Classification of service delivery to airport passengers using data mining

Reza Moosaei [1], Ali Asghar Safaei [2],*

[1]Department of Computer Engineering, Qeshm International Branch, Islamic Azad University, Qeshm, Iran
[2]Department of Medical Informatics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

A B S T R A C T

In this paper a model has been offered with the aim of improving quality for classifying airport passengers using data mining offering services. The provided model is able to separate and classify passengers according to performance and places them in different classes. As a result the offered airport services would be according to these classifications which can be useful in customer satisfaction and efficient services and resource efficiency. This paper focuses on customer's satisfaction and data mining methods and feature selection has been used for improving classifying services to passengers. The Random Tree method has been used as the best method for learning and feature selection method for interest information (IG) with 93.88% efficiency has been used. 12 features from 49 features have been selected by using interest information. Results indicated that proposed model for classifying offering services for passengers this model has good and acceptable performance. Also PCA method has improved the simple, LIBSVM, and Random Forest method.

1. Introduction

Removing policy rules in airport industry has reduced some pressures from airport. And also it has become to an intense competition across the world. Normally an airport that uses of efficient facilities and top-level services (LOS) would be selected by passengers. Thus airport managers are faced with new rivals in the commercial pressures (Suykens and Vandewalle, 1999). One of the main and most important challenges of them is that they are unable to attract passengers for airports. In previous studies more focus was on operational standards according to queuing time, service processing time, space, physical facilities etc. Of course recently researchers are willing to attract passengers through aviation transport (Büyüközkan and Çifçi, 2012). Although different various aviation agencies such as Federal Aviation Administration (FAA), Airports Council International (ACI) and Transport Canada (TC) have begun to develop a methodology for high level evaluating and improving services yet despite these efforts and proposed standards this method has been criticized by airport professional. The most of previous studies have focused only in the airport passenger, terminal building parts (for example checking entrance door, the departure lounge etc.)

and have ignored other components such as parking, transport links (Michie et al., 1994).

The main idea of this research is improving the quality of airport services and passenger attraction which in spite of other methods (that have paid attention to typical data analysis) will use of data mining techniques. The process of improving quality service would stimulate passengers to airport selection. In this article a model will offered for classifying passengers that will help airport managers to create new strategies and useful frameworks for organizing airport systems that help them to provide services to passengers (Yao and Zhao, 2009). The passengers are classified according to some characteristics such as age, profession, degree education and some other features. And the services to each category will offer according to provided criteria. Therefore by offering classified and objective services for passengers while better efficiency of airport resources, the better efficiency would be obtained (Chang and Lin, 2011).

Therefore in this paper learning and classifying machine algorithm such as SVM and Kiminz over the learning data collection would be used. accordingly by classifying passengers to different classes , providing differentiated services and fits with each class and category with passengers a decision support for airport managers would design and proposed which help them in offering services and will be lead to better resource efficiency, better performance and satisfaction of the system of airport

* Corresponding Author.
Email Address: aa.safaei@modares.ac.ir (A. Safaei)

passengers. the necessity and motivation to engage in this research in the first step fixing problems services for passengers and then creating solutions for more attracting of passengers for airports. The entries that we will have in this study the records of passengers during so many years' passengers' records in so many years and we will analyze them based on a series of algorithms.

The necessity and motivation to do this study primarily to solve services problems to passengers and then creating solutions to better attraction of passengers to airports; The inputs which we have in this research are passenger's experiences for many years and our analysis is according to some algorithms and in the end we would have the output of a bunch of passengers and service improvement for each category. This research enjoys a special

importance for airport managers. The managers are needed to identify the requirements of their passengers in different categories to have planning such as improving the welfare and services. This study provides this need of managers. Also in the airport the speed of access to passenger's information is very important.in this paper the access speed to information would be increased by passengers categorize and separation. The other importance of this research is helping to choose the different approaches of quality improvement.

The required steps for improving offered services according to analysis of registered data and creating classified model by using data mining is indicated in Fig. 1.
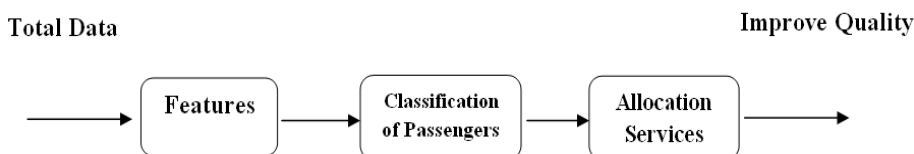

Fig. 1: the chain of improving service offering

2. Research literature

One of the most important researches that had been done in the field of airport classification we can mention classification of the Florida airport which has been developed for Florida department of transportation and it is prepared through council's airports of Florida and a consortium of development of air systems in 2006.

This classification which has been performed based on education and service levels with offering standards for each of the services and facilities and classifying each of them into 7 class ,has classified and evaluated 129 airport of Florida from F to A. In this classification, the classification parameter is on the basis of facilities and 1 to 7 is according to service levels. The classification of service levels of airport includes services of passengers and the pilots and due to broad service level, they are divided into two service types in mandatory and desirable level (Hall et al., 2009).

Another research has been performed in relation to ranking airports at Virginia. Virginia airport in 2003 .Classification system OF Virginia airports determines duties and performance of the airport, the optimal reference code of the airport and leading strategy toward offering better services. According to this classification, the type of airport performance and roles of the airport can be classified into 5 groups that are as following:
1. Commercial service airports;
2. Airports and utilities;
3. Regional Airport general aviation;
4. General aviation airports
5. Local service airport

In this article the efficient methods for assessing airport performance and determine a suitable

pattern for improving performance of airport. The analysis is according to data envelopment analysis, partial productivity index, overall productivity and customer satisfaction which is a valuable recommendation for operational managers in determining the effectiveness and efficiency of current operations and future goals to improve. The emphasis in this paper is on satisfaction of customers yet we has used data envelopment analysis for determine efficiency relationship of airports and then we will use overall productivity index and partial productivity index models for determine pattern for improving performance of airport. In this methods the entries are include operating costs and number of employees and outputs include operating revenues, number of passengers and number of flights. This paper presents a framework for evaluating customer service in the air transport sector in the United Arabic Emirates and also deals to compare the three main airports in the United Arabic Emirates. In a new approach, an alternative method provided to assessing the quality of services. Specifically a phase model for service concept servers is considered for the basic service quality standards while the multi-criteria method has used for making decisions of ELECTRE III which refers to quality ranking for service replacements. An empirical study on evaluating the service quality in order to showing proposed method has done from three international airports in Sicily, Italy with a precise design for improving passenger service. Results indicated that only a few aspects services play a key role in quality services of airports. In addition the effects in evaluating quality services calculated as customer's uncertainty (Sebastiani, 2002).

3. Proposed model

In most studies the efficient methods are proposed for evaluating airport performance and determine appropriate and helpful pattern to improve the performance of airports. The analysis is according to data envelopment analysis, partial productivity index, overall productivity index and customer satisfaction which are valuable proposition for operational managers in determining efficiency and effectiveness of current operations and future goals. The necessity and motivation to engage in this article are primarily fix problems of passenger's services then creating solutions for more attracting passengers for airports. The entries which we would have in this research are records of passengers for many years and we will analysis according to a series of algorithms and finally the output of a series of categories of passengers and improve services for each category.

This article enjoys of particular importance for airport managers, managers are required to identify the needs of their passengers in different categories in order to have planning's for improving the welfare and services.

In spite of statistical methods machine learning techniques is good in conditions which there is no primary knowledge about data patterns for this reason sometimes this methods are called down to up , One of the main benefits of using machine learning techniques is that act very quickly and are cost-effective. Machine learning techniques in data mining generally divided into two categories with supervision and without supervision. An issue that has been studied in this article should be used from with supervision technique. Also in learning phases of decision tree J48, Random Fours, Random Tree, SVM and LIBSVM, naïve Bays has been used for data learning and experiment.

3.1. Flowcharts of steps of proposed model

The overall steps in creating a decision model by using data mining techniques and learning machine can be divided into two steps education (learning) and testing. In the learning phase, the model are formed and evolves using the training data then in test phase the performance of created model will assessed with test data. Steps taken to develop and evaluate proposed classification model is indicated in Fig. 2.

As stated previously, data collection is used of Sanfransisco International Airport .this data collection is the most up to date data collection which has been collected from 2011.in this dataset first after receiving dataset it should preprocessing done .preprocessing is include removing duplicate data, delete no value data, delete Data that is entered by mistake instead of the number and vice versa. After all data were similar to each other it should popular features such as method of information extraction methods (IG) and principal component analysis (PCA) used. In this section features are so

practical that are more useful in service classification (Adeli and Hung, 1994). And finally waste and other useful features are removed in order to in better classification documented. In following learning algorithms should be used. Learning methods should be divided into two classes. Firstly the data are educated then and finally the rest data experimented. In this case by default the same 40-60 has been used. In this research we used the more practical algorithms which we would describe them in detail at next session. The used parameters in classifications Precision, Recall،F-measure, Error classification were described which the most important parameter is Accuracy. He more this parameter is that the system we proposed would be more credible.

WEKA tool is include a collection of machine learning algorithms for data mining tasks, algorithms can also apply directly to data collection or may be called by Java code. Also machine learning tool WEKA includes a tool for primary preprocess classification, regression, clustering, association rules, and visualization. This tool can be useful for improving new designs of learning machine. WEKA tool include other tools such as the probe, testing, data flow and knowledge, a simple command line interface and java interfaces. IG and PCA methods are used for better extraction of features. Following the first review will be done with simple methods and the results are examined with extraction methods.
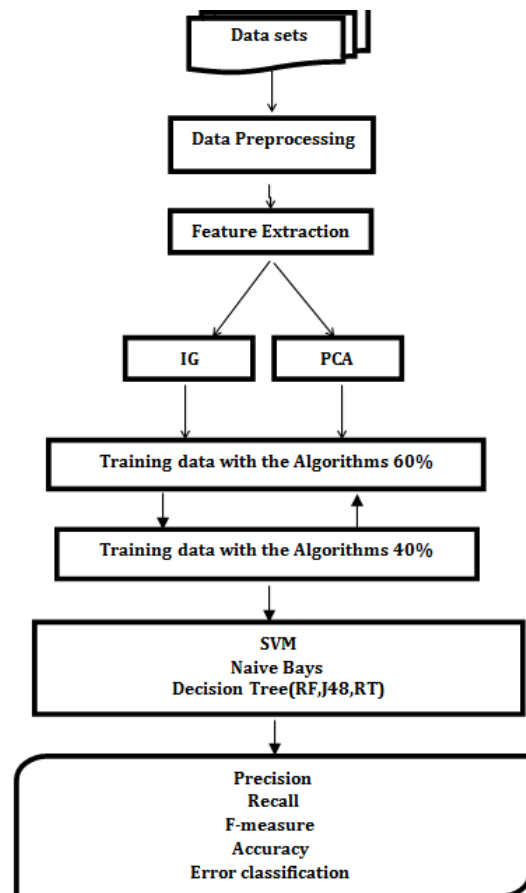

Fig. 2: flowchart of the steps to proposed classification model

3.2. Setting the parameters of learning algorithms

Before learning the implementation of the necessary parameters are set as follows: According to the tests for SVM algorithm the linear kernel and C-SVC type was used for vector machine. In vector machine support for c parameter which in fact establishes a relation between border width and wrong separation? According to test different values, the value of 3 was selected. In fact considering tested experiments for SVM algorithm, this parameter is called regulator which it is the user's choice. According to the tests for KNN algorithm the value K = 3 is selected from linear kernel. For algorithm Random Forest, the depth of the tree is considered a maximum of 100, this parameter is default and we had not change it. For random tree algorithm variance parameter for regression trees is considered as 0.001 maximum, this parameter is default and not changed. One of the most important parts that need to be considered in the classification is evaluation of classification models. This section will examine the assessment criteria for classification. Evaluation of a classification model can be done according the training and test samples.

3.3. Effectiveness model of evaluating parameters

For assessment a label that classifying model attributed to recorded attribute should be compared with a label that recorded is belonged to it. Occurring of different states for records and classes according a collection of entering data with values TP, FP, TN, FN is indicated for two classes positive and negative and shown in Table 1.

Table 1: Different moods for categories and people

estimated		
positive	negative	
TP	FN	positive
FP	TN	negative

The most important standard for determine the performance of an algorithm is accuracy classification standard. This measure calculates the accuracy of a classification. This measure indicates that what percentage of the total data set is classified correctly. The Eq. 1 indicates the calculations of correction standard.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

Two values TN and TP are the most important values that must be maximized. In order to the efficiency of classification be maximized. Considering this point which in classifying issues may be between the numbers of samples in different categories no balance is not established. It is possible a category have more samples from other categories as a result a final model driven to the category with

the highest sample therefore a category which have the lowest samples will not have a great impact on improving and lack improve performance. It can be resulted that accuracy standard is not a proper standard in data collection that have unbalanced categories. Classification standard error is achieved of the Eq. 2.

$$ER = \frac{FN+FP}{TP+FP+FN+TN} = 1 - \text{Accuracy} \tag{2}$$

Precision standards show the document percentages that which is classified properly. In other words, the accuracy of classification categories to suit all cases indicate that sticker i for sample classification suggested by j. How to calculate this metric in was shown Eq. 3. Subscript i on these parameters mean that the parameters must be calculated for each category of i.

$$Precision_i = \frac{TP_i}{TP_i+FP_i} \tag{3}$$

Recall criteria for each category, the percentage of records that all the documents belonging to that category, is properly classified. In other words, the accuracy of classification categories i based on the total samples shows labeled i. How to calculate this metric is shown in Eq. 4.

$$Recall_i = \frac{TP_i}{TP_i+FN_i} \tag{4}$$

Significant point is that the performance criteria are classified according to the number of category I. While Precision criteria based on the accuracy of prediction and suggest that to what extent we can trust the output classify. F1 standard obtained of combination Precision and Recall metrics and is used in cases that could not distinguish between each of two criteria Precision and recall. Eq. 5 shows how to calculate this measure.

$$F1_i = \frac{2*Precision_i*Recall_i}{Precision_i+Recall_i} \tag{5}$$

4. Results and discussion

In this article it has been used of the most widely used algorithms such as tree algorithm J48, support machine, on simple, Random Forest, Random tree which data have been analyzed with category parameters such as Precision, Recall, F-measure, and Error classification which the most important parameter is Accuracy which the more this parameter our system would be more reliable. In the following results of each case reported.

4.1. The results of simple method

In this sector we provide a report on the methods with the entire feature (Fig. 3).

In the Fig. 3 the results of comparison all learning methods have offered which the best method is

Random Tree among other methods; The interesting point is the efficiency of decision trees which after Random Tree the J48 algorithm and Random Forest can be better selected.

4.2. The results of methods to extract useful IG features

In this part feature selection feature or IG are using .in the following table all features alongside the value of feature, number of feature and feature name are shown in Table 2.

Table 2: The value of each feature by IG method

Value and Rating Feature	Feature number	Feature
0.39748	2	RESPNUM
0.13649	9	AIRLINE
0.09374	18	Q7
0.08706	11	Q2_1
0.06624	12	Q3_1
0.06034	6	DESTGEO
0.05475	46	Q17
0.05223	31	Q8M
0.04334	29	Q8K
0.04324	42	Q14
0.04279	44	Q15B
0.04036	3	CCGID
0.04023	37	Q10E
0.03906	27	Q8I
0.03789	38	Q10F
0.03523	28	Q8J
0.03387	39	Q11
0.03373	10	GATE
0.03298	21	Q8C
0.03262	5	DESTMARK
0.03222	34	Q10B
0.03178	20	Q8B
0.03072	23	Q8E
0.02968	25	Q8G
0.02914	19	Q8A
0.02894	15	Q4C
0.02812	32	Q8N
0.02771	30	Q8L
0.02746	24	Q8F
0.02728	35	Q10C
0.02717	1	RUNID
0.0265	36	Q10D
0.02353	26	Q8H
0.02312	40	Q12
0.02241	7	TERM
0.02239	4	LATECODE
0.02051	33	Q10A
0.0199	14	Q4B
0.01986	22	Q8D
0.01626	45	Q16
0.01431	48	Q20
0.01411	41	Q13A1
0.01405	17	Q6
0.01137	47	Q19
0.01064	16	Q5
0.00988	43	Q15A
0.00976	13	Q4A
0.00807	8	ATYPE

4.2. The useful extracted features

In the following the useful extracted features among all features select and they has used for decision making for reaching proper quality (Table 3).

Table 3: List of selected features using feature extraction IG

Value and Rating Feature	Feature number	Feature
0.39748	2	RESPNUM
0.13649	9	AIRLINE
0.09374	18	Q7
0.08706	11	Q2_1
0.06624	12	Q3_1
0.06034	6	DESTGEO
0.05475	46	Q17
0.05223	31	Q8M
0.04334	29	Q8K
0.04324	42	Q14
0.04279	44	Q15B
0.04036	3	CCGID

In the Fig. 4 you see the overall results which the most important method is Random Tree method among other methods. The interesting point is the efficiency of decision trees. After Random Tree, j48 and Random Forest algorithm can be selected better.

You see in Fig. 5 which in some methods the efficiency of interest information is weakened but in the decision tree, Random Tree and Random Forest are still the best and the efficiency has more better yet in the LIBSVM and SVM method the efficiency had been better in comparison with a method that was assessed with all features. But it has not the highest efficiency in comparison with Dassin Tree.

4.4. Results methods with extract useful features

In this section feature selection method can be used for decision making to classifying of people to reaching to suitable quality.

The overall result you can see in Fig. 6 that the best method is the Random Tree method. The interesting point is the efficiency of decision trees which after Random Tree, J48, and Random Forest algorithm can be selected better. Also Libsvm method has improved very much.

4.5. Comparison of studied learning methods

In Fig. 7 you can see the overall result, which the best method is Random Tree method and interest information of feature selection (IG). With using the information feature selection method was used to select 12 features from 49 features. Therefore is suitable for the decision to classify the passenger's services. In this Fig. you can see that PCA method causes to improving Naive Bays method, LIBSVM and Random Forest (Table 4).

5. Conclusion

Those who involved in the airline industry believe that they had sold credit tickets by this method and they are able to planning on their potential passengers, this process was continued until airlines sales managers recognized that they should affiliate more privileges of potential passengers to themselves more than past. Service quality by means of maintaining customer support has impact on a firm's competitive advantage; finally, will follow profitability and market share. Providing high quality services as increasing competition in the airline is converted into market needs. Airline services are consisted of different process. The passengers may have distinct expectations in the different steps of service offering chain. In order to provide better service to airline passengers the companies are required to understand their expectations and their needs.

Studies on other sections show those Judgments concerning customer satisfaction and quality service including customers who want to compare their expectations with their actual performance. In the context of air transport factors such as service value and corporate image are tend to be ignored. Such neglect can trigger some problems due to wrong diagnosis and weak predictive.

As usual key variables that are considered during modeling decision-making process of passengers are include Service expectations, service perception, service value, customer satisfaction, and image of airline corporation. Understanding what consumers expect from a service organization is important because expectations provide standard criteria for judging consumers. Aerial services are integrated chain of services which they are divided into a number of processes.
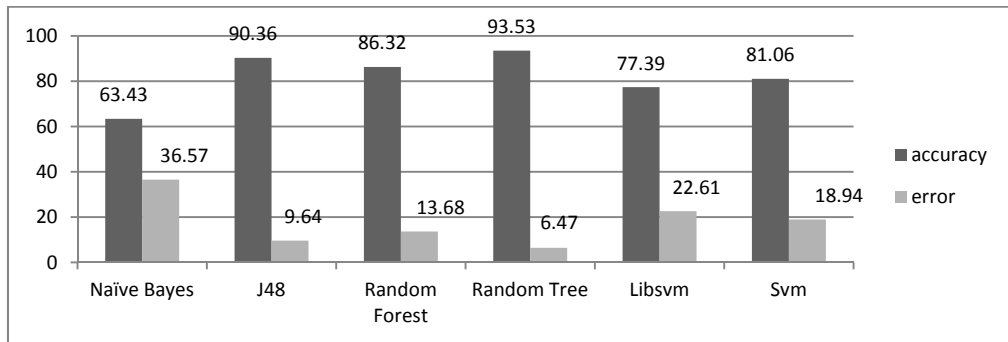

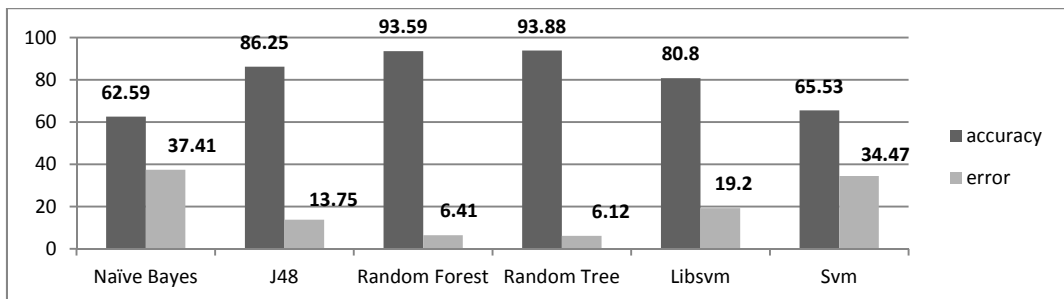**Fig. 3:** Compare studied learning techniques (with all the functionalities)

Fig. 4: Compare all methods with 12 features
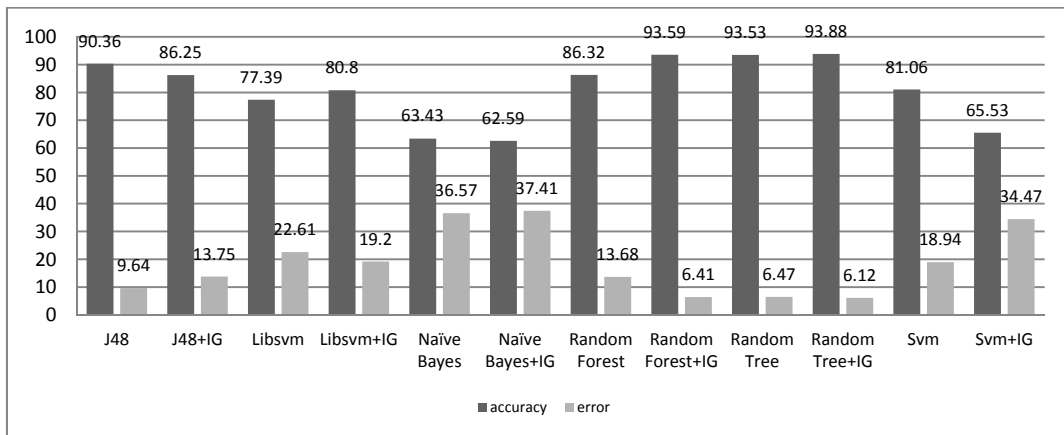

Fig. 5: Compare all methods of feature selection method
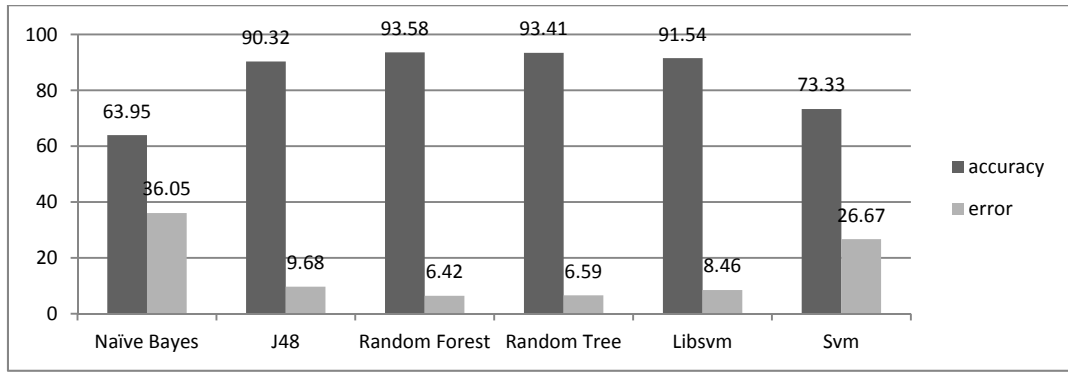
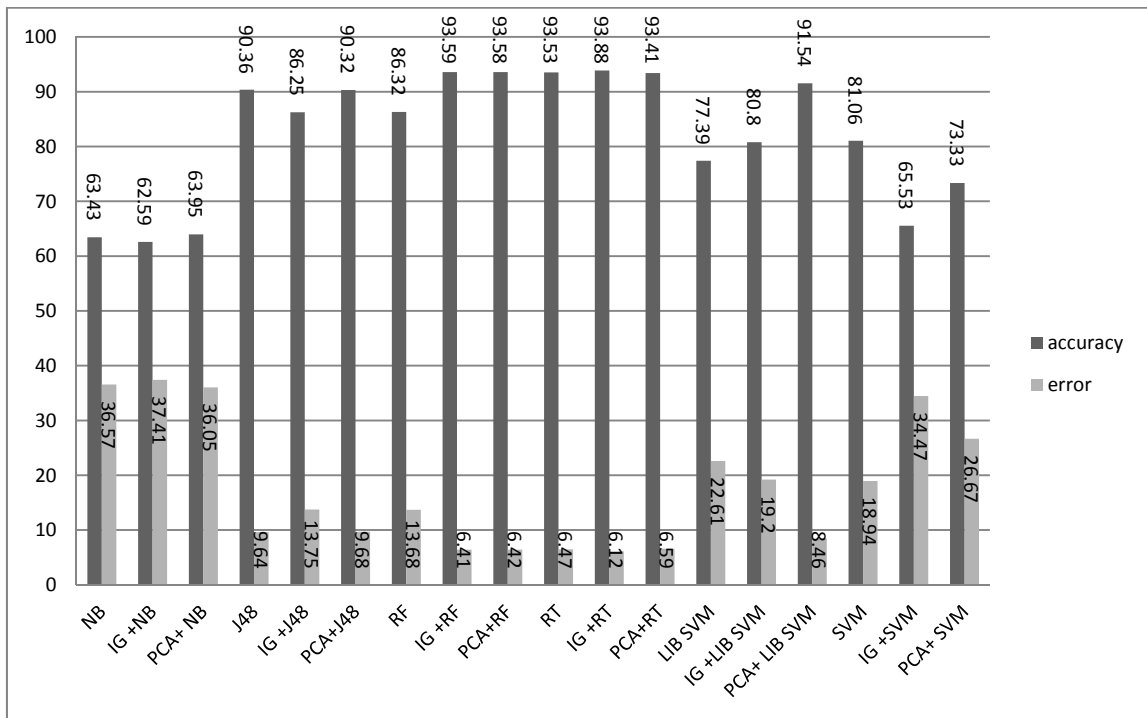**Fig. 6:** Compare all methods with PCA feature selection methods

Fig. 7: the status of studied learning methods in terms of accuracy to the error

Table 4: The status of the studied learning methods in terms odd accuracy

Learning method	Accuracy
Naïve Bayes	63.43
IG +Naïve Bayes	62.59
PCA+ Naïve Bayes	63.95
J48	90.36
IG +J48	86.25
PCA+J48	90.32
Random Forest	86.32
IG +Random Forest	93.59
PCA+ Random Forest	93.58
Random Tree	93.53
IG +Random Tree	93.88
PCA+ Random Tree	93.41
Libsvm	77.39
IG +Libsvm	80.8
PCA+ Libsvm	91.54
Svm	81.06
IG +Svm	65.53
PCA+ Svm	73.33

Passenger's expectations of service quality, in various stages of service delivery process are different. Given the importance of air transportation industry, an air travel is divided into two phases:

Ground services and in-flight services for example, ground services include providing information, keeping fit and ticket sales, passenger and cargo services and post-flight services. As air passengers, especially on international routes, spend most of their time in flight the fight service quality required more attention from airline corporations. In cases where a lot of activity in the airport is expected before the final decision, detailed analysis of the types of projects must be undertaken. The geometry of the front of the plane, make-up and capacity levels and dimensions is effectively to select the passenger terminal. However, in the initial assessment and the airport prior to locating the various elements of the design must be limited to two or more designs. Quantitative measures of the plan walking distance the required surface area etc. must be compared and efficiency of passenger flow and airplane ground vehicles and sequencing, functional and operational components to be investigated .in airports with very actively Aviation using simulation technology could be justified.

Considering the importance of providing services in the airports to passengers we can recommend some research topics such as the following items:
- Providing a model for optimal management of airports based on their performance
-to prioritize the strategies and optimize them to fix weak points and shortcomings
-to provide a model for optimal allocation of funding of airports by considering the maximization of resource efficiency of the airport

References

Adeli H and Hung SL (1994). Machine learning: neural networks, genetic algorithms, and fuzzy systems. John Wiley and Sons, Inc. (In Persian).

Büyüközkan G and Çifçi G (2012). A combined fuzzy AHP and fuzzy TOPSIS based strategic analysis of electronic service quality in healthcare industry. Expert Systems with Applications, 39(3): 2341-2354.

Chang CC and Lin CJ (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3): Article 27, 27 pages.

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P and Witten IH (2009). The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter, 11(1): 10-18.

Michie D, Spiegelhalter DJ and Taylor CC (1994). Machine learning, neural and statistical classification. Ellis Horwood Series in Artificial Intelligence, New York, NY: Ellis Horwood,| c1994, edited by Michie, Donald; Spiegelhalter, David J.; Taylor, Charles C., 1

Sebastiani F (2002). Machine learning in automated text categorization. ACM Computing Surveys (CSUR), 34(1): 1-47.

Suykens JA and Vandewalle J (1999). Least squares support vector machine classifiers. Neural Processing Letters, 9(3): 293-300.

Yao Y and Zhao Y (2009). Discernibility matrix simplification for constructing attribute reducts. Information Sciences, 179(7): 867-882.