



A NEW RATIONAL CLASSIFICATION APPROACH BY THE NEW MIXED DATA BINARIZATION METHOD

Keywords

*Binarization,
Classification,
Data Mining,
IRIS Data Set,
Decision Models,*

Abstract

Classification algorithm is a supervised learning technique that is used to identify the category of new observations. However, in some cases, quantitative and qualitative data must be used together. With this approach, we tried to overcome the problems encountered in using quantitative and qualitative data together. In this paper, we model a new classification technique by converting all types of data to binary data because in real world, data are classified in different types such as binary, numeric or categorical. By this way, we develop a more accurate and efficient mixed data binarization approach for multi-attribute data classification problems. First, we determine the classes from available dataset and then we classify the new instances into these predetermined classes by using the new proposed data binarization approach. We show how each step of this algorithm could be performed efficiently with a numeric example. Then, we apply the proposed approach on a well-known iris dataset and our model show promising results and improvements over previous approaches.

KARMA VERİ İKİLİLEŞTİRME YÖNTEMİ İLE YENİ BİR RASYONEL SINIFLANDIRMA YAKLAŞIMI

Anahtar Kelimeler

*İkilileştirme,
Sınıflandırma,
Veri Madenciliği,
IRIS Veri Seti,
Karar Modelleri.*

Öz

Sınıflandırma algoritması, yeni gözlemlerin kategorisini belirlemek için kullanılan denetimli bir öğrenme tekniğidir. Ancak bazı durumlarda nicel ve nitel verilerin birlikte kullanılması gerekir. Bu yaklaşımla nicel ve nitel verilerin birlikte kullanılmasında karşılaşılan sorunlar aşılmaya çalışılmıştır. Bu çalışmada, gerçek dünyada veriler ikili, sayısal veya kategorik gibi farklı türlerde sınıflandırıldığından, tüm veri türlerini ikili verilere dönüştürerek yeni bir sınıflandırma tekniği modellenmektedir. Bu sayede çok özellikli veri sınıflandırma problemleri için daha doğru ve verimli bir karma veri ikilileştirme yaklaşımı geliştirilmiştir. Öncelikle mevcut veri setinden sınıfları belirlenmektedir ve ardından yeni önerilen veri ikilileştirme yaklaşımını kullanarak yeni örnekleri bu önceden belirlenmiş sınıflara sınıflandırılmaktadır. Bu algoritmanın her adımının nasıl verimli bir şekilde gerçekleştirilebileceğini sayısal bir örnekle gösterilmiştir. Ardından, önerilen yaklaşımı iyi bilinen bir iris veri kümesine uygulamış ve modelimiz önceki yaklaşımlara göre umut verici sonuçlar ve iyileştirmeler verdiği gösterilmiştir.

Alıntı / Cite

Soyad1, A1., Soyad2, A2., Soyad3, A3., (2021). Türkçe Makaleler İçin Türkçe Başlık Olmalı ve Kelimelerin İlk Harfi Büyük Olmalıdır, Mühendislik Bilimleri ve Tasarım Dergisi, 9(1), ppp-ppp.

Yazar Kimliği / Author ID (ORCID Number)

A. Soyad, 0000-0000-0000-0001
A. Soyad, 0000-0000-0000-0002
A. Soyad, 0000-0000-0000-0003

Makale Süreci / Article Process

Başvuru Tarihi / Submission Date DD.MM.YYYY
Revizyon Tarihi / Revision Date DD.MM.YYYY
Kabul Tarihi / Accepted Date DD.MM.YYYY
Yayın Tarihi / Published Date DD.MM.YYYY

1. Introduction

Data Classification is one of the most important working areas in data mining. Data classification is a supervised learning strategy that categories the data in distinct classes. It deals to identify the patterns and classify the new samples into known classes (Silva and Zhao, 2012; Schwenker and Trentin 2014). Classification problems have been studied by different researchers including computer scientists, engineers, statisticians, biologist, and

economists (Jouni et al., 2014; Buniyamin et al., 2016; Loh, 2011; Graur et al., 2015). There are variety of methods for solving classification problems such as neural networks, K-nearest neighbor approach, support vector machines, linear programming, and fuzzy logic (Pratikakis et al., 2017; Cover and Hart, 1967; Zhang et al., 2018; Pal and Foody, 2010; Bai, 2020; Melin et al., 2013). Many data classification methods have been developed till now, but each of them has some shortcomings and difficulties which make them unattractive. Therefore, researchers are focusing on developing more accurate and efficient methods or trying to improve the existing methods. This leads different classification models to be applied in different fields in the literature including finance, risk management, health care, sports, engineering, and science (Zhang et al., 2013; Singhal et al., 2011; Faes, 2019; Russo et al., 2019; Waltman and van Eck, 2012). Also, data classification has a wide range of customer segmentation related applications as well. Characterization of customer segmentation into groups with similar behaviour and predict customer purchasing behavior such as buying a car, can be identified by classification (Sutcu, 2020).

Data classification is also very popular with the advances in technology which helps to increase capability of both generating and collecting data. This leads to a trend for data, its size and dimensionality grow. The widespread use of labeling techniques like barcodes, the computerization of businesses and advances in the data collection tools have provided us a huge amount of data. Millions of databases are now used by companies, governments, and universities (UCI, 2007; University of Toronto, 2003). It is noted that the number of these databases continue to grow rapidly because of the high-tech database systems. So, mining the unrefined data and convert it into useful information and knowledge is very important. However, there are a lot of different attributes related to stored data. Some of them have small some others have big impacts on decisions. Thus, splitting up the relevant and irrelevant data becomes an important issue.

Separation of relevant and irrelevant attributes becomes important because irrelevant attributes contain little or no information, for example, students' ID is often irrelevant to the task of predicting students' GPA. Also "redundant attributes" duplicate much, or all of the information contained in one or more other attributes, for example, purchase price of a product and the amount of sales tax paid. Moreover, we should create new attributes that can capture the important information on a data set which are much more efficiently than the original attributes.

A large number of data classification methods have been developed, but they have some obstacles and difficulties which make them unattractive (Carnevali and Miguel, 2008). Hence, researchers are focusing on developing more accurate and more efficient methods or improving the existing methods. Some of the previous methodologies have computational difficulties for large data sets, also these methodologies are time consuming and too expensive. Also, previous approaches are limited to either continuous or discrete case. Moreover, previous approaches are not enough to handle them, and they don't predict efficiently the classes if the dataset includes different variable types. In order to overcome these difficulties, we construct a new classification technique by converting all types of data to binary data because in real world, data are classified in different types such as binary, numeric or categorical. This issue leads us to develop a more accurate and efficient mixed data binarization approach for multi-attribute data classification problems.

The remainder of this paper is structured as follows: In section 2, we present methodology and theory of the new classification approach. Section 3 discusses the results of the application of the new classification approach to buying computer example. Section 4 contains concluding remarks.

2. Material and Method

In this section, we first discuss the new proposed approach and then explain and motivate the new approach with synthetic data. We then describe the matrix types used in our new method and we will give their mathematical formulations. We finally explain applicability of the proposed model with a motivating example.

2.1. New Data Classification Approach

In this subsection, we discuss the theory of the binarization approach.

2.1.1. Theory of the Binarization Approach

The data classification problem is considered in two parts as training part and testing part. Determination of the characteristics of the instances that belong to a certain class and differentiating them from the instances that belong to other classes are the main objectives of the training part. After the classes are determined, then by using the testing part, we can find the classes of the new instances.

Given the classes, the attributes and values of the attributes are sufficient to solve the classification problem with binarization approach. The all data set is represented by set Z. We divide the dataset Z into two groups; training data set which is represented by X, and test data set which is represented by Y. X_i where $i=1, \dots, n$ and Y_j where $j=1, \dots, m$ show the data in each set. There is totally $(n+m)$ amount of data in the set Z. Then, the attributes are shown by

$$\text{Values of instance } X_i = \{x_{i1}, x_{i2}, \dots, x_{ik}\} \quad (1)$$

$$\text{Values of instance } Y_j = \{y_{j1}, y_{j2}, \dots, y_{jk}\} \quad (2)$$

where k =number of attributes, x_{ik}, y_{jk} are the attribute values of k th attribute of i th and j th data. We have k attributes, and each attribute has at least one value. For instance, we have 3 attributes A, B, and C. A and B have 2 attribute values, C has 3 attribute values. So, we can represent them as $A = \{a_1, a_2\}$, $B = \{b_1, b_2\}$, and $C = \{c_1, c_2, c_3\}$.

For a given dataset, each data in dataset is written as binary values. So, “1” or “0” is used as an attribute value instead of using the real values. Assume instance-1 has the attribute values like $X_1 = \{a_1, b_2, c_3\}$ so, this attribute values are converted to a binary case by $A = [1, 0]$, $B = [0, 1]$, $C = [0, 0, 1]$ and then in the same logic, the binary vector is written as $X_1^b = [1, 0, 0, 1, 0, 0, 1]$ where X_1^b is the binary vector of instance-1.

2.1.2. Classification of Unclassified Instances

For determining the classes of unclassified instances, we use the metric distance approach. Generally, the aim is to minimize the distance between two classified instances, while maximize the distance between different classes. The most commonly used metrics to measure the distance of a sample from a given training set $X = \{x_1, x_2, \dots, x_n\}$ are as follows:

$$d(X, X^*) = \left(\sum_{i=1}^n |x_i - x_i^*|^r \right)^{\frac{1}{r}} \quad (3)$$

$$d(X, X^*) = \sqrt{\sum_{i=1}^n (x_i - x_i^*)^2} \quad (4)$$

$$d(X, X^*) = \left(\sum_{i=1}^n |x_i - x_i^*| \right) \quad (5)$$

$$d(X, X^*) = \max_i^n |x_i - x_i^*| \quad (6)$$

where X^* is the unclassified instance data set, x_i^* is the i th unclassified instance. Also, here “d” represents the distance measure.

In this study, the Euclidean Distance approach is used to measure the performance of the proposed model. The distance between unclassified instance and each class is calculated to find the nearest class for the instance. For each class, $d(X_j, X^*) = (\sum_{i=1}^n (x_{ji} - x_i^*)^2)$ is used as distance measure where X^* is the unclassified data, j is j th class and i is the i th value of the class j and the value of unclassified data. After explaining the previous distance methods, we now introduce our new measurement approach which is called multiplication approach by using a synthetic data set.

2.2. Example of the New Binarization Model with Synthetic Data

We use a synthetic dataset to explain and motivate the new approach. The data table is shown in Table 1. There are four attributes, of which X and Y are integer, Y is continuous, and Z is categorical. There are 5 classes in this example.

Table 1. Motivating Example Data Set

INITIAL MIXED DATA SET				
X	Y	Z	W	
3	0.342	A	3	X=Integer
5	0.51	C	120	Y=Continuous
2	0.104	E	141	Z= Categorical
1	0.779	B	6	W=Integer
4	0.628	D	130	

Firstly, attribute "W" is converted to nominal data by clustering with k-Nearest Neighbor algorithm. In this algorithm, the centroid of a cluster is selected as its center point. The centroid can be defined in various ways such as by the mean or medoid of the objects (or points) assigned to the cluster. The difference between an object $p \in C_i$ and c_i , the representative of the cluster, is measured by $dist(p, c_i)$, where $dist(x, y)$ is the Euclidean distance between two points x and y. The quality of cluster C_i can be measured by the within cluster variation, which is the sum of squared error between all objects in C_i and the centroid c_i , defined as

$$E = \sum_{i=1}^k \sum_{p \in C_i} dist(p, c_i)^2 \quad (7)$$

where E is the sum of the squared error for all objects in the data set; p is the point in space representing a given object; and c_i is the centroid of cluster C_i (both p and c_i are multi-dimensional). Here we have the values {3, 120, 141, 6, 130} for the attribute W. Intuitively, by visual inspection we may imagine the points partitioned into the two clusters, if equation-7 is applied and if we select k as 2, the partitioning {3, 6} and {120, 141, 130} has the within cluster variation as following:

$$(3 - 4.5)^2 + (6 - 4.5)^2 + (120 - 130.3)^2 + (130 - 130.3)^2 + (141 - 130.3)^2 = 238.37 \quad (8)$$

given that the mean of cluster {3, 6} is 4.5 and the mean of {120, 141, 130} is 130.3. Therefore, 238.37 is the lowest distance between classes when we partition the set into 2 clusters. Then, the set is converted into numeric values as $w_1 = \{3, 6\}$ and $w_2 = \{120, 141, 130\}$. So, the new dataset converts to nominal values by k-NN and now an updated data set with the attribute values w1 and w2 of attribute W is generated. Table-II shows the values of the updated version of the dataset.

Table 2. Motivating Example Converted Data Set

X	Y	Z	W
3	0,342	A	W ₁
5	0,51	C	W ₂
2	0,104	E	W ₂
1	0,779	B	W ₁
4	0,628	D	W ₂

The second step is converting the continuous attributes by splitting method as shown in Table-3. If the training dataset is big, we could assume that these values are distributed normally. So, we can find the mean and the variance of an attribute and then convert each value of an attribute to a probability value. Then we use the splitting method and partition the continuous valued attributes into 2, 3 or more branches. For instance, if we partition into two, then two branches are grown, corresponding to $A \leq split_{point}$ and $A > split_{point}$. Then, the data values are converted to binary values. (Assume $split_{point} = 0,50$)

Table 3. The Binary Data Set for New Approach

X					Y		Z					W	
X ₁	X ₂	X ₃	X ₄	X ₅	Y ₁	Y ₂	Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	W ₁	W ₂
0	0	0	1	0	1	0	1	0	0	0	0	1	0
0	0	1	0	0	0	1	0	0	0	1	0	0	1
1	0	0	0	0	1	0	0	1	0	0	0	0	1
0	0	0	0	1	0	1	0	0	0	0	0	1	0
0	1	0	0	0	0	1	0	0	1	0	0	0	1

The final step is the finding the most suitable class for the unclassified instance. For this, Euclidean Distance method is used to solve the minimization problem. The class is found by

$$\text{Minimum distance} = \min_j \left(d(X_j, X^*) = \left(\sum_{i=1}^n (x_{ji} - x_i^*)^2 \right) \right) \quad (9)$$

2.3. Class Matrix, Preference Matrix and Decision Matrix

In this section, we will describe the matrix types used in our new method and we will give their mathematical formulations.

Class matrix shows each class and their attribute values. It is a matrix where rows show the classes and columns show the attribute values.

$$\text{Class_Matrix } C_{(m \times n)} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} \quad (10)$$

where m=number of classes, n=total number of values of attributes.

Preference matrix shows the preferences of decision maker as a column vector.

$$Preference_Matrix P_{(n \times 1)} = \begin{pmatrix} p_{11} \\ p_{21} \\ \dots \\ p_{n1} \end{pmatrix} \text{ where } n = \left(\sum_{i=1}^x v_i \right) \quad (11)$$

where v_i is the number of values of an attribute.

Finally, the decision matrix is the multiplication of the class matrix and preference matrix which shows the results of the distances between each class and unclassified distance. We can easily read the distances from the decision matrix. We can find the most suitable class by reading the maximum value of each value of the matrix.

$$DecisionMatrix D_{(m \times 1)} = C_{(m \times n)} \times P_{(n \times 1)} = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \dots & c_{mn} \end{pmatrix} \times \begin{pmatrix} p_{11} \\ p_{21} \\ \dots \\ p_{n1} \end{pmatrix} = \begin{pmatrix} d_{11} \\ d_{21} \\ \dots \\ d_{n1} \end{pmatrix} \quad (12)$$

We then look at the best class for a customer's preferences by $c^* = \text{argmax}_i = (d_{i1})$ where c^* shows the best suitable class.

2.4. The Motivating Example of the New Classification Approach

The motivating example includes 4 attributes where attributes "age" and "income" have three values, attributes "student" and "credit" have 2 values. So, for this motivating example, there are different classes however handling with these all classes is difficult, expensive and time consuming. Because this is a small example, but it has 72 different classes. As handling with a bigger problem with so many attributes and values, the number of classes and difficulty of the problem increases exponentially.

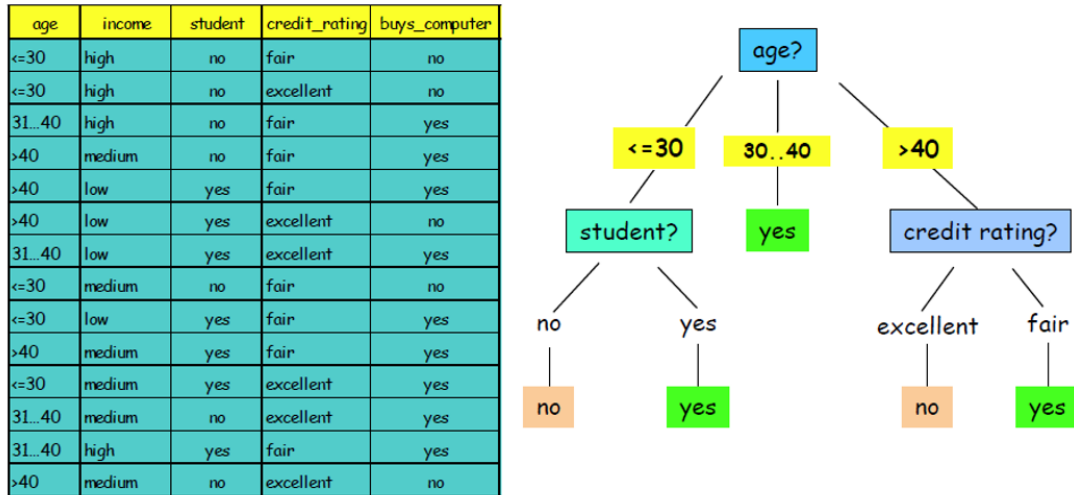


Figure 1. Classification of the Motivation Example by Decision Tree Induction

Due to the difficulty of the problem, we prune the number of classes to fourteen instead of seventy-two. The attributes are given as

$$Age = \{\text{under 30}, 31 - 40, \text{above 40}\} \quad (13)$$

$$Income = \{low, medium, high\}$$

$$Student = \{no, yes\}$$

$$Credit_rating = \{fair, excellent\}$$

Next step, we convert the values of each class into binary values. Therefore, the new table is shown in Table 4.

Table 4. The Binary Data Set for Buying Computer Case

	AGE			INCOME			STUDENT		CREDIT		OUTCOME	
	Age-T	Age-M	Age-O	Income-L	Income-M	Income-H	Student-N	Student-Y	Credit-F	Credit-E	Buy	Not buy
Class 1	1	0	0	0	0	1	1	0	1	0	0	1
Class 2	1	0	0	0	0	1	1	0	0	1	0	1
Class 3	0	1	0	0	0	1	1	0	1	0	1	0
Class 4	0	0	1	0	1	0	1	0	1	0	1	0
Class 5	0	0	1	1	0	0	0	1	1	0	1	0
Class 6	0	0	1	1	0	0	0	1	0	1	0	1
Class 7	0	1	0	1	0	0	0	1	0	1	1	0
Class 8	1	0	0	0	1	0	1	0	1	0	0	1
Class 9	1	0	0	1	0	0	0	1	1	0	1	0
Class 10	0	0	1	0	1	0	0	1	1	0	1	0
Class 11	1	0	0	0	1	0	0	1	0	1	1	0
Class 12	0	1	0	0	1	0	1	0	0	1	1	0
Class 13	0	0	0	0	0	1	0	1	1	0	1	0
Class 14	0	0	1	0	1	0	1	0	0	1	0	1

Now, our model is ready for the unclassified instances. For instance, one of the unclassified instances is given a teen-age, low-income student with a fair credit score. So, it can be written $X = \{teen, low, yes, fair\}$. Then we convert it to binary values by

Customer Preferences	1	0	0	1	0	0	0	1	1	0
	Age			Income			Student		Credit	

Finally, we calculate the distances between the instance and 14 classes one by one. The results are shown in Table 5. Also, the outcomes of each class are shown in Table 6.

Table 5. Euclidean Distances of Unclassified Instance

Class 1	6	Class 8	6
Class 2	8	Class 9	2
Class 3	4	Class 10	4
Class 4	6	Class 11	6
Class 5	2	Class 12	6
Class 6	4	Class 13	3
Class 7	2	Class 14	8

Table 6. Outcomes of each Class

Class 1	not buy	Class 8	not buy
Class 2	not buy	Class 9	Buy
Class 3	Buy	Class 10	Buy
Class 4	Buy	Class 11	Buy
Class 5	Buy	Class 12	Buy
Class 6	not buy	Class 13	Buy
Class 7	Buy	Class 14	not buy

We can see from the example, without spending too much time and money, we can easily find the class of the unclassified instance. Here, we can see that the instance would be in class-5, class-7 or class-9. In all the options, we can say that the customer is going to buy a computer. Also, if we compare our findings with the test data (the unclassified instance was taken from the test data set), the customer's behavior is to buy a computer. Also, we try several options and for each case our findings are 100% correct.

3. An Application on IRIS Flower Dataset

The IRIS data set were used by Fisher in his development of the linear discriminant function and is still one of the standard discriminant analysis examples used in explaining or testing most current approaches and methodologies (Fisher, 1936). In this problem, three classes of IRIS flowers are to be discriminated using four continuous valued features that represent physical characteristics of the flowers. The data set consists of 150 cases, 50 for each class. The four attributes are Sepal Length, Sepal Width, Petal Length, and Petal Width.

The attributed that already been predicted belongs to the class of IRIS plant. The list of attributes present in the IRIS can be described as categorical, nominal, and continuous. The experts have mentioned that there isn't any missing value found in any attribute of this data set. The data set is complete. This project makes use of the well-known IRIS dataset, which refers to three classes of 50 instances each, where each class refers to a type of IRIS plant. The first of the classes is linearly distinguishable from the remaining two, with the second two not being linearly separable from each other. The 150 instances, which are equally separated between the three classes, contain the following four numeric attributes:

1. sepal length – continuous
2. sepal width – continuous
3. petal length – continuous
4. petal width – continuous

and the fifth attribute is the predictive attributes which is the class attribute that means each instance also includes an identifying class name, each of which is one of the following: IRIS Setosa, IRIS Versicolour, or IRIS Virginica the IRIS dataset (downloaded from the UCI repository, www.ics.uci.edu, which is a 150×4 matrix, is taken as the input data) (UCI, 1988).

In the analysis part, we first convert all the attributes to binary attributes as Table 6 below.

Table 6. Binary Attributes of IRIS Dataset

	Att-1	Att-2	Att-3	Att-4	Class	Attribute-1			Attribute-2			Attribute-3			Attribute-4			
						V-1	V-2	V-3	V-1	V-2	V-3	V-1	V-2	V-3	V-1	V-2	V-3	
1	5.1	3.5	1.4	0.2	Iris-setosa	1	0	0	0	0	1	1	0	0	1	0	0	Setosa
2	4.9	3	1.4	0.2	Iris-setosa	1	0	0	0	1	0	1	0	0	1	0	0	Setosa
3	4.7	3.2	1.3	0.2	Iris-setosa	1	0	0	0	0	1	1	0	0	1	0	0	Setosa
4	4.6	3.1	1.5	0.2	Iris-setosa	1	0	0	0	0	1	1	0	0	1	0	0	Setosa
5	5	3.6	1.4	0.2	Iris-setosa	1	0	0	0	0	1	1	0	0	1	0	0	Setosa
6	5.4	3.9	1.7	0.4	Iris-setosa	0	1	0	0	0	1	0	1	0	0	1	0	Versicolor
54	5.5	2.3	4	1.3	Iris-versicolor	0	1	0	1	0	0	0	1	0	0	1	0	Versicolor
55	6.5	2.8	4.6	1.5	Iris-versicolor	0	0	1	1	0	0	0	0	1	0	0	1	virginia
56	5.7	2.8	4.5	1.3	Iris-versicolor	0	1	0	1	0	0	0	0	1	0	1	0	Versicolor
104	6.3	2.9	5.6	1.8	Iris-virginica	0	0	1	0	1	0	0	0	1	0	0	1	virginia
105	6.5	3	5.8	2.2	Iris-virginica	0	0	1	0	1	0	0	0	1	0	0	1	virginia
106	7.6	3	6.6	2.1	Iris-virginica	0	0	1	0	1	0	0	0	1	0	0	1	virginia

As binarization is the process of transforming data features of any entity into vectors of binary numbers to make classifier algorithms more efficient, we now transform all the attributes of the dataset into binary vectors to represent all the data in the dataset as binary attributes. After converting all the data to suitable binary vectors based on each attribute, we run the proposed model in order to measure the performance of it and compare our model with existing classification models, SVM, k-NN and decision trees using IRIS dataset. The performance results are shown on Table 7.

Table 7. Comparison of Classification Models with Proposed Approach

Classification	Accuracy Rate	# of Correctly	# of Incorrectly
Method	(%)	Classified	Classified
k-Nearest Neighbors	82%	123	27
Decision Tree	79%	118	32
Support Vector Machine	86%	129	21
<i>Binarization Method</i>	86%	129	21

As can be seen in Table-7, the proposed model we defined performed well on IRIS dataset. In general, our model gives a high accuracy rate of 86% where k-NN is around 82%, and decision trees gives an accuracy rate of 79%. The results show that our model gave the same results as SVM and gave better performance than k-NN and decision tree models.

4. Discussion and Conclusion

This paper presents a novel binary classification approach for multi-class mixed data classification. A binarization algorithm is developed for finding the classes of the unclassified instances. In order to overcome the difficulties for large data sets, we convert the data set to a binary value matrix. The performance of the model is tested by applying the instances in the testing data set.

The new model can be used for all kinds of data including integer, discrete, continuous, binary, or categorical. High classification accuracies are observed the application of different cases. From the computational time and cost of the model perspective, the proposed model is acceptable and applicable to obtain solutions for large mixed data classification problems.

Furthermore, the testing algorithm is computationally tractable for high dimensional data sets. As observed from the examined data sets, total computational time for proposed approach is reasonable. The proposed approach in our paper gives high accuracy values on the different data sets. Hence, the developed multi-class mixed data classification model is at least as accurate as the other models including NN, SVM, Decision Trees, K-Nearest Neighbor, Logistic Regression, Bayesian Classifier, etc.

As a conclusion, by the development of this new approach, solutions to multi-class mixed data classification problems can be obtained and the prediction accuracies can be improved. In addition to this, the simplicity and the understandability of the proposed model are preferable.

Acknowledgement

The authors would like to thank the anonymous reviewers and editor for their comments and suggestions.

Conflict of Interest

No conflict of interest was declared by the authors.

References

- Bai, Jing, Anran Yuan, Zhu Xiao, Huaji Zhou, Dingchen Wang, Hongbo Jiang, and Licheng Jiao. 2020. "Class Incremental Learning With Few-Shots Based on Linear Programming for Hyperspectral Image Classification." *IEEE Transactions on Cybernetics*.
- Buniamin, Norlida, Usamah bin Mat, and Pauziah Mohd Arshad. 2016. "Educational Data Mining for Prediction and Classification of Engineering Students Achievement." In *2015 IEEE 7th International Conference on Engineering Education, ICEED 2015*, 49–53. Institute of Electrical and Electronics Engineers Inc.
- Carnevali, Jose A., and Paulo Cauchick Miguel. 2008. "Review, Analysis and Classification of the Literature on QFD-Types of Research, Difficulties and Benefits." *International Journal of Production Economics* 114 (2): 737–54.
- Cover, T. M., and P. E. Hart. 1967. "Nearest Neighbor Pattern Classification." *IEEE Transactions on Information Theory* 13 (1): 21–27.
- Faes, L, M K Schmid, S K Wagner Bmbch, Liu Mbchb, R Chopra Bsc, N Pontikos, Sim Phd, et al. 2019. "Automated Deep Learning Design for Medical Image Classification by Health-Care Professionals with No Coding Experience: A Feasibility Study." *Articles Lancet Digital Health* 1: 232–74.
- Fisher. 1988. "Iris Data Set." UCI Center for Machine Learning and Intelligent Systems. July 1, 1988. <https://archive.ics.uci.edu/ml/datasets/Iris>.
- FISHER, R. A. 1936. "THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS." *The Annals of Human Genetics*, September, 179–88.
- Graur, Dan, Yichen Zheng, and Ricardo B.R. Azevedo. 2015. "An Evolutionary Classification of Genomic Function." *Genome Biology and Evolution* 7 (3): 642–45. <https://doi.org/10.1093/gbe/evv021>.
- Jouini, Mouna, Latifa Ben Arfa Rabai, and Anis ben Aissa. 2014. "Classification of Security Threats in Information Systems." In *Procedia Computer Science*, 32:489–96. Elsevier B.V.
- Loh, Wei Yin. 2011. "Classification and Regression Trees." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (1): 14–23.
- Melin, Patricia, Frumen Olivas, Oscar Castillo, Fevrier Valdez, Jose Soria, and Mario Valdez. 2013. "Optimal Design of Fuzzy Classification Systems Using PSO with Dynamic Parameter Adaptation through Fuzzy Logic." *Expert Systems with Applications* 40 (8): 3196–3206.
- Pal, Mahesh, and Giles M. Foody. 2010. "Feature Selection for Classification of Hyperspectral Data by SVM." *IEEE Transactions on Geoscience and Remote Sensing* 48 (5): 2297–2307.
- Pratikakis, I, F Dupont, and M Ovsjanikov. 2017. "Exploiting the PANORAMA Representation for Convolutional Neural Network Classification and Retrieval." *Eurographics Workshop on 3D Object Retrieval*.
- Russo, Mohammad Ashraf, Laksono Kurnianggoro, and Kang-Hyun Jo. 2019. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE.
- Schwenker, Friedhelm, and Edmondo Trentin. 2014. "Pattern Classification and Clustering: A Review of Partially Supervised Learning Approaches." *Pattern Recognition Letters* 37 (1): 4–14.
- Silva, Thiago Christiano, and Liang Zhao. 2012. "Network-Based High Level Data Classification." *IEEE Transactions on Neural Networks and Learning Systems* 23 (6): 954–70.
- Singhal, Piyush, Gopal Agarwal, and Murali Lal Mittal. 2011. "Supply Chain Risk Management: Review, Classification and Future Research Directions." *Journal of Business Science and Applied Management*. Vol. 6.
- Sutcu, Muhammed. 2020. "Effects of Total Cost of Ownership on Automobile Purchasing Decisions." *Transportation Letters* 12 (1): 18–24.
- UC Irvine. 2007. "UCI Machine Learning Repository." UCI. 2007. <https://archive.ics.uci.edu/ml/datasets.php>.
- University of Toronto. 2003. "Data for Evaluating Learning in Valid Experiments." 2003. <https://www.cs.toronto.edu/~dave/>.
- Waltman, Ludo, and Nees Jan van Eck. 2012. "A New Methodology for Constructing a Publication-Level Classification System of Science." *Journal of the American Society for Information Science and Technology* 63 (12): 2378–92.
- Zhang, Liang, Lingling Zhang, Weili Teng, and Yibing Chen. 2013. "Based on Information Fusion Technique with Data Mining in the Application of Finance Early-Warning." In *Procedia Computer Science*, 17:695–703. Elsevier B.V.
- Zhang, Shichao, Xuelong Li, Ming Zong, Xiaofeng Zhu, and Ruili Wang. 2018. "Efficient KNN Classification with Different Numbers of Nearest Neighbors." *IEEE Transactions on Neural Networks and Learning Systems* 29 (5): 1774–85.