International journal of applied exercise physiology
ISSN: 2322-3537
Vol-11 Issue-02 Aug 2022

Using machine Learning Approaches for Prediction of the Types of Asthmatic Allergy across the Turkey

Mr. Potharaju Srinivas
Assistant Professor, Department of CSE,
Malla Reddy College of Engineering for Women.,
Maisammaguda., Medchal., TS, India

Abstract— **Allergies are now considered a significant contributor to the rising incidence of illness in our modern culture. As a result, the primary goal of allergy researchers is to investigate the relationship between patient characteristics such as age, sex, and the kind of allergic illness, such as asthma, allergic rhinitis, food allergy, allergic dermatitis, and so on. In this study, we propose to use well-known machine learning algorithms such as Decision Tree, Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbor (kNN) and ensemble classifiers to design an intelligent diagnostic assistant for the automatic prediction of the type of allergic disease in Turkey. A dataset from the Kocaeli University Research and Application Hospital is used in studies. The highest possible accuracy rating of 77% in recognising 18 distinct allergy diagnoses is therefore reached by majority voting.**

Keywords: Allergy Classification Algorithms, Ensemble Classifiers, Machine Learning.

I. INTRODUCTION

When it comes to several fields, the amount of data is growing at an astronomical rate. It is becoming more difficult and crucial to analyse data automatically as the volume of data grows. One of the most important revolutions in healthcare, as well as many other fields, is the use of data analysis. Using the data to benefit patients and physicians is a significant step forward in the field of healthcare. Conventional techniques of processing and analysing medical data are ineffective because the data is too large and complicated. In order to employ these sets of data in decision-making, machine learning approaches give the means and technology. In particular, illness prediction, treatment efficacy evaluation, health service management, and customer relationship management are the key areas in which machine learning and its applications in health services are explored. However, the use of machine learning to analyse medical data presents a number of obstacles. It's important to engage with an expert to make sense of the data since there are a number of problems, such as: data stored in diverse sources, data in various formats, and heterogeneity of data (numerical or categorical values). In the medical field, data processing is very crucial.

A common use of machine learning techniques in medicine is to anticipate and investigate the origins of allergy illness. Allergies are among the most common and debilitating long-term conditions [2]. Asthma, rhinitis, eczema, dietary gastroenteritis, and colitis have all seen a surge in recent years as a result of pollution in the environment and urban living [3]. Asthma and asthma exacerbations are common in industrialised nations [4].

Allergenicity is predicted in a number of research in the literature. By analysing protein sequences, the vast majority of them arrive at their conclusions. According to the literature, SVM is the most widely utilised machine learning approach in these investigations. Furthermore, Zorzet et al. [6] used the kNN method to classify amino acid sequences for the prediction of food allergenicity. There are three distinct supervised algorithms utilised to predict allergenicity in Soeria-Atmadja and colleagues' [7] study: the logistic regression classifier, kNN, and a quadratic Gaussian linear classifier. Dimitrov et al. [8] created artificial neural network (ANN)-based algorithms for the prediction of allergenicity using protein sequences, and these algorithms were applied to 2427 known allergens and 2427 non-allergens, respectively. Predicting allergenic proteins was the goal of Dang and Lawrence.

There are just a few studies in the literature that use actual patient data to predict allergy diagnosis. Allergy symptoms in Taiwanese children may be predicted using neural networks, decision trees, and support vector machines. Ambrosia pollen concentration may be estimated using deep learning and ensemble learners, according to Zewdie et al. [11]. Several machine learning techniques were used

International journal of applied exercise physiology
ISSN: 2322-3537
Vol-11 Issue-02 Aug 2022

by Fontanella et al [12] to study the complicated link between particular immunoglobulin E and asthma. Lee et al. [13] suggest an alternative strategy that uses streaming Twitter data to forecast allergy levels. In order to identify the most prevalent forms of allergies, they employ text mining and machine learning.

Skin test data of 872 individuals from an allergy-testing centre were used to create a clinical decision support system (CDSS) by Christopher et al. [14]. CDSS was classified according to a set of rules.

Children and adults with allergies were the subject of our research, which looked at the prevalence of allergic disorders. To develop our app, we utilised a database of 28,031 patients' allergy diagnoses from throughout Turkey.

Research and Application Hospital of Kocaeli University provided the data for this dataset.

According to this structure, the paper will proceed. As a result of these investigations, algorithms and ensemble approaches for classification are discussed in Section 2. Setup and outcomes are detailed in Sections 3 and 4. Section 4 concludes the study with a discussion.

II. PREDICTION METHODS

In this section, the prediction methods, which are used in our experiments, are summarized.

A. Decision Tree

Deciduous trees divide training sets into subsets, each of which consists mostly of samples from a single class. An entropy-based metric known as information gain [15] is used to pick the characteristic that will best split samples into specific classes in the decision tree model partitioning procedure. In our studies, each attribute's information gain is computed using an entropy-based decision tree. Test attributes are selected based on the information gained from a certain training set. The quality of a tree is determined by the accuracy of its categorization and the size of the tree.

B. kNN

One other well-known machine learning approach that uses supervised, non-parametric classification is known as kNN. During the classification process, kNN calculates the Euclidian distances between an unknown data item and each of the training data objects [17]. It is therefore possible to compare the estimated equidistant distances of unknown data items to known data objects and assign them to the corresponding class.

C. SVM

Sustained machine learning approach SVM splits n-dimensional space with n-dimensional hyperplane into two areas in a manner that the hyperplane has the maximum distance from training vectors of two classes, termed support vectors [18]. By using kernel approaches, SVM may be used to classify data in a non-linear fashion. It's standard practise in machine learning to use kernel algorithms, which implicitly transform data into high-dimensional feature spaces that can be divided up linearly. For the SVM classifier, the employment of multiple kernel functions allows for a wide range of classifiers with distinct decision bounds. Due to its strong performance in terms of radial basis function (RBF) and polynomial kernel, linear kernel function is chosen in this article for SVM training.

D. Logistic Regression

Logistic regression is a common technique for classifying binary data. On the basis of this assumption, logistic regression is a kind of statistical analysis. An allergy diagnosis may be predicted if we suppose X is the input set of the independent variables (xi...,xk) that belong to patients and Y is the dependent variable we are attempting to predict by looking at X. An allergic diagnostic like "allergenic rhinitis" has a logistic distribution as a conditional probability provided by this assumption (1).

$$P(Y = allergenicrhinitis \mid X = x_i). \qquad (1)$$

The function in (1) is called as logistic regression function we need to predict Y.

E. Ensemble Classifiers

Ensemble models in machine learning bring together a number of ineffective learners to create a more effective one. To put it another way, ensemble models combine many alternative hypotheses to produce a more robust theory.

International journal of applied exercise physiology

ISSN: 2322-3537
Vol-11 Issue-02 Aug 2022

[19] Breiman and Cutler created Random Forest (RF) as a classification technique that employs an ensemble of decision tree learners. One of the most effective learning methods, ensemble literature has a very accurate classifier when used to several prediction models. RF bootstraps the training data to build each decision tree model. Using a random selection of characteristics, an impurity measure is utilised to divide data [20]. A entropy-based impurity measure is employed in all decision tree models in our tests.

Extremely Randomized Decision Tree (extra-trees) [21] is a decision tree-based ensemble model. Using randomness as a foundation allows us to build an extra-tree ensemble. After drawing each node of the tree, the best performing rule is assigned to that node depending on a predetermined threshold.

In ensemble learning, majority voting is the simplest and most extensively used method for combining several categorization predictions [22]. An unlabeled instance is classified with the class label that receives the most votes in a majority voting system.

III. EXPERIMENTAL SETUP AND RESULTS

A. Dataset

Real-world data from Turkey's Kocaeli University Research and Application Hospital is used for the experiments in this research. 28,021 patients were included in the dataset, which includes allergy diagnoses based on several factors. In the dataset's initial form, each patient record included eight distinct characteristics to consider. Patients are divided into 18 categories based on their characteristics, each of which represents a distinct allergy diagnosis. All of the characteristics are either numerical or categorical, and there is no missing data. Table I and Table II describe the dataset's characteristics and class labels, respectively.

Birthdate and ApplicationDate are maintained in date format whereas characteristics like NameHomeCity, Sex, Complaint, Diagnosis1 and 2 are nominal values.

TABLE I. THE FEATURES OF DATASET.

Feature	Value of Feature	Feature Characteristic
Birthdate	Birthdate of patients	Date
Name	Name of patients	Categorical
ApplicationDate	Day of Application	Date
HomeCity	Home city of patients	Categorical
Sex	Male, female	Categorical
Complaint	Patient's first complaint	Categorical
Diagnosis1	Initial diagnosis about patient	Categorical
Diagnosis2	Second stage diagnosis about patient	Categorical
Diagnosis3 (Class Label)	Definitive diagnosis about patient	Categorical

The dataset seems to be uneven when the number of samples in each class is taken into account. Some courses have much fewer training opportunities than others. There are less than ten training instances for the classes "J30.0" and "Z88.6," but there are 10612 training instances for the class "J30.2." One of the most important issues in machine learning is class imbalance. One of the key reasons for this is that classifiers tend to favour the majority classes when establishing the decision boundary. Minority-class cases are more likely to be misclassified as a result of this. This is confirmed by the findings of our study.

TABLE II. CLASS LABELS OF THE DATASET.

No	Class Labels (Turkish/English)	Number of Samples
D69.0	Alerji, tanımlanmamış (Allergy, Unspecified)	1140
L23.4	Alerjik kontakt dermatit, boyalara bağlı (Allergic contact dermatitis due to dyes, cosmetics)	16
L23.8	Alerjik kontakt dermatit, diğer ajanlara bağlı (Allergic contact dermatitis due to other agents)	124
L23.5	Alerjik kontakt dermatit, kimyasal ürünlere bağlı (Allergic contact dermatitis due to chemical products)	12
L23.0	Alerjik kontakt dermatit, metallere bağlı (Allergic contact dermatitis due to metals)	22
L23.9	Alerjik kontakt dermatit, tanımlanmamış nedenler (Allergic contact dermatitis, unspecified causes)	592
D69.0	Alerjik purpura (Allergic purpura)	16
J30.1	Alerjik rinit, polene bağlı (Allergic rhinitis due to pollen)	32
J30.4	Alerjik rinit, tanımlanmamış (Allergic Rhinitis, Unspecified)	7725
K52.2	Alerjik ve diyetetik gastroenterit ve kolit (Allergic and dietary gastroenteritis and colitis)	965
J45.0	Astım, alerjik (Asthma, allergic)	6658
J45.1	Astım, intrensek, alerjik olmayan (Asthma, intrinsic (non-allergic))	49
Z88.6	Kişisel ağrı kesici ajan alerjisi öyküsü (Personal history of pain relief allergy)	9
Z88.3	Kişisel diğer anti-enfektif ajanlara alerji öyküsü (Personal history of allergy to other anti-infective agents)	22
Z88.9	Kişisel tanımlanmamış uyuşturucu, ilaç ve biyolojik madde alerji öyküsü (Personal unspecified history of allergy to drugs, drugs and biological substances)	7
J30.2	Mevsimsel alerjik rinit, diğer (seasonal allergic rhinitis)	10613
T39.8	Nonopioid ağrı kesici, ateş düşürücü ve antiromatizmaller ile zehirlenme, tanımlanmamış	14
J30.0	Vazomotor rinit (Vasomotor rhinitis)	5

B. Pre-processing

The characteristics of patients are examined in depth for feature engineering, and it is discovered that all of them take numerical values. It is necessary to transform them numerically. The pre-processing procedures used in this investigation are outlined below.

Day is a new feature that stores the difference between a patient's day of birth and the date of their application. "Birthdate" and "ApplicationDate" are removed from the application. The "day" feature is then normalised using z-scores.

One hot encoding approach transforms the category feature "HomeCity" into a 1-dimensional numerical vector.

International journal of applied exercise physiology
ISSN: 2322-3537
Vol-11 Issue-02 Aug 2022

Male and female are represented numerically by the 0 and 1 numbers, respectively, for "sex."

For example, "J00-J99," "K00-K93," and "S00-S98" are all examples of the "Complaint" characteristic, which indicates a patient's initial complaint. The letters "J," "K," and "S" are used to represent these values.

For further information on the first and second stages of the diagnostic process, see the "Diagnosis1" and "Diagnosis2" features. Diagnoses 1 and 2 are kept with the special codes J30 and J30.0, respectively, since they are both classified as "upper respiratory disorders" for the patient.

C. Experimental Results

Pre-processed data is used to train machine learning algorithms, and it is later discovered that more tweaks to model parameters are needed to improve algorithm accuracy. Python's "sklearn" module is used to run these algorithms.

The impurity metric employed in decision tree implementation is entropy-based information gain. The model's accuracy, recall, and F-measure were all determined to be 0.63, 0.64, and 0.63 correspondingly once the tree had reached maturity. In addition, the train set had an accuracy of 0.89, whereas the validation set had an accuracy of 0.64. Decision tree model was found to be overfitting the training data when the computed train and validation set errors were compared. Pre-pruning, an early halting strategy in a decision tree model, was used to tackle this issue. As long as the decision node's partitioning would result in less than the set threshold, pruning takes place. Pre-processing in a decision tree is what we mean by this. The growth process may be halted early by pre-pruning. Pre-pruning is used in our trials, with the impurity reduction parameter set to a minimum of 0.1. If a split causes a drop in impurity greater than or equal to 0.01, a decision node will be split. An improvement in this adjustment has resulted in an increase in the accuracy of both the validation and training sets.

The number of neighbours is set to 5, and the distance is measured in Euclidian units for the kNN classifier.

Both underfitting and overfitting levels may be controlled using SVM penalty parameters. It has been found to be 10 in our testing. In SVM learning, the behaviour of the kernel technique is critical in finding the best hypothesis. Because the data points are not localised, the kernel approach is specified as a linear kernel.

Logistic regression classifiers use a penalty parameter, which is set at 100 by default. Additionally, the Newton-Conjugate-Gradient technique is used to optimise the logistic regression classifier.

Table III shows the accuracy, recall, and F-measure values for each method.

TABLE III. EVALUATION RESULTS.

Classifier	Precision	Recall	F-Measure
Decision tree	0.79	0.77	0.76
kNN	0.70	0.70	0.70
SVM	0.78	0.77	0.76
Logistic Regression	0.74	0.75	0.74
Random Forest	0.75	0.75	0.75
Extra-trees	0.77	0.76	0.76
Majority voting	0.79	0.77	0.77

Pre-processed data is used to train machine learning algorithms, and it is later discovered that more tweaks to model parameters are needed to improve algorithm accuracy. Python's "sklearn" module is used to run these algorithms.

The impurity metric employed in decision tree implementation is entropy-based information gain. The model's accuracy, recall, and F-measure were all determined to be 0.63, 0.64, and 0.63 correspondingly once the tree had reached maturity. In addition, the train set had an accuracy of 0.89, whereas the validation set had an accuracy of 0.64. Decision tree model was found to be overfitting the training data when the computed train and validation set errors were compared. Pre-pruning, an early halting strategy in a decision tree model, was used to tackle this issue. As long as the decision node's partitioning would result in less than the set threshold, pruning takes place. Pre-processing in a decision tree is what we mean by this. The growth process may be halted early by pre-pruning. Pre-pruning is used in our trials, with the impurity reduction parameter set to a minimum of 0.1. If a split causes a drop in impurity greater than or equal to 0.01, a decision node will be split. An improvement in this adjustment has resulted in an increase in the accuracy of both the validation and training sets.

ISSN: 2322-3537
Vol-11 Issue-02 Aug 2022

The number of neighbours is set to 5, and the distance is measured in Euclidian units for the kNN classifier.

Both underfitting and overfitting levels may be controlled using SVM penalty parameters. It has been found to be 10 in our testing. In SVM learning, the behaviour of the kernel technique is critical in finding the best hypothesis. Because the data points are not localised, the kernel approach is specified as a linear kernel.

Logistic regression classifiers use a penalty parameter, which is set at 100 by default. Additionally, the Newton-Conjugate-Gradient technique is used to optimise the logistic regression classifier.

Table III shows the accuracy, recall, and F-measure values for each method.

TABLE IV. CONFUSION MATRIX OF DECISION TREE CLASSIFIER.

	Precision	Recall	F1-Score	Support
D69.0	0.00	0.00	0.00	4
J30.0	0.00	0.00	0.00	1
J30.1	0.00	0.00	0.00	6
J30.2	0.82	0.54	0.65	2102
J30.4	0.57	0.84	0.68	1537
J45.0	0.99	1.00	1.00	1341
J45.1	0.00	0.00	0.00	10
K52.2	1.00	1.00	1.00	196
L23.0	0.00	0.00	0.00	6
L23.4	0.00	0.00	0.00	2
L23.5	0.00	0.00	0.00	1
L23.8	0.00	0.00	0.00	26
L23.9	0.74	1.00	0.85	131
T39.8	0.00	0.00	0.00	3
T78.4	0.99	1.00	0.99	230
Z88.3	0.00	0.00	0.00	4
Z88.6	0.00	0.00	0.00	2
Z88.9	0.00	0.00	0.00	2
micro avg	0.77	0.77	0.77	5604
macro avg	0.28	0.30	0.29	5604
weighted avg	0.79	0.77	0.76	5604

J45.0 (Allergic Asthma), K52.2 (Allergic and dietary gastroenteritis and colitis), and T78.4 (Unspecified Allergy) had a f1-score of about 1.0. J30.2, L23.9 and other classes are anticipated to have a f1-score of at least 0.65-2. There are between 592 and 10612 training instances for all of these courses when they are taken into account. Their numbers make them a majority. Minority classes may be shown in Table 2 by looking at the 0.0 f1-score classes.

TABLE V. CONFUSION MATRIX OF MAJORITY VOTING CLASSIFIER.

	Precision	Recall	F1-Score	Support
D69.0	1.00	1.00	1.00	4
J30.0	0.00	0.00	0.00	1
J30.1	0.00	0.00	0.00	6
J30.2	0.79	0.56	0.66	2102
J30.4	0.57	0.80	0.67	1537
J45.0	0.99	1.00	1.00	1341
J45.1	0.00	0.00	0.00	10
K52.2	1.00	1.00	1.00	196
L23.0	0.00	0.00	0.00	6
L23.4	0.00	0.00	0.00	2
L23.5	0.00	0.00	0.00	1
L23.8	0.00	0.00	0.00	26
L23.9	0.78	1.00	0.88	131
T39.8	1.00	1.00	1.00	3
T78.4	1.00	1.00	1.00	230
Z88.3	0.75	0.75	0.75	4
Z88.6	0.00	0.00	0.00	2
Z88.9	1.00	1.00	1.00	2
micro avg	0.77	0.77	0.77	5604
macro avg	0.49	0.51	0.50	5604
weighted avg	0.79	0.77	0.77	5604

Table V displays the majority voting classifier's confusion matrix. Ensemble model achieves a 1.0 f1-score for minorities such as D69.0, T39.8 and Z88.9, which is a small achievement. In the decision tree model, the accuracy of the majority classes is as great. Models like random forest, extra-trees, and SVM are also effective in predicting both majority and minority groups.

Our prediction accuracy is high despite the unbalanced data.

IV. CONCLUSIONS

An intelligent diagnostic helper for allergy patients was the goal of this work, which used a variety of classification algorithms and ensemble integration methodologies. Classification methods include decision trees, SVM, logistic regression, and kNN, as well as ensembles of these algorithms. Single classifiers such as decision tree and support vector machine (SVM) have the best accuracy, while the majority voting model has the best success rate of any technique.

In terms of its advantages, this study may assist researchers forecast and investigate the sort of allergy sickness more simply. In addition, research demonstrates that machine learning has become an indispensable tool for helping clinicians properly grasp their patients' needs. [pagebreak]

Data on aspects like air pollution, population density, and more will be included into future research. Additional research will focus on predicting the kinds of asthma allergies and variables that contribute to the condition in each city.

ISSN: 2322-3537
Vol-11 Issue-02 Aug 2022

REFERENCES

[1] C. Lagor, W. P. Lord, N. W. Chbat, J. D. Schaffer and T. Wendler, Advances in Healthcare Technology. Netherlands. Springer, 2006, ch. 22.

[2] R. Pawankari "Allergic diseases and asthma: a global public health concern and a call to action", World Allergy Organ J, vol. 7, pp. 1-3, May 2014.

[3] A. Mari, E. Scala, P. Palazzo, S. Ridolfi, D. Zennaro and G. Carabella, "Bioinformatics applied to allergy: Allergen databases, from collecting sequence information to data integration. The Allergome platform as a model", Cel. Immunol., vol. 244, no. 2, pp. 97-100, Dec. 2006.

[4] G. Devereux, "The increase in the prevalence of asthma and allergy: food for thought", Nat Rev Immunol., vol. 6, no. 11, pp. 869-874, Nov. 2006.

[5] K. Kadam, S. Sawant, V. K. Jayaraman and U. Kulkarni-Kale, Bioinformatics– Updated Features and Applications. London. UK: IntechOpen, 2016, ch. 4.

[6] A. Zorzet, M. Gustafsson and U. Hammerling, "Prediction of Food Protein Allergenicity: A Bio-informatic Learning Systems Approach", In Silico Biol., vol. 2, no. 4, pp. 525-534, 2002.

[7] D. Soeria-Atmadja, A. Zorzet, M. G. Gustafsson and U. Hammerling, "Statistical Evaluation of Local Alignment Features Predicting Allergenicity Using Supervised Classification Algorithms", Int Arch Allergy Immunol., vol. 133, no. 2, pp. 101-112, Feb. 2004.

[8] I. Dimitrov, L. Naneva, I. Bangov and I. Doytchinova, "Allergenicity prediction by artificial neural network", J Chemometr, vol. 28, no. 4, pp. 282-286, Jan. 2014.

[9] H. X. Dang, C. B. Lawrence, "Allerdictor: fast allergen prediction using text classification techniques", Bioinformatics, vol. 30, no. 8, pp. 1120-1128, Apr. 2014.

[10] H. F. Ng, H. M. Fathoni and I. C. Chen, "Prediction of allergy symptoms among children in Taiwan using data mining", 2009.

[11] G. K. Zewdie, D. J. Lary, E. Levetin and G. F. Garuma, "Applying Deep Neural Networks and Ensemble Machine Learning Methods to Forecast Airborne Ambrosia Pollen", Int J Environ Res Public Health., vol. 16, no. 11, Jun. 2019.

[12] S. Fontenella, C. Frainay, C. S. Murray, A. Smpson and A. Custovic, "Machine learning to identify pairwise interactions between specific IgE antibodies and their association with asthma: A crosssectional analysis within a population-based birth cohort", PLoS Med, vol. 15, no. 11, pp. 1-22, Nov. 2018. [Online]. Available: https://doi.org/10.1371/journal.pmed.1002691

[13] K. Lee, A. Agrawal and A. Choudhary, "Mining Social Media Streams to Improve Public Health Allergy Surveillance", in Proc. 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Paris, 2015, pp. 815–822.

[14] J. J. Christopher, H. K. Nehemiah and A. Kannan, "A clinical decision support system for diagnosis of Allergic Rhinitis based on intradermal skin tests", Comput Biol Med., vol. 65, pp. 76-84, Oct. 2015.

[15] J. R. Quinlan, "Induction of decision trees", Mach Learn., vol. 1, no. 1, pp. 81-106, Mar. 1986.

[16] M. S. Chen, J. Han and P. S. Yu, "Data Mining: An Overview from a Database Perspective", IEEE T Knowl Data En, vol. 8, no. 6, pp. 866-883, Dec. 1996.

[17] H. Zhuang, Y. Ni and S. Kokot, "Combining HPLC–DAD and ICP-MS data for improved analysis of complex samples: Classification of the root samples from Cortex", Chemometr Intell Lab., vol. 135, pp. 183-191, July 2014.

[18] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Min Knowl Disc., vol. 2, no. 2, pp. 121-167, June 1998.

[19] L. Breiman, "Random Forests", Mach Learn., vol. 45, no. 1, pp. 5-32, Oct. 2001.

[20] E. Ekinci, S. İlhan Omurca and N. Acun, "A Comparative Study on Machine Learning Techniques Using Titanic Dataset", in Proc. 7th International Conference on Advanced Technologies, Antalya, 2018, pp. 411-416.

[21] P. Geurts, D. Ernst and L. Wehenkel, "Extremely randomized trees", Mach Learn., vol. 63, no. 1, pp. 3-42, Apr. 2006.

[22] Z. H. Kilimci and S. Omurca, "Enhancement of the Heuristic Optimization Based Extended Space Forests with Classifier Ensembles", Int Arab J Inf Techn., vol. 17, no. 2, Mar. 2020.

[23] H. G. Nguyen, A. Bouzerdoum and S. L. Phung. Pattern Recognition, Vukovar, Croatia: In-The, 2009, pp. 193-208.

[24] Z. H. Kilimci and S. İlhan Omurca, "Extended Feature Spaces Based Classifier Ensembles for Sentiment Analysis of Short Texts", Inf Technol Control, vol. 47, no. 3, pp. 457-470, 2018.