Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01432-w

ORIGINAL RESEARCH

A living environment prediction model using ensemble machine learning techniques based on quality of life index

Zülfiye Erdoğan[1] · Ersin Namlı[2]

Received: 26 August 2018 / Accepted: 1 August 2019

Abstract
The living environment is an area that has the features necessary for all people to keep living their lives. Nevertheless, as time has progressed, needs and economic capacities of people have changed; consequently, the living environment has begun to express different meanings for different people. This differentiation brings the following question to mind: Can every person live in the same living environment under the same conditions? Of course not! People differ from each other in their preferences, needs, economic capacities, and many other aspects, so they must be assigned living environments that best fit their unique features. In this study, machine learning (ML) techniques and ensemble machine learning (EML) techniques are used for the establishment of a living environment prediction model. The data required for the quality of life index are obtained from a questionnaire. This questionnaire was prepared to consider people's desires and economic capacities. Quality of life index is computed taking the weighted arithmetic mean of questionnaire data. Calculated quality of life indexes of the individuals is assigned to available quality of life classes of the cities. ML-based classification techniques have been used for prediction of assigned classes of the cities. A living environment prediction model is proposed in this study by using an ML and EML techniques-based quality of life index. For predicting living environment, ML-based methodologies include an artificial neural network (ANN), support vector machines (SVM) and EML-based methodologies include stacking and voting. The prediction results of ensemble models are better than prediction results of individual ML models, especially the stacking-based model composed of SMO + LMT and which reached the best performance values using the 80% split method.

Keywords Living environment · Quality of life index · Ensemble machine learning · Artificial intelligence

1 Introduction

Quality of life is a concept that is rather amorphous, difficult to measure, multidimensional, multifaceted, affecting and interacting with various living environments (Walker and Lowenstein 2009). Quality of life can be defined as the sum of an individual's biological and physiological conditions, lifestyles, and personal and social relations (Seker 2011). There are many definitions of quality of life in the literature. However, in general, the components of quality of life and the methods used to calculate quality of life index are very similar and, in most studies, the same methods and components are used. Many components constitute quality of life, including health, education, and the environment. The quality of life index should include all of these components. Furthermore, the weighted average is taken using the weight of each component in calculating the quality of life index. The weights of the components are determined by experts. This method and weights of the components are used in the first stage of the proposed methodology.

The living environment is one of the main condition that exerts an influence on people's life quality. The living environment has been studied and examined from different perspectives for various disciplines such as the health economics, environment, etc. The livability in a natural environment is defined as pollution, global warming and degradation of nature by ecologists. City planners can associate livability in the structured environment to sewage systems, traffic congestion (Michalos 2014). In general, there is no specific

✉ Ersin Namlı
enamli@istanbul.edu.tr

Zülfiye Erdoğan
zulfiye.erdogan@iste.edu.tr

[1] Department of Industrial Engineering, Iskenderun Technical University, Hatay, Turkey

[2] Department of Industrial Engineering, Istanbul University-Cerrahpasa, Istanbul, Turkey

definition of living environment. Living environment is a complex and multidimensional concept representing spatial and non-spatial, economic, social and environmental dimensions (Michalos 2014; Carmona and de Magalhaes 2007). People's attitudes and behavior are directly related to the environments in where they live (Lewin 1939). The reason for this is the continuous interaction with their surroundings. People's expectations of the living environment are different. It is important for individuals to live in a living environment that can meet their needs. However, the individual has the economic capacity to meet the requirements of this environment. Maintaining the balance between these requirements and responding to requests is essential. People should be assigned to the living environments that have better standards by which they can live in terms of their development and life standards in the direction of their current features and possibilities. The living environment is a multifaceted concept that affects the health and development of people both psychologically and physically. It is important to choose the most suitable living environment for people. Therefore, in the first phase of the proposed methodology, the qualities of life of the cities that individuals want to live were calculated. The calculated indexes and the current quality of life indexes of cities were compared. Then people are assigned to the most suitable living environment.

Quality of life in literature, it is seen that studies have been used to measure the life qualities of cities. Therefore, these studies do not predict; they only measure current situations. However, this study aims to predict living environment, which has best-fit standards by which individuals can live using the measured life qualities of cities. The predictability of the city quality that obtained from previous calculations are used for building a ML based predictive model. The represented new model has eliminated the calculation step and offers a suitable living environment to individuals who without any mathematical calculations.

ML is a process consisting of three parts: representation, prediction of information, and learning mechanism (Kovalerchuk and Vityaev 2000). In the literature, the ML and EML are used in many areas (Unlu and Xanthopoulos 2019a; Erdal et al. 2013; Namli et al. 2016; Erdal 2015) including credit scoring models (Erdal and Karahanoğlu 2016; Zhou et al. 2010), stock market analysis (Hassan et al. 2007), prediction models such as energy consumption (Namli et al. 2019; Chou and Ngo 2016; Namli and Yucel 2018; Turkan et al. 2016; Ekinci and Erdal 2015), purchasing prediction (Gupta and Pathak 2014), housing prices (Park and Bae 2015), climatic conditions (Kisi et al. 2016), and tender offer amounts for projects (Chou et al. 2015) etc. However, ML and EML techniques are not used in studies of quality of life index. Therefore, this study is different in terms of method and goal from the studies related to the other living environment in the literature.

The rest of this paper is organized as follows. A brief overview about the studies related to ML, EML, and quality of life in the literature is provided in Sect. 2. ML and EML techniques and performance evaluation methods are described in Sect. 3. In the Sect. 4, modeling experiments examines the experimental settings and predicts the results to be obtained from each of the techniques as compared to various other methods are given. Finally, the results are reported and discussion of these results and conclusions that drawn from the study.

2 Related work

Machine learning as a field of artificial intelligence recently has a rapidly spreading popularity due to its accurate prediction outcomes. ML algorithms have the ability to learn from past data and they are very useful tools as decision support systems for scientists and researchers. Baesens et al. (2002) used models based on statistical and neural network algorithms for modeling Purchase incidence. Hao et al. (2004), Sangwan et al. (2015), Suresh et al. (2011), and Avramović et al. (2015) applied models based on genetic algorithms (GA)-ANN in optimization problems. Hassan et al. (2007) forecasted using a fusion model by combining the hidden Markov model, ANN, and GA for financial market behavior. This method is used for large-scale learning. Zhou et al. (2010) applied least squares-SVM (LSSVM) for building credit scoring models. Panda et al. (2010) applied discriminative multinomial Naive Bayes with filtering analysis. An incident information management framework was built using data integration, data mining, and multi-criteria decision making (MCDM) in the study (Peng et al. 2011). Karimi and Yousefi (2012) offered a nanofluids density prediction model using ANN and GA. SVM, an adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS), and regression techniques were used for estimation of reference evapotranspiration (Tabari et al. 2012). Suspended sediment was modeled using LSSVM. Two different algorithms were used for ANN applications, and the LSSVM model predicts better results than the ANN models (Kisi 2012). ANN with an artificial bee colony (ABC) algorithm was used for modeling a discharge-suspended sediment relationship by Kisi et al. (2012). Khademolqorani and Hamadani (2013) proposed an integrated approach that included data mining and MCDM in decision support systems. Kou et al. (2014) used evaluation of clustering algorithms for financial risk analysis based on MCDM (TOPSIS, DEA, and VIKOR). Chou and Pham (2014) proposed a GA-based SVR model for predicting bridge scour depth. The monthly temperature values are modeled by MLR, ANN, and ANFIS models (Cobaner et al. 2014). Aghdaie et al. (2014) offered an integrated approach that includes cluster analysis, SWARA, and VIKOR as two multiattribute decision makers in supplier clustering and ranking. The extreme learning machine was offered as a learning algorithm to ANNs by Xue et al. (2014). Zuo et al. (2014) used SVM for studying extreme consumer purchasing behavior. Gupta and Pathak (2014) applied a k-means clustering algorithm for grouping customers. A novel ensemble model of the extreme learning machine was proposed for predicting reservoir properties by Anifowose et al. (2015). Bose and Chen (2015) proposed a fuzzy c-means clustering algorithm that examines how customers dynamically move between clusters. Therefore, purchase behavior of online customers is predicted by customer segmentation on dynamic pricing. The EML techniques (Voting, BSA, Stacked, and MetaCost) and classification models (Naive Bayes, logistic regression, decision trees, and SVM) are applied in pay-per-click campaign management by King et al. (2015). Kisi et al. (2015) proposed a novel approach based on SVM coupled with firefly algorithm (FA). FA predicts SVM parameters. The extreme learning machine models are used for prediction of reference evapotranspiration (Abdullah et al. 2015). A hybrid approach, including MCDM techniques and ML algorithms were proposed for multi-attribute inventory classification by Kartal et al. (2016). The Chi-squared automatic interaction detector, neural networks, classification, and regression tree were used to predict daily pan evaporation (Kisi et al. 2016). This method is used for building a network intrusion detection system. Maghrebi et al. (2016) proposed an ensemble learning algorithm for matching experts' decisions in concrete delivery dispatching centers. Heinermann and Kramer (2016) analyzed heterogeneous ML ensembles for wind power prediction. A parallel ensemble of the online sequential extreme learning machine algorithm based on MapReduce was proposed by Huang et al. (2016). Raikar et al. (2016) investigated the model based on GA-ANN in the prediction of scour depth within channel contractions.

On the other hand, increasing quality of life is key concept for governments, municipalities and policy makers so especially in recent years, many researchers and scientists have been studying quality of life and related studying fields. Endicott et al. (1993) evaluated the performance characteristics of quality of life and satisfaction through a questionnaire for adult patients who wanted to be treated for depression. Felce and Perry (1995) suggested a model of quality of life that combines objective and subjective indicators, living environments, and individual values. The World Health Organization Quality of Life (WHOQOL) Group (1995) explained a tool of the WHO to improve the quality of life in the study. The trends of quality of life in the living environment of Slovenia were examined by Christensen (1996). The urban quality of life in Singapore between 1997 and 1998 was assessed by Seik (2000). Cummins et al. (2013), developed a national index of subjective wellbeing and applied it to Australian. Avci and Pala (2004)evaluated the quality of life of research assistants and specialist doctors working in medical faculties. Quality of life of Guwahati in an urban environment was examined by Das (2008). Abdullah and Tap (2009) computed the quality of life index of three selected states in Peninsular Malaysia by using a mathematical model based on the fuzzy set theory. Lazim and Osman (2009) presented a mathematical model based on the fuzzy set theory and a new way of expressing the quality of life index based on Maslow's hierarchy of needs. Das et al. (2012) investigated the potential effects of living environment on quality of life. Kilic and Keklik (2012) studied the quality of life of health workers and investigated the effects of quality of life on job motivation. The quality of life indexes of 39 districts in Istanbul were computed by Seker (2011, 2015). Chen and Chen (2015) investigated the potential risks and effects of perceived living environments and neighborhood safety on mental health. Streimikiene (2015) deal with quality of life in terms of environmental quality and developed a system of indicators to evaluate it. Robustelli and Whisman examined the associations between gratitude and three domains of life satisfaction, including satisfaction in relationships, work, and health, and overall life satisfaction, in the United States and Japan. They suggest that gratitude is uniquely associated with specific domains of life satisfaction (Robustelli and Whisman 2018). Doré and Bolger (2018) used nonlinear multilevel models and proved that stressful life events like divorce, unemployment, and spousal loss are associated with decreased well-being. Garau and Pavan (2018) developed a method based on investigative checklists, objective and subjective indicators analyzed the urban quality of life in the city of Cagliari they presented an adaptable method for similar urban contexts. Kazemzadeh-Zow, et al. (2018) used multi criteria decision making (MCDM) method and fuzzy logic to develop a method based on satellite images and GIS for spatiotemporal modelling of Urban quality of life. Gibbons et al. (2018) used items from the World Health Organization Quality of Life Assessment (100-item version) and they calibrated model parameters of Quality of Life according to different cultures. Kaklauskas, et al. (2018a) studied environmental sustainability—Ecological Footprint (EF) and the Environmental Performance Index (EPI) —and Quality of Life Index in 15 republics of the former USSR, they strong correlations obtained between the EPI, EF and Quality of Life indicators on one side and the macroeconomic, values-based, human development and well-being factors. Kaklauskas et al. (2018b) applied multiple criteria analysis method to compare Quality of Life Index and the INVAR methods by using data from the 2012–2016 surveys on the Quality of Life in European Cities. Biagi et al. (2018) focused on accessibility to services, individual allocation of time, and the social interactions people enjoy parameters and their effect on residents' perception of quality of life in cities. Gavrilidis et al.

(2016) calculated Urban Landscape Quality Index via using visual assessment of the landscape and compared the results between the visual assessment of the urban landscape and survey targeting the landscape quality perception of the locals finally they reached 75% accuracy in their study.

3 Methodology

In the literature, the studies have calculated the quality of life by considering various factors with different methods. The results of these calculations are not used very effectively. We focus on a different working area of the quality of life that has not been discussed in the literature is emphasized. It was investigated whether the quality of life would be successful in prediction of the suitable living environment for individuals. The flow chart of the proposed method is shown in Fig. 1. As seen in Fig. 1, in the first stage, some desired features of the cities such as entertainment, social activity, social life, cultural life etc. are asked to individuals. Then quality of life indexes of the cities that they want to live were calculated. The calculated index and the current quality of life indexes of cities were compared. The people are assigned to the most suitable cities for themselves, taking into account their wishes and economic conditions. Then, individual and ensemble machine learning techniques were used to determine whether the previously calculated qualities of cities could be predicted by considering the responses of the individuals without these calculations. Therefore, a living environment prediction model is proposed using ML and EML techniques based on quality of life index. For predicting living environment, ML-based methodologies, including ANN, SVM, and EML-based methodologies include stacking and voting.

A section of the data set, which is used in the study, is given in Table 1. Proposed machine learning techniques use attribute values in other columns for predicting the index values as a dependent label in the last column of the data set. The output value in the last column is the quality of life classes of the cities that people want to live. Cities are divided into five classes in terms of quality of life. The best class value is 1 and the worst class value is 5. The index values corresponding to the classes are as follows: 5 (− 0, 6 ≤), 4 (− 0, 6 < and ≤ − 0, 2), 3 (− 0, 2 < and ≤ 0, 2), 2 (0, 2 < and <= 0, 6), 1 (0, 6 >).

3.1 Quality of life index

The quality of life has been studied with different point of views in the literature. The effects of psychology, physical health and economy on quality of life are not examined as

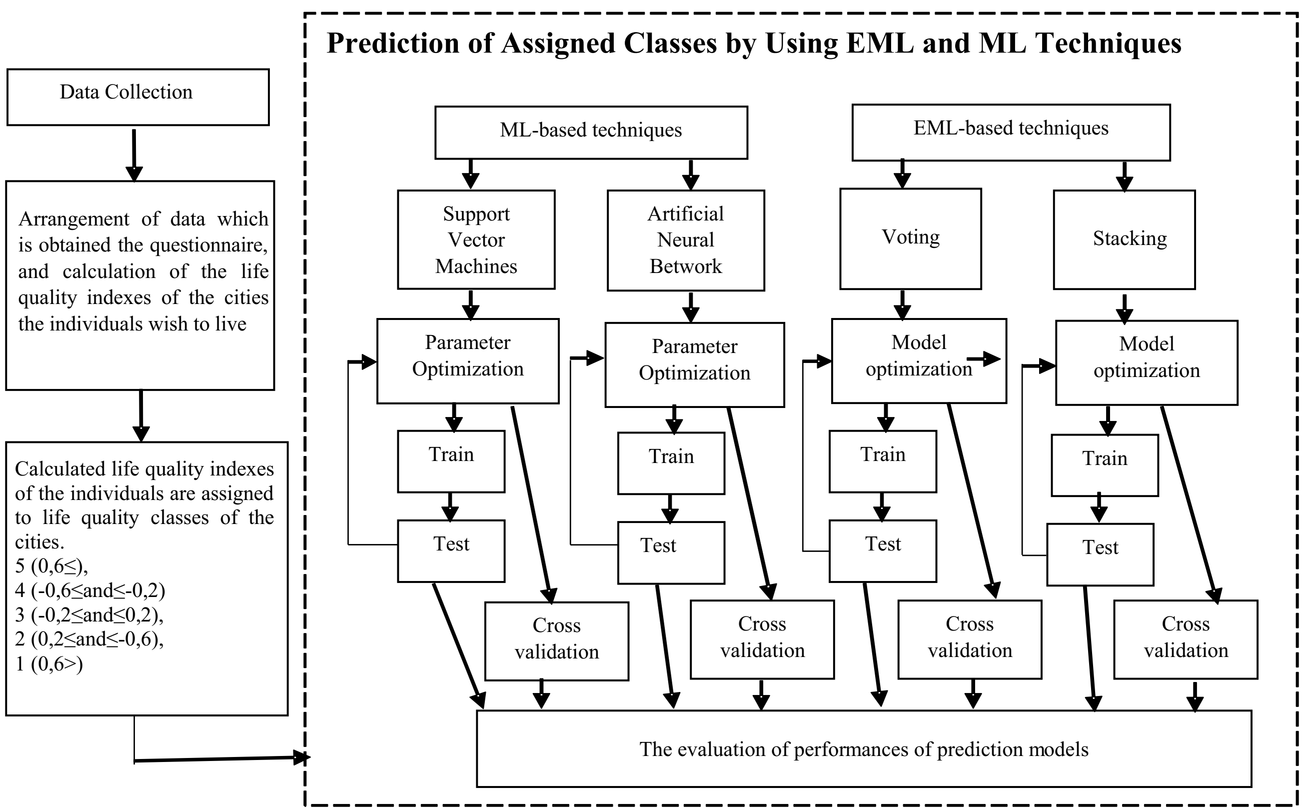

Fig. 1 Flow chart of model

Table 1 A section of the questionnaire study

Educational Status	Monthly household income	Number of people living in the house-hold	…	Boredom and gloom at work	Opportunities for professional development at work	Work health and safety conditions	Index values
High School	4001–6000	3–6	…	2	2	3	3
University and higher	4001–6000	0–3	…	3	5	3	2
University and higher	10001–12000	3–6	…	4	4	4	2
…	…	…	…	…	…	…	…
University and higher	4001–6000	0–3	…	5	5	5	2
University and higher	2001–4000	0–3	…	3	5	5	3
University and higher	6001–8000	3–6	…	1	3	3	4

a whole in the literature. In the studies, only one or two of each factor is considered. However, it is not enough to consider these factors separately to improve or measure of the quality of life. Having very good physical health psychology, health and economic situation is not enough to be happy. All of these factors need to be examined. Because these factors are components that contribute to the quality of life in different rates. Quality of life is a result of the reaction of these components. Therefore, a web based online survey study was conducted in order to determine the quality of life of the cities that people want to live in, by considering all these factors.

These questionnaire items were prepared under four headings as the main indicators (demographic structure, education, health, social life, and security), economic structure, environmental structure (infrastructure and environment, and transportation), and urban and personal life (urban life and personal satisfaction) (Seker et al. 2014). The questionnaire consists of sixty questions.

Data obtained from the questionnaire were transformed to a N (0, 1), which is the normal distribution of zero mean and 1 standard derivation by the following equation:

$$z = \frac{X - \mu}{\sigma} \tag{1}$$

where μ and σ are arithmetic mean and standard deviation, respectively.

Quality of life index was computed taking the weighted arithmetic mean of the transformed data. The Weights of the data were determined by experts (Seker et al. 2014).

$$\text{Index} = \frac{\sum iA_{ji}X_{ji}}{\sum iA_{ji}} \tag{2}$$

where A_{ji} is the weight of Ith subvariable of jth main variable and X_{ji} is transformed the value of ith subvariable of the jth main variable.

Calculated quality of life indexes of individuals are assigned to quality of life classes of the cities. Quality of life classes of the cities are 5 (− 0, 6 ≤), 4 (− 0, 6 < and ≤ − 0, 2), 3 (− 0, 2 < and ≤ 0, 2), 2 (0, 2 < and <=0, 6), 1 (0,6 >).

3.2 Machine learning techniques

The data used in the study is suitable for use of supervised learning methods due to the knowledge of class labels (Jiawei and Kamber, 2001; Unlu and Xanthopoulos 2019b). For this reason, classification algorithms, which are supervised learning methods, are preferred to set a prediction model. Individual classification algorithms deal with the various classification methods such as decision trees, artificial neural networks, statistic and distance based algorithms. All of the available methods were tested by trial and error method with using different parameter settings in order to find the algorithms that gives the best classification performance results from the individual classification algorithms. According to test results, classification algorithms that give the best performance on existing data were LMT, simple logistic, j48 and SVM methods.

When the results of the algorithms were examined, it was observed that individual algorithms were generally more successful for prediction of the each class label. Ensemble machine learning techniques are methods in which multiple learners solve the same problem that are trained to obtain a composite global model to obtain more reliable predictions rather than using a single model (Kargupta et al. 2008). For this reason, ensemble classification algorithms were used to complete the deficiency part of the individual algorithms in process of building prediction models. Voting and stacking algorithms are preferred to use individual machine learning techniques to be preferred in the class label in which they give more accurate results. Voting algorithm prefers the algorithm that has the most correct class label from an individual machine learning algorithm. On the other hand the stacking increases the weight of the algorithm that has the most accurate prediction results. The both methods are used in this study to tolerate the deficiencies of individual ML techniques. The methods that have the most accurate classification percentage value were preferred as subclassifier. The selected subclassifier methods are LMT, simple logistic, J48 and SVM. Different combinations of these methods were run as subclassifiers in the two ensemble machine learning techniques used in this study.

At this stage the use of ML techniques are tested for the estimation of the quality of life of the cities that individuals were assigned. The predictability of previously calculated city qualities by taking into consideration the answers of the people without these calculations were evaluated using individual and ensemble machine learning techniques. In this study, a living environment prediction model is proposed using ML and EML techniques for predicting quality of life index. Furthermore predicting living environment via ML-based methodologies, including ANN, SVM, and EML-based methodologies include stacking and voting.

3.2.1 Artificial neural network

Artificial neural network (ANN) is a powerful learning model for solving very complex problems inspired by the central nervous system of the human brain (Qiu et al. 2017). Multilayer perceptron (MLP) neural networks are standard neural network models whose structure mainly consist of one input layer, one or more hidden layers, and one output layer containing one computation node representing the living environment (Li et al. 2017). The hybrid models based on ANN are used for financial market behavior (Hassan et al. 2007), prediction problems (Hao et al. 2004; Sangwan et al. 2015; Suresh et al. 2011; Karimi and Yousefi 2012; Raikar et al. 2016), and optimization problems (Chou and Ngo 2016).

3.2.2 Support vector machines

Support vector machines (SVM) is the most robust and accurate methods among data mining algorithms (Wu and Kumar 2009). SVM which is a supervised learning algorithm for classification and regression analysis was first developed by Cortes and Vapnik (1995). SVM is one of the kernel methods which find applicable methods to process, analyze, and compare types of data (Shoombuatong et al. 2012). The hybrid models based on SVM used to create a prediction or a regression model (Tabari et al. 2012; Kisi et al. 2015).

3.2.3 Logistic model tree (LMT)

The logistic model tree (LMT) that combines decision tree learning methods and logistic regression (LR) is a classification model (Quinlan 1993). LogiBoost algorithm (Landwehr et al. 2005) produces an LR model at every node in the tree. The classification and regression tree (CART) algorithm is used to prune the tree (Breiman et al. 1984).

3.2.4 J48

J48 is an extension of the ID3 algorithm proposed by Ross Quinlan (1993). J48 offers additional features to overcome problems that ID3 was unable to deal with (Panda and Patra 2008).

J48 is an implementation of the C4.5 algorithm (Witten et al. 2011; Kaur and Chhabra 2014). Missing values are estimated using attribute values for the other records (Patil and Sherekar 2013). Each sample that represents attributes or features of the sample is a vector. J48 from decision trees is displayed for a high degree of accuracy and performance for a large amount of data (Kumar and Rathee 2011).

3.2.5 Simple logistic regression

The data that can be adapted to the LR are quite extensive (Cohen and Cohen 2008). The simple LR method has some advantages: their formulas are well known and do not require specialized software (Hsieh et al. 1998).

3.3 Ensemble machine learning techniques

3.3.1 Stacking

Stacked generalization or stacking, for short, is a hierarchical method of constructing multi-level classifiers (Wolpert 1992). Inputs for the second level of classifiers are outputs of base classifiers (Chou et al. 2014). The first level of stacking is learning how the base classifier makes errors and the outputs of the first level classifiers are used to train the second level classifier, then the second level classifier tries to overcome the errors (Chen et al. 2017). When bagging and boosting are used to combine models of the same type, stacking is used to construct models built by different learning algorithms (Witten et al. 2011).

By combining two different individual classifiers, stacking obtained three different ensemble classifiers. The first level of stacking is used for combining different base learning methods. In this study, an ensemble of two different classifiers was used: LMT + simple logistic, SMO + LMT, LMT + J48. LMT was used as the meta classifier.

3.3.2 Voting

Voting combining multiple classifiers is a simple and efficient method (Chen et al. 2017). The voting method pools the outputs of the individual classifiers in the cases of prediction and the final classification decision is the class that has the largest number of votes (Onan et al. 2016). Each algorithm runs the inducers, which are taken as input multiple times by changing the distribution of training set instances (Bauer and Kohavi 1999). Voting methods can be examined according to two types: boosting methods and bagging methods (Witten et al. 2011).

By combining two different individual classifiers, this study obtained three different ensemble classifiers. The ensemble of two different classifiers was used: SMO + simple logistic, SMO + LMT, SMO + simple logistic.

3.3.3 Cross validation

The cross validation method is preferred to minimize bias associated with the random training set and hold out observations (Wolpert 1992). If the best is algorithmic, cross validation can be used (Taniar 2008). The training set is selected as n − 1 out of n observations in cross validation. This process is applied in each of the training observations (Fernandez 2010; Ma et al. 2006). This study used a tenfold cross-validation.

In cross-validation method where the dataset N is randomly split k different subsets $(N, N_2, \ldots, N_k)$ such that:

$$\bigcup_{i=1}^{k} \left\{N, N, \ldots, N_k\right\} = N. \tag{3}$$

Each N_i is defined as test set and $N - N_i$. is used for the training set. Prediction result is calculated as the average of k runs.

$$R^* = \frac{\sum_{i=1}^{k} R_i}{k} \tag{4}$$

where $(R_i, i = 1, 2, \ldots, k)$ is the performance result in ith iteration.

K-folds cross-validation scheme is shown in Fig. 2 where k is chosen as 10.

3.4 Performance evaluation methods

$\sum SI$ ($\sum$Synthesisindex) was used as a comprehensive performance measure.

$$\sum \mathrm{SI} = (1 - \mathrm{ErrorSI}) + \mathrm{PerformanceSI} \tag{5}$$

where Error SI (E-SI) and performance SI (P-SI) are calculated by the following formula (Huang et al. 2016):

$$SI = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{P_i - P_{\min,i}}{P_{\max,i} - P_{\min,i}} \right) \tag{6}$$

where E-SI is calculated via four statistical measures; MAE, RMSE, RAE, and RRSE were derived. P-SI was calculated via twelve statistical measures; Accuracy, TPRate, FPRate, precision, recall, F-measure, MCC, ROC area, PRC area, G-mean1 and G-mean2 were derived by being inspired from (Chou et al. 2014). Where P_i is ith performance or error measure. The SI range was 0–1.

$$\mathrm{MAE} = \frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{n} \tag{7}$$

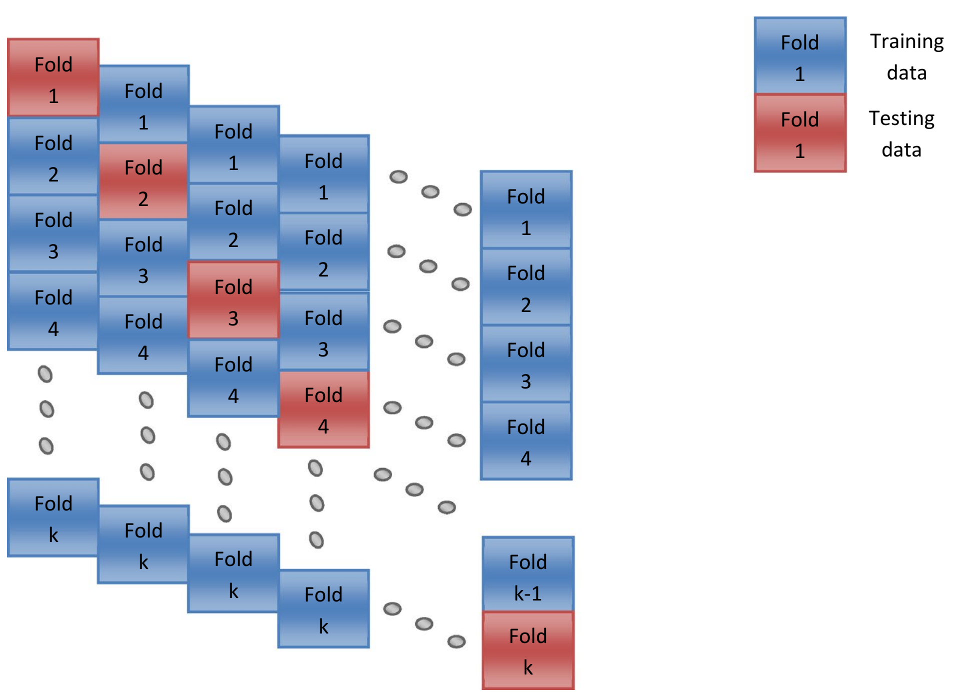

Fig. 2 K-fold cross validation method

where p is the actual value and a is the predict value.

$$\text{RMSE} = \sqrt{\frac{(p_1 - a_1)^2 + \cdots + (p_n - a_n)^2}{n}} \tag{8}$$

$$\text{RAE} = \frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{|a_1 - \bar{a}| + \cdots |a_n - \bar{a}|} \tag{9}$$

$$\text{RRSE} = \sqrt{\frac{(p_1 - a_1)^2 + \cdots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \cdots + (a_n - \bar{a})^2}} \tag{10}$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

$$F - \text{measure} = \frac{\left(1 + \beta^2\right) \times \text{recall} \times \text{precision}}{\beta^2 \times \text{recall} \times \text{precision}} \tag{14}$$

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}. \tag{15}$$

G-mean1 and G-mean2 performance measures are calculated by the following formula (Ma et al. 2006):

$$G - \text{mean1} = \sqrt{(\text{recall} \times \text{precision})} \tag{16}$$

$$G - \text{mean2} = \sqrt{(\text{recall} \times \text{TNR})}.$$

Table 2 Items related to main indicators

Maın indıcators	Variables	Values
Demographic structure	Monthly household income	1 (0–2000 TL), 2 (2001–4000 TL), 3 (4001–6000 TL), 4 (6001–8000 TL), 5 (8001–10,000 TL), 6 (10,001–12,000 TL), 7 (12,001–15,000 TL), 8 (15,001–20,000 TL), 9 (20,001 TL and higher)
	Number of people living in the household	
Education	Educational status	1 (illiterate), 2 (literate), 3 (primary school), 4 (middle school), 5 (high school), 6 (university and higher)
Health	Adequate number of health institutions	1(yes), 2(no)
	Cleanliness and hygiene	
	Treatment satisfaction	
	The healthcare staff is sufficiently interested in their patients	
	Treatment is obtained without too much waiting	
Social Life	Services for the cultural, artistic, and sports activities	The importance ratings are between 1 and 5
	Parks and playgrounds	
	Social services for the elderly	
	Social services for young people	
	Social services for women	
	Social services for poor and disabled	
	Traffic rules	
	Respect for other people in the public domain	
	Mutual respect in neighborhood relations	
	Pollution in the environment	
Security	The intervention speed and adequacy of police	
	A safe feeling while alone at home, the safety of the living environment when walking alone at night	
	The behavior of security officers against the citizens	
	Freedom from worry when the car is parked on the street (for safety reasons)	
	Lack of exposure to theft, snatching, harassment	

Table 3 Items related to economic capacity

Variables	Values
Economic capacity	
The heating/cooling the family home to comfortable levels	1 (yes), 2 (no)
Travelling without staying with relatives or acquaintances for once a week	
Replacing older furniture	
Eating meals consisting of beef, chicken, or fish every other day	
Welcoming friends or relatives into the home once a week	

4 Data description and modeling experiments

For the selection of the living environment what was taken into consideration were the factors that may be important for people in the cities where they live and whether there are economic opportunities that can provide these factors. A questionnaire combining demands and economic capacities was prepared by examining surveys used in the other studies of quality of life. This questionnaire polled nearly 200 people living on the European side of Istanbul.

The questionnaire items were prepared under four headings as the main indicators, economic structure, environmental structure, and urban and personal life (Seker et al. 2014). The questionnaire consists of sixty questions.

Quality of life index was computed taking the weighted arithmetic mean of the questionnaire data.

Weights of data were determined by experts (Seker et al. 2014). Weights of demographic structure, education, health, social life, security, economic structure, infrastructure and environment, transformation, urban life, and personal life were 10%, 10%, 10%, 5%, 5%, 30%, 8%, 7%, 7%, and 8%, respectively.

The questionnaire items are shown in Tables 2, 3, 4 and 5.

The quality of life of the cities where the people want to live is calculated using the data obtained from this questionnaire. Table 6 shows distribution of respondents according to their index class.

Parameter settings of proposed ML techniques are given in Table 7. Parameters of ML algorithms are optimized via grid search method to obtain best combination. Minimum and maximum prediction performance results are given in Table 8.

5 Results and discussion

The data used in the study were obtained from the questionnaire. In likert scale questions, nominal responses converted into numerical values. Five point alternatives having an inherent order 1, 2, 3, 4 and 5 represent, respectively very important, slightly important, moderately important, important and very important for importance Likert scale questions and for Satisfaction Likert scale questions again respectively very dissatisfied, dissatisfied, neutral, satisfied and very satisfied.

As seen in Table 9 and Fig. 3, data interpretation results with reference to Likert scale responses shows that as the satisfaction increases from health services, social state services, social norm quality, sense of security, expenditure-related qualification, satisfaction for municipal services and finally the general satisfaction degree life of quality index increases.

In addition, It is observed that people living with high income but living alone are in the class 3 and People living

Table 4 Items related to environmental structure

Environmental structure	Variables	Variables
The infrastructure and environment	Environmental waste collection service	The importance ratings are between 1 and 5
	Sewage service	
	Main water service	
	Road/sidewalk construction service	
	Green space ratio	
	The fight against the air pollution	
Transformation	Public transportation services	
	Traffic regulation	
	Parking services	

Table 5 Items related to urban and personal life

The urban and personal life	Variables	Variables
Urban life	Air quality and environmental pollution	The importance ratings are between 1 and 5
	Security	
	Educational institutions	
	Public institutions	
	Illumination of streets	
	Professional skills development and employment services	
Personal life	Education levels	The satisfaction ratings are between 1 and 5
	Current jobs	
	Family homes	
	Their countries	
	Family life	
	Health	
	Social life	
	Acquisitions	
	Personal development	
	Spiritual life	
	Job stress and worry	
	Boredom and gloom at work	
	Work health and safety conditions	
	Opportunities for professional development at work	

Table 6 Respondents' class based distribution

Index class	Sample value	Sample mix (%)
1	7	3.68
2	76	40.00
3	54	28.42
4	26	13.68
5	27	14.21
Total	190	100.00

Table 7 Parameter settings of ML models

Model	Parameter	Setting
MLP	Hidden layer	a
	Learning rate	0.4
	Momentum	0.3
	Training/time	500
	Validation threshold	20
SVM	C	1
	Kernel	Poly
Stacking	Classifiers	LMT + simple logistic, SMO + LMT, LMT + J48
	metaClassifier	LMT
Voting	Classifiers	simple logistic + LMT, SMO + LMT, SMO + simple logistic
	Combination Rule	Average of probabilities

in the crowd and have low income levels have poor quality of life.

Prediction performances of the applied ML techniques are shown in Table 2. A tenfold cross-validation method and 80%, 90% split methods were applied for each of the models. All prediction performance results of the model are shown in Table 10.

As shown in Table 4, Voting (simple logistic + LMT) performed the best over MAE and RAE using the tenfold cross-validation method. Stacking (SMO + LMT) performed the best over RMSE, RRSE, accuracy, kapa statistic, TPR, precision, recall, F-measure, MCC, G-mean1, and G-mean2 using an 80% split method. SMO performed the best over FPR using 90% split methods. Voting (simple logistic + LMT) performed the best over ROC area and PRC area using 80% split methods. The Voting method showed the best results by combining the Simple Logistic method and the LMT method. The Stacking method showed the best results by combining the SMO method and the LMT method. SMO performed the best over E-SI using a 90% split method. Stacking (SMO + LMT) performed the best over P-SI and $\sum$SI using an 80% split method. The multilayer perceptron method showed the best outcomes using the tenfold cross-validation method. SI values of the models illustrated in Fig. 4.

Table 8 The best and the worst values of prediction performances of models

	Min	Method	Max	method
MAE	0.1206	Vote (simple logistic + LMT) (FOLD 10)	0.2611	SMO (90% split)
RMSE	0.2402	Stacking (SMO + LMT) (80% split)	0.3531	Multilayer perceptron (90% split)
RAE	41.7687	Vote (simple logistic + LMT) (FOLD 10)	93.3847	SMO (90% split)
RRSE	64.9244	Stacking (SMO + LMT) (80% split)	96.5534	Multilayer perceptron (90% split)
Accuracy	0.42918	SMO (90% split)	0.76324	Stacking (SMO + LMT) (80% split)
Kappa	0.42918	SMO (90% split)	0.76324	Stacking (SMO + LMT) (80% split)
TPR	0.63158	SMO (90% split)	0.842105	Stacking (SMO + LMT) (80% split)
FPR	0.06601	Vote (SMO + LMT) (90% split)	0.182113	SMO (90% split)
Precision	0.6274	SMO (90% Split)	0.857202	Stacking (SMO + LMT) (80% split)
Recall	0.63158	SMO (90% split)	0.842105	Stacking (SMO + LMT) (80% split)
F-Measure	0.62180	Multilayer perceptron (90% split)	0.83797	Stacking (SMO + LMT) (80% split)
MCC	0.47646	SMO (90% split)	0.77556	Stacking (SMO + LMT) (80% split)
ROC Area	0.778	SMO (90% split)	0.952	Vote (simple logistic + LMT) (90% split)
PRC Area	0.581	SMO (90% split)	0.895	Vote (simple logistic + LMT) (90% split)
G-mean1	0.63210	SMO (90% split)	0.84962	Stacking (SMO + LMT)(%80 split)
G-mean2	0.71872	SMO (90% split)	0.8815	Stacking (SMO + LMT) (80% split)
E-SI	0.03983	Stacking (SMO + LMT) (80% split)	0.9716	SMO (90% split)
P-SI	0.18337	SMO (90% split)	0.93028	Stacking (SMO + LMT) (80% split)
$\sum$SI	0.21176	SMO (90% split)	1.89045	Stacking (SMO + LMT) (80% split)

Table 9 Average results of likert scale data via numerical representation

Labels	Average social state services	Average social norm quality	Average security	Average municipality services	Average private quality of life index
1	32.43	19.43	30.00	78.57	59.14
2	28.95	18.58	28.74	74.07	49.84
3	27.67	17.31	26.83	67.57	48.43
4	24.00	15.81	24.42	60.62	41.85
5	18.96	13.37	20.63	52.74	36.67

SMO performed the worst over MAE, RAE, accuracy, kappa statistic, TPR, precision, recall, MCC, Roc area, PRC area, G-mean1, G-mean2, E-SI, P-SI, $\sum$SI using a 90% split method. Multilayer perceptron performed the worst over RMSE, RRSE, and the F-measure, using the 90% split method.

Fig. 3 Trends of class based average values for likert scale data via numerical representation

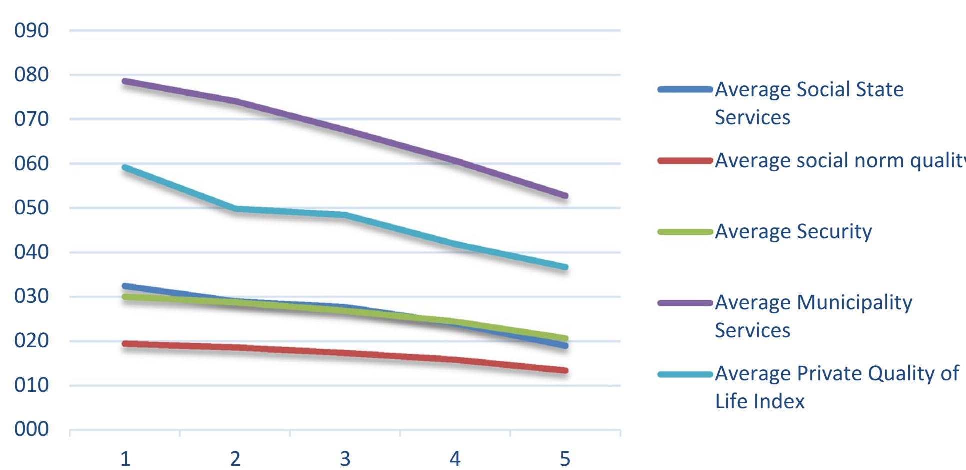

Table 10 Prediction performances of models

	Accuracy	Kappa	MAE	RMSE	RAE	RRSE	TPR	FPR	Precision	Recall	F-measure	MCC	ROC Area	PRC Area	G-mean1	G-mean2	E-SI	P-SI	$\sum$SI
Stacking (LMT + simple logistic) (FOLD 10)	75.79	0.66	0.14	0.28	50.16	73.71	0.76	0.09	0.73	0.76	0.74	0.67	0.90	0.70	0.74	0.83	0.24	0.58	1.34
Stacking (LMT + simple logistic) (80% split)	71.05	0.58	0.18	0.28	62.76	75.42	0.71	0.08	0.78	0.71	0.73	0.64	0.94	0.85	0.74	0.81	0.37	0.58	1.21
Stacking (LMT + simple logistic) (90% split)	78.95	0.67	0.14	0.27	49.41	71.84	0.79	0.11	0.82	0.79	0.78	0.71	0.91	0.82	0.81	0.84	0.17	0.76	1.59
Stacking (SMO + LMT) (FOLD 10)	75.26	0.65	0.14	0.27	49.02	70.96	0.75	0.10	0.73	0.75	0.73	0.65	0.92	0.76	0.74	0.82	0.18	0.61	1.43
Stacking (SMO + LMT) (80% split)	84.21	0.76	0.13	0.24	46.14	64.92	0.84	0.08	0.86	0.84	0.84	0.78	0.94	0.89	0.85	0.89	0.04	0.93	1.89
Stacking (SMO + LMT) (90% split)	78.95	0.67	0.14	0.30	48.92	81.35	0.79	0.11	0.82	0.79	0.78	0.71	0.86	0.71	0.81	0.84	0.32	0.70	1.38
Stacking (LMT + J48) (FOLD 10)	72.11	0.60	0.15	0.29	51.93	76.87	0.72	0.10	0.68	0.72	0.70	0.61	0.89	0.69	0.70	0.81	0.31	0.47	1.16
Stacking (LMT + J48) (80% split)	71.05	0.58	0.18	0.28	62.76	75.42	0.71	0.08	0.78	0.71	0.73	0.64	0.94	0.85	0.74	0.81	0.37	0.58	1.21
Stacking (LMT + J48) (90% split)	68.42	0.50	0.17	0.32	61.49	86.52	0.68	0.16	0.78	0.68	0.67	0.59	0.88	0.77	0.73	0.76	0.53	0.47	0.95
Vote (simple logistic + LMT) (FOLD 10)	74.21	0.64	0.12	0.29	41.77	77.55	0.74	0.09	0.73	0.74	0.73	0.65	0.91	0.75	0.73	0.82	0.22	0.58	1.36
Vote (simple logistic + LMT) (80% split)	65.79	0.50	0.14	0.31	48.03	82.43	0.66	0.11	0.74	0.66	0.68	0.58	0.93	0.84	0.70	0.77	0.34	0.43	1.09
Vote (simple logistic + LMT) (90% split)	73.68	0.61	0.15	0.27	52.25	73.44	0.74	0.07	0.85	0.74	0.76	0.72	0.95	0.89	0.79	0.83	0.23	0.69	1.47
Vote (SMO + LMT) (FOLD 10)	73.68	0.63	0.19	0.29	65.08	75.34	0.74	0.10	0.72	0.74	0.73	0.64	0.92	0.75	0.73	0.82	0.41	0.57	1.15
Vote (SMO + LMT) (80% split)	65.79	0.50	0.20	0.30	69.12	80.01	0.66	0.11	0.74	0.66	0.68	0.58	0.91	0.84	0.70	0.77	0.51	0.42	0.91
Vote (SMO + LMT) (90% split)	73.68	0.61	0.20	0.29	72.82	79.83	0.74	0.07	0.85	0.74	0.76	0.72	0.95	0.87	0.79	0.83	0.53	0.68	1.16
Vote (SMO + simple logistic) (FOLD 10)	74.21	0.64	0.19	0.28	65.16	78.00	0.74	0.09	0.73	0.74	0.73	0.65	0.92	0.76	0.73	0.82	0.43	0.58	1.15

Table 10 (continued)

	Accuracy	Kappa	MAE	RMSE	RAE	RRSE	TPR	FPR	Precision	Recall	F-measure	MCC	ROC Area	PRC Area	G-mean1	G-mean2	E-SI	P-SI	∑SI
Vote (SMO + simple logistic) (80% split)	65.79	0.50	0.20	0.30	69.12	80.01	0.66	0.11	0.74	0.66	0.68	0.58	0.91	0.84	0.70	0.77	0.58	0.42	0.91
Vote (SMO + simple logistic) (90% split)	73.68	0.61	0.20	0.29	72.82	79.83	0.74	0.07	0.85	0.74	0.76	0.72	0.95	0.87	0.79	0.83	0.53	0.68	1.16
SMO (FOLD 10)	68.42	0.56	0.26	0.34	88.47	89.45	0.68	0.12	0.67	0.68	0.68	0.56	0.87	0.62	0.68	0.78	0.86	0.37	0.49
SMO (80% split)	76.32	0.65	0.26	0.34	90.21	91.90	0.76	0.09	0.80	0.76	0.77	0.70	0.86	0.71	0.78	0.84	0.91	0.63	0.72
SMO (90% split)	63.16	0.43	0.26	0.35	93.38	95.06	0.63	0.18	0.66	0.63	0.62	0.48	0.78	0.58	0.65	0.72	0.97	0.18	0.21
Multilayer Perceptron (FOLD 10)	70.53	0.58	0.13	0.32	46.10	84.10	0.71	0.12	0.69	0.71	0.70	0.59	0.89	0.71	0.70	0.79	0.37	0.46	1.09
Multilayer perceptron (80% split)	68.42	0.54	0.13	0.32	46.45	86.65	0.68	0.10	0.76	0.68	0.71	0.60	0.91	0.82	0.72	0.78	0.39	0.49	1.09
Multi Layer perceptron (90% split)	63.16	0.45	0.15	0.35	52.58	96.55	0.63	0.11	0.81	0.63	0.62	0.58	0.94	0.86	0.71	0.75	0.60	0.41	0.81

The analysis presented in Table 3 shows that the outcomes of the ensemble models were better than the outcomes of the individual models. The stacking-based SMO + LMT, especially, had the best performance values using the 80% split method. Finally attribute selection process was performed for the selected model.

Attributes are ranked by their individual evaluations. Info gain method used for attribute evaluator this method evaluates the worth of an attribute by measuring the information gain with respect to the class.

As seen in Table 11, total 14 attributes do not have any contribution to the classification model. These redundant input attributes removed from the model. Then, same classification accuracy obtained by stacking method that is superior to other models. The redundant attributes that removed from the model did not have a positive or negative effect on classification performance. Moreover, there are many advantages of attribute selection process such as the number of questions increases in surveys, the number of respondents who participate the questionnaire decreases also less data means shorter training time and low computational complexity.

In this study, a multi-class classification problem has been addressed. Many classification algorithms have been tried to obtain accurate prediction results. The data set contains 60 input variables which means the dimension of the search space is very high when compared to several data sets. Also, it contains categorical and numerical values.

According to the characteristics of this data set combination of LMT and SVM algorithms via stacking method approached the most accurate classification results. LMT algorithm is a tree based algorithm so it can works efficiently even in little data preparation also can handle both numerical and categorical data finally Performs well when the number of features is big but they are not stable enough. Memory efficient SVM algorithm that works perfectly in high dimensional spaces increased the prediction accuracy by making the model more stable. Besides SVM reduces the overfitting by maximizing the margin between data points.

6 Conclusion

Proposed study aims to predict the living environments that have the best fit standards by which individuals can live using the measured quality of life of cities. The data required for the quality of life index were obtained from the questionnaire and the values of the quality of life index were computed taking the weighted arithmetic mean of these questionnaire data. The calculated indexes and the current quality of life indexes of cities were compared. The individuals are assigned to the most suitable cities for

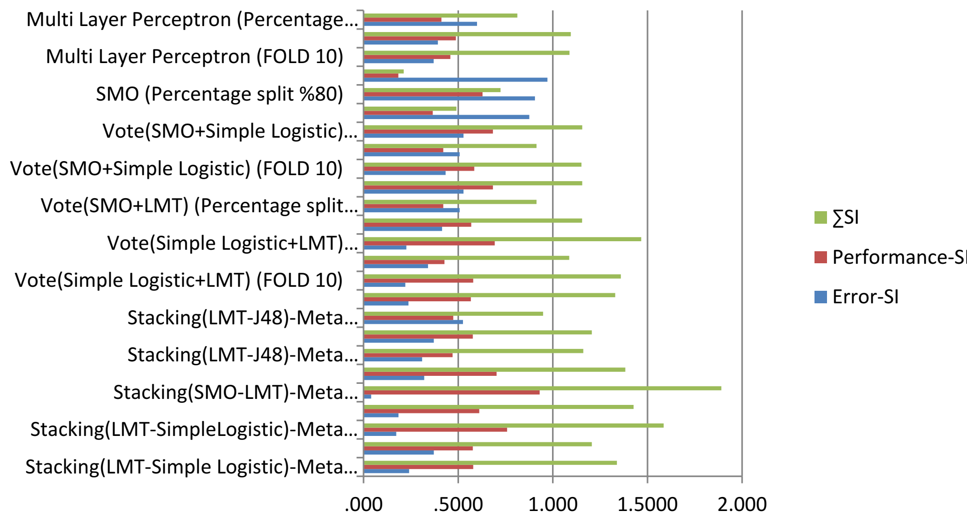

Fig. 4 SI values of prediction models

Table 11 Attribute ranking values

Sort order	Attribute	Rank value	Sort order	Attribute	Rank value	Sort order	Attribute	Rank value
1	41	0.4032	21	37	0.2166	41	48	0.1095
2	32	0.3665	22	33	0.2146	42	44	0.1068
3	38	0.3645	23	39	0.2121	43	49	0.1059
4	35	0.3252	24	25	0.2108	44	18	0.1058
5	42	0.309	25	43	0.2107	45	1	0.1049
6	3	0.2773	26	20	0.1952	46	50	0.0946
7	40	0.2712	27	15	0.185	47	4	0
8	19	0.2662	28	55	0.1835	48	59	0
9	21	0.2659	29	28	0.1799	49	5	0
10	36	0.2647	30	14	0.1783	50	58	0
11	23	0.2544	31	47	0.1722	51	7	0
12	24	0.2514	32	30	0.1674	52	2	0
13	27	0.2483	33	29	0.1657	53	6	0
14	17	0.242	34	45	0.1606	54	8	0
15	22	0.2418	35	26	0.1591	55	46	0
16	34	0.2364	36	53	0.1518	56	57	0
17	16	0.2334	37	10	0.1375	57	51	0
18	9	0.2325	38	13	0.1233	58	52	0
19	31	0.2262	39	11	0.1174	59	56	0
20	54	0.2212	40	12	0.1164	60	60	0

themselves, taking into account their wishes and economic conditions. Then, individual and ensemble machine learning techniques were used to determine whether the previously calculated qualities of cities could be predicted by considering the responses of the individuals without these calculations. Subsequently, this study performed a comprehensive comparison of various individual learning techniques and ensemble learning techniques. Two different ML techniques included SVMs and multilayer perceptron and two different EML techniques including stacking and voting were used to predict living environment. The parameters in Table 2 were used to develop these individual and ensemble models. The outcomes showed that the EML techniques have the best performance values in predicting living environment. Stacking-based SMO + LMT showed the best results using the 80% split method. SMO showed the worst results using the 90% split method. When the results of the method are examined, it is observed that machine learning techniques are suitable for estimating the optimal living environment for individuals. The proposed methodology has successful estimates as 70–80% rates.

This study emphasizes the importance of the quality of life index from which we will benefit regarding many issues, as well as the functional aspect that it can be used in the selection of living environments. Considering the application field of the methodology and proposed methodology, the study is the first model in the literature. In literature, it is seen that the studies have been conducted to measure the life qualities of cities. Therefore, these studies do not predict: instead, they only measure current situations. The study showed that the quality of life is not only a result of a measurement, but also an important parameter in assignments of living environments quality of life.

For future studies, enriching the living environment prediction models via adding relevant input attributes can give more accurate prediction results.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Abdullah ML, Tap AOM (2009) Quality of life index of three selected states in the Peninsular Malaysia: Fuzzy sets approach. Int J Ecol Econ Stat 11(S08):90–99

Abdullah SS, Malek MA, Abdullah NS, Kisi O, Yap KS (2015) Extreme learning machines: a new approach for prediction of reference evapotranspiration. J Hydrol 527:184–195

Aghdaie MH, Zolfani SH, Zavadskas EK (2014) Synergies of data mining and multiple attribute decision making. Procedia Soc Behav Sci 110:767–776

Anifowose FA, Labadin J, Abdulraheem A (2015) Ensemble model of non-linear feature selection-based extreme learning machine for improved natural gas reservoir characterization. J Natl Gas Sci Eng 26:1561–1572

Avci K, Pala K (2004) Uludağ Üniversitesi Tıp Fakültesinde çalışan araştırma görevlisi ve uzman doktorların yaşam kalitesinin değerlendirilmesi

Avramović JM, Veličković AV, Stamenković OS, Rajković KM, Milić PS, Veljković VB (2015) Optimization of sunflower oil ethanolysis catalyzed by calcium oxide: RSM versus ANN-GA. Energy Convers Manag 105:1149–1156

Baesens B, Viaene S, Van den Poel D, Vanthienen J, Dedene G (2002) Bayesian neural network learning for repeat purchase modelling in direct marketing. Eur J Oper Res 138(1):191–211

Bauer E, Kohavi R (1999) An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Mach Learn 36(1–2):105–139

Biagi B, Ladu MG, Meleddu M (2018) Urban quality of life and capabilities: an experimental study. Ecol Econ 150:137–152

Bose I, Chen X (2015) Detecting the migration of mobile service customers using fuzzy clustering. Inform Manag 52(2):227–238

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC Press, Boca Raton

Carmona M, de Magalhaes C (2007) Local environmental quality: a new view on measurement. Department for Communities and Local Government, London

Chen J, Chen S (2015) Mental health effects of perceived living environment and neighborhood safety in urbanizing China. Habitat Int 46:101–110

Chen XM, Zahiri M, Zhang S (2017) Understanding ridesplitting behavior of on-demand ride services: an ensemble learning approach. Transp Res Part C Emerg Technol 76:51–70

Chou JS, Ngo NT (2016) Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. Appl Energy 177:751–770

Chou JS, Pham AD (2014) Hybrid computational model for predicting bridge scour depth near piers and abutments. Autom Constr 48:88–96

Chou JS, Tsai CF, Pham AD, Lu YH (2014) Machine learning in concrete strength simulations: multi-nation data analytics. Constr Build Mater 73:771–780

Chou JS, Lin CW, Pham AD, Shao JY (2015) Optimized artificial intelligence models for predicting project award price. Autom Constr 54:106–115

Christensen BV (1996) Quality of life in the living environment of Slovenia. Druž Razpr 22–23:41–58

Cobaner M, Citakoglu H, Kisi O, Haktanir T (2014) Estimation of mean monthly air temperatures in Turkey. Comput Electron Agric 109:71–79

Cohen Y, Cohen JY (2008) Statistics and data with R: an applied approach through example. Wiley, Chichester

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Cummins RA, Eckersley R, Pallant J, Van Vugt J, Misajon R (2013) Developing a national index of subjective wellbeing: the Australian Unity Wellbeing Index. Soc Indic Res 64(2):159–190

Das D (2008) Urban quality of life: a case study of Guwahati. Soc Indic Res 88(2):297–310

Das D, Borbora S, Dutta MK (2012) Living environment and quality of life: a case study. Int J Ecol Econ Stat 25(2):71–84

Doré B, Bolger N (2018) Population-and individual-level changes in life satisfaction surrounding major life stressors. Soc Psychol Personal Sci 9(7):875–884

Ekinci A, Erdal H (2015) Optimizing the monthly crude oil price forecasting accuracy via bagging ensemble models. J Econ Int Finance 7(5):127–136

Endicott J, Nee J, Harrison W, Blumenthal R (1993) Quality of life enjoyment and satisfaction questionnaire: a new measure. Psychopharmacol Bull 29:321–326

Erdal H (2015) Makine öğrenmesi yöntemlerinin inşaat sektörüne katkısı: basınç dayanımı tahminlemesi. Pamukkale Üniv Mühendis Bilimleri Derg 21(3):109–114

Erdal H, Karahanoğlu İ (2016) Bagging ensemble models for bank profitability: an emprical research on Turkish development and investment banks. Appl Soft Comput 49:861–867

Erdal HI, Karakurt O, Namli E (2013) High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform. Eng Appl Artif Intell 26(4):1246–1254

Felce D, Perry J (1995) Quality of life: its definition and measurement. Res Dev Disabil 16(1):51–74

Fernandez G (2010) Statistical data mining using SAS applications, 2nd edn. CRC Press, Boca Raton

Garau C, Pavan VM (2018) Evaluating urban quality: indicators and assessment tools for smart sustainable cities. Sustainability 10(3):575

Gavrilidis AA, Ciocănea CM, Niţă MR, Onose DA, Năstase II (2016) Urban landscape quality index–planning tool for evaluating urban landscapes and improving the quality of life. Procedia Environ Sci 32:155–167

Gibbons CJ, Skevington SM, WHOQOL Group (2018) Adjusting for cross-cultural differences in computer-adaptive tests of quality of life. Quality of Life Research 27(4):1027–1039
Gupta R, Pathak CA (2014) Machine learning framework for predicting purchase by online customers based on dynamic pricing. Procedia Comput Sci 36:599–605
Hao Z, Qian X, Cen K, Jianren F (2004) Optimizing pulverized coal combustion performance based on ANN and GA. Fuel Process Technol 85(2):113–124
Hassan MR, Nath B, Kirley M (2007) A fusion model of HMM, ANN and GA for stock market forecasting. Expert Syst Appl 33(1):171–180
Heinermann J, Kramer O (2016) Machine learning ensembles for wind power prediction. Renew Energy 89:671–679
Hsieh FY, Bloch DA, Larsen MD (1998) A simple method of sample size calculation for linear and logistic regression. Stat Med 17(14):1623–1634
Huang S, Wang B, Qiu J, Yao J, Wang G, Yu G (2016) Parallel ensemble of online sequential extreme learning machine based on mapreduce. Neurocomputing 174:352–367
Jiawei H, Kamber M (2001) Data mining: concepts and techniques, 2nd edn. Morgan Kaufmann, San Francisco **(ISBN: 10: 1-55860-901-6)**
Kaklauskas A, Herrera-Viedma E, Echenique V, Zavadskas EK, Ubarte I, Mostert A, Podviezko A (2018a) Multiple criteria analysis of environmental sustainability and quality of life in post-Soviet states. Ecol Ind 89:781–807
Kaklauskas A, Zavadskas EK, Radzeviciene A, Ubarte I, Podviezko A, Podvezko V, Bucinskas V (2018b) Quality of city life multiple criteria analysis. Cities 72:82–93
Kargupta H, Han J, Philip SY, Motwani R, Kumar V (eds) (2008) Next generation of data mining. CRC Press, Boca Raton
Karimi H, Yousefi F (2012) Application of artificial neural network–genetic algorithm (ANN–GA) to correlation of density in nanofluids. Fluid Phase Equilib 336:79–83
Kartal H, Oztekin A, Gunasekaran A, Cebi F (2016) An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. Comput Ind Eng 101:599–613
Kaur G, Chhabra A (2014) Improved J48 classification algorithm for the prediction of diabetes. Int J Comput Appl 98(22):13–17. https://doi.org/10.5120/17314-7433
Kazemzadeh-Zow A, Darvishi Boloorani A, Samany NN, Toomanian A, Pourahmad A (2018) Spatiotemporal modelling of urban quality of life (UQoL) using satellite images and GIS. Int J Remote Sens 2018:1–22
Khademolqorani S, Hamadani AZ (2013) An adjusted decision support system through data mining and multiple criteria decision making. Procedia Soc Behav Sci 73:388–395
Kilic R, Keklik B (2012) Sağlık çalışanlarında iş yaşam kalitesi ve motivasyona etkisi üzerine bir araştırma
King MA, Abrahams AS, Ragsdale CT (2015) Ensemble learning methods for pay-per-click campaign management. Expert Syst Appl 42(10):4818–4829
Kisi O (2012) Modeling discharge-suspended sediment relationship using least square support vector machine. J Hydrol 456:110–120
Kisi O, Ozkan C, Akay B (2012) Modeling discharge–sediment relationship using neural networks with artificial bee colony algorithm. J Hydrol 428:94–103
Kisi O, Shiri J, Karimi S, Shamshirband S, Motamedi S, Petković D, Hashim R (2015) A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. Appl Math Comput 270:731–743
Kisi O, Genc O, Dinc S, Zounemat-Kermani M (2016) Daily pan evaporation modeling using Chi squared automatic interaction detector, neural networks, classification and regression tree. Comput Electron Agric 122:112–117
Kou G, Peng Y, Wang G (2014) Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Inf Sci 275:1–12
Kovalerchuk B, Vityaev E (2000) Data mining in finance: advances in relational and hybrid methods. Kluwer Academic, New York
Kumar V, Rathee N (2011) Knowledge discovery from database using an integration of clustering and classification. Int J Adv Comput Sci Appl 2(3):29–33
Landwehr N, Hall M, Frank E (2005) Logistic model trees. Mach Learn 59(1–2):161–205
Lazim MA, Osman MTA (2009) A new Malaysian quality of life index based on fuzzy sets and hierarchical needs. Soc Indic Res 94(3):499
Lewin K (1939) Field Theory and experiment in social psychology: concepts and methods. Am J Sociol 44(6):868–896
Li J, Song N, Gongliu Y, Li M, Qingzhong C (2017) Improving positioning accuracy of vehicular navigation system during GPS outages utilizing ensemble learning algorithm. Inform Fusion 35:1–10
Ma, Y, Guo, L, Cukic, B (2006) A statistical framework for the prediction of fault-proneness. Adv Mach Learn Appl Softw Eng pp 237–265
Maghrebi M, Waller T, Sammut C (2016) Matching experts' decisions in concrete delivery dispatching centers by ensemble learning algorithms: tactical level. Autom Constr 68:146–155
Michalos AC (ed) (2014) Encyclopedia of quality of life and well-being research. Springer Netherlands, Dordrecht
Namli E, Yucel M (2018) Yapay Zekâ Modelleri ile Betonarme Yapılara ait Enerji Performans Sınıflarının Tahmini. Uludağ Univ J Fac Eng 22(3):325–346
Namli E, Erdal HI, Erdal H (2016) Dalgacık dönüşümü ile beton basınç dayanım tahmininin iyileştirilmesi. Politek Derg 19(4):471–480
Namli E, Erdal H, Erdal HI (2019) Artificial intelligence-based prediction models for energy performance of residential buildings. Recycling and reuse approaches for better sustainability. Springer, Cham, pp 141–149
Onan A, Korukoğlu S, Bulut H (2016) A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Syst Appl 62:1–16
Panda M, Patra MR (2008) A comparative study of data mining algorithms for network intrusion detection. In: IEEE conference on emerging trends in engineering and technology (ICETET'08), pp 504–507
Panda M, Abraham A, Patra MR (2010) Discriminative multinomial naive bayes for network intrusion detection. In: IEEE conference on information assurance and security (IAS), pp 5–10
Park B, Bae JK (2015) Using machine learning algorithms for housing price prediction: the case of Fairfax County, Virginia housing data. Expert Syst Appl 42(6):2928–2934
Patil TR, Sherekar SS (2013) Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int J Comput Sci Appl 6(2):256–261
Peng Y, Zhang Y, Tang Y, Li S (2011) An incident information management framework based on data integration, data mining, and multi-criteria decision making. Decis Support Syst 51(2):316–327
Qiu X, Ren Y, Suganthan PN, Amaratunga GA (2017) Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. Appl Soft Comput 54:246–255
Quinlan J (1993) C4. 5: Programs for machine learning. C4. 5-programs for machine learning. Morgan Kaufmann Publishers, San Mateo

Raikar RV, Wang CY, Shih HP, Hong JH (2016) Prediction of contraction scour using ANN and GA. Flow Meas Instrum 50:26–34

Robustelli BL, Whisman MA (2018) Gratitude and life satisfaction in the United States and Japan. J Happiness Stud 19(1):41–55

Sangwan KS, Saxena S, Kant G (2015) Optimization of machining parameters to minimize surface roughness using integrated ANN-GA approach. Procedia CIRP 29:305–310

Seik FT (2000) Subjective assessment of urban quality of life in Singapore (1997–1998). Habitat Int 24(1):31–49

Seker M (2011) İstanbul'da yaşam kalitesi araştırması. İTO

Seker M (2015) Quality of life index: a case study of Istanbul. Ekonom İstat e-Derg 23:1–15

Seker M, SaldanIı A, Bektas H, Eyiler R (2014) TR63 bölgesi yaşam kalitesi analizi

Shoombuatong W, Hongjaisee S, Barin F, Chaijaruwanich J, Samleerat T (2012) HIV-1 CRF01_AE coreceptor usage prediction using kernel methods based logistic model trees. Comput Biol Med 42(9):885–889

Streimikiene D (2015) Environmental indicators for the assessment of quality of life. Intellect Econ 9(1):67–79

Suresh MVJJ, Reddy KS, Kolar AK (2011) ANN-GA based optimization of a high ash coal-fired supercritical power plant. Appl Energy 88(12):4867–4873

Tabari H, Kisi O, Ezani A, Talaee PH (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. J Hydrol 444:78–89

Taniar D (2008) Data mining and knowledge discovery technologies. IGI Global, Pennsylvania

Turkan YS, Aydogmus HY, Erdal H (2016) The prediction of the wind speed at different heights by machine learning methods. Int J Optim Control Theor Appl (IJOCTA) 6(2):179–187

Unlu R, Xanthopoulos P (2019a) A weighted framework for unsupervised ensemble learning based on internal quality measures. Ann Oper Res 276(1–2):229–247

Unlu R, Xanthopoulos P (2019b) Estimating the number of clusters in a dataset via consensus clustering. Expert Syst Appl 125:33–39

Walker A, Lowenstein A (2009) European perspectives on quality of life in old age. Springer, Berlin, pp 61–66

Whoqol Group (1995) The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med 41(10):1403–1409

Witten I, Frank E, Hall M (2011) Data mining: practical machine learning tools and techniques, 3rd edn. Morgan Kaufmann, Burlington

Wolpert DH (1992) Stacked generalization. Neural Netw 5(2):241–259

Wu X, Kumar V (2009) The top ten algorithms in data mining. CRC Press, Boca Raton

Xue X, Yao M, Wu Z, Yang J (2014) Genetic ensemble of extreme learning machine. Neurocomputing 129:175–184

Zhou L, Lai KK, Yu L (2010) Least squares support vector machines ensemble models for credit scoring. Expert Syst Appl 37(1):127–133

Zuo Y, Ali AS, Yada K (2014) Consumer purchasing behavior extraction using statistical learning theory. Procedia Comput Sci 35:1464–1473