scale (HSS) has a positive coefficient. This result indicates that as the general perception of health deteriorates, the likelihood of dissatisfaction with healthcare services increases. Individuals' negative perceptions of their general health status may lead them to evaluate the service provided more critically. The positive and significant nature of the Hospital Visit Frequency (HVF) variable indicates that the risk of dissatisfaction increases among individuals who visit healthcare institutions more frequently. This situation can be explained by the fact that individuals who have more contact with the healthcare system experience systemic failures more intensely and become more aware of them.

One of the noteworthy findings of the study is the high impact level of the Health Personnel Shortage (STAFF_P) variable. The coefficient of this variable indicates that shortages in personnel numbers and quality are among the factors that most sharply increase the risk of dissatisfaction across the system. This finding highlights that human resource planning and workload distribution in the healthcare system are critical not only in terms of service efficiency but also in terms of patient satisfaction. Furthermore, this finding is consistent with research findings showing that it is vital in reducing patient dissatisfaction levels and improving the overall perception of healthcare services (29).

When examining the economic dimension, the high positive coefficient of the Drug Price Issues (DRUG_P) variable indicates that drug costs are a significant source of dissatisfaction for patients. Perceptions regarding the cost of healthcare services greatly influence patient outcomes. This result highlights the negative impact of out-of-pocket healthcare expenditures on patient perceptions and points to the need to strengthen financial protection mechanisms. Similarly, the decisive role of financial burden on patient satisfaction is consistent with findings in the literature (30).

Overall, the results of this study reveal that patient satisfaction is shaped not only by clinical service quality but also by the holistic impact of organisational, economic, and sociodemographic factors. The combined use of machine learning and classical statistical methods offers the potential to produce more accurate, reliable and policy-guiding results in the field of healthcare management. In this context, it can be said that policies aimed at reducing waiting times in the healthcare system, improving clinical processes, increasing staff competence and reducing drug costs are priority areas for intervention in increasing patient satisfaction.

Declarations

Conflict of Interest

The authors have declared no conflicts of interest or financial support.

Funding

This study was conducted without any external funding.

Ethics Approval

This study does not require ethical committee approval, as it used secondary data. Permission to use the data was obtained from the Turkish Statistical Institute with application number 588938.

Availability of Data and Materials

The datasets generated and/or analyzed during the current study are available from the TÜİK.

Author's Contributions

All tasks related to the study, such as study design, data collection, data analysis, study supervision, manuscript writing, and critical revisions for important intellectual content, were performed by C.Ç.

Acknowledgments

None.

Referances

1. Birinci S, Guven SC, Ata N, Ulgu MM, Kucuksahin O. Patient satisfaction with health care services in Türkiye: A glimpse from nationwide Ministry of Health database. North Clin Istanb. 2024;11(6). doi:10.14744/nci.2024.84584
2. Batbaatar E, Dorjdagva J, Luvsannyam A, Savino MM, Amenta P. Determinants of patient satisfaction: A systematic review. Perspect Public Health. 2017;137(2):89–101. doi:10.1177/1757913916634136

3. Ferreira DC, Vieira I, Pedro MI, Caldas P, Varela M. Patient satisfaction with healthcare services and the techniques used for its assessment: A systematic literature review and a bibliometric analysis. Healthcare (Basel). 2023;11(5):639. doi:10.3390/healthcare11050639
4. Schmid L, Roidl M, Kirchheim A, Pauly M. Comparing statistical and machine learning methods for time series forecasting in data-driven logistics—A simulation study. Entropy (Basel). 2024;27(1):25. doi:10.3390/e27010025
5. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380:1347–1358. doi:10.1056/NEJMra1814259
6. Javaid M, Haleem A, Singh RP, Suman R, Rab S. Significance of machine learning in healthcare: Features, pillars and applications. Int J Intell Netw. 2022;3:58–73. doi:10.1016/j.ijin.2022.05.002
7. Çakmak MA, Kurt ME, Çakmak C. Makine öğrenmesi algoritmaları ile COVID-19 hastalarının mortalite risklerinin hesaplanması. Süleyman Demirel Üniv Vizyoner Derg. 2022;13(35):994–1011. doi:10.21076/vizyoner.1074212
8. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8–17. doi:10.1016/j.csbj.2014.11.005
9. Obermeyer Z, Emanuel EJ. Predicting the future—Big data, machine learning, and clinical medicine. N Engl J Med. 2016;375:1216–1219. doi:10.1056/NEJMra1606181
10. Oraini B. The effect of artificial intelligence capability on patient satisfaction. Int J Data Netw Sci. 2024;8(3):1429–1436. doi:10.5267/j.ijdns.2024.3.022
11. Chen J. A machine learning approach grounded in empathy theory. J Med Internet Res. 2024;26(4):e11686637. doi:10.2196/11686637
12. Uğurluoğlu Ö, Ürek D, Demir İB. Evaluation of individuals' satisfaction with health care services in Turkey. Health Policy Technol. 2019;8(1):24–29. doi:10.1016/j.hlpt.2019.02.003
13. Polatlı LÖ. Sağlık hizmetlerinde güncel makine öğrenmesi algoritmaları. DergiPark. 2023. Available from: https://dergipark.org.tr/en/download/article-file/2432612
14. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant. 2007;41(5):673–690. doi:10.1007/s11135-006-9018-6
15. Mahesh B. Machine learning algorithms – a review. International Journal of Science and Research (IJSR). 2018;7(6):381–386. ISSN: 2319-7064..
16. Raj A, D'Souza R. A review on machine learning algorithms. International Journal of Science and Engineering Development Research (IJRASET). 2019;7(4):792-796. Accessed February 10, 2026
17. Ezugwu AE, Ho YS, Egwuche OS, Ekundayo OS, Van Der Merwe A, Saha AK, Pal J. Classical machine learning: seventy years of algorithmic learning evolution. Data Intelligence. 2024. doi:10.3724/2096-7004.di.2024.0051
18. Cohen G. Age and health status in a patient satisfaction survey. *Soc Sci Med.* 1996;42(7):1085–1093.
19. Jaipaul CK, Rosenthal GE. Are older patients more satisfied with hospital care than younger patients? J Gen Intern Med. 2003 Jan;18(1):23-30.
20. Tian Y, Wong ELY, Qiu H, Wang H, Cheung AWL, Yeoh EK. Gender differences in healthcare experience: evidence from Hong Kong hospital surveys (2010-2020). Eur J Public Health. 2025 Oct 27;35(Suppl 4)
21. Alhasan MS, Alahmadi IM, Alhilali HF, Alyoubi NM, Alkhoriji NA, Aljazairi AM, Othman BM, Alshehri AM, Alhejaili MO, Alhejaili B, Alanazi OI, Azzam AY, Alhasan AS. Determinants of patient satisfaction among outpatients with chronic illnesses in the region of medina, Saudi Arabia: a cross-sectional study. BMC Health Serv Res. 2025 Jul 14;25(1):953. doi: 10.1186/s12913-025-13116-7. PMID: 40660224; PMCID: PMC12257647.
22. Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: A meta-analysis. Soc Sci Med. 1990;30(7):811–818. doi:10.1016/0277-9536(90)90205-7
23. Thi HM, Lee WY, Lin CC. Factors determining patient satisfaction with nursing care in a teaching hospital. J Nurs Res. 2002;10(3):169–178. doi:10.1016/S0277-9536(01)00045-4
24. Paul P, Hakobyan M, Valtonen H. The association between self-perceived health status and satisfaction with healthcare services: Evidence from Armenia. BMC Health Serv Res. 2016;16(1):67. doi:10.1186/s12913-016-1309-6
25. Mosleh SM, Alsereidi AR, Aldhanhani AA, Alnaqbi HM, Alhouti RS, Alshehhi SS. A descriptive study on patient satisfaction with waiting time in emergency departments: Insights from hospitals in the Northern Emirates. Int Emerg Nurs. 2025;78:101564. doi:10.1016/j.ienj.2024.101564
26. Alarcon-Ruiz CA, Heredia P, Taype-Rondan A. Association of waiting and consultation time with patient satisfaction: Secondary-data analysis of a national survey in Peruvian ambulatory care facilities. BMC Health Serv Res. 2019;19(1):439. doi:10.1186/s12913-019-4288-6
27. Semiz BB, Semiz T, Güneş K. Sağlık kurumlarında algılanan hizmet kalitesinin ve hasta memnuniyetinin ölçülmesine yönelik bir araştırma. Bus Manag Stud Int J. 2023;11(3):1130–1151. doi:10.15295/bmij.v11i3.2289
28. Senić V, Marinković V. Patient care, satisfaction and service quality in health care. Int J Consum Stud. 2013;37(3):312–319. doi:10.1111/j.1470-6431.2012.01132.x
29. Vatica JR, Nur'aini NA, Lubis M. Effect of health service quality toward patients satisfaction. J La Medihealtico. 2021;2(1):63–72. doi:10.37899/journallamedihealtico.v2i1.308
30. Nguyen T, Nguyen H, Dang A. Determinants of patient satisfaction: Lessons from large-scale inpatient interviews in Vietnam. PLoS One. 2020;15(9):e0239306. doi:10.1371/journal.pone.0239306