IOJPE

ISSN: 1300 – 915X
www.iojpe.org

International Online Journal of Primary Education 2024, volume 13, issue 1

MACHINE LEARNING FOR ENHANCED CLASSROOM HOMOGENEITY IN PRIMARY EDUCATION

Faruk BULUT
Associate Professor, Istanbul Esenyurt University, İstanbul, Türkiye
ORCID: https://orcid.org/0000-0003-2960-8725
farukbulut@esenyurt.edu.tr

İlknur DÖNMEZ
Senior Researcher, Turkish Scientific and Technical Research Council, Gebze, Türkiye
ORCID: https://orcid.org/0000-0002-8344-1180
ilknur.donmez@tubitak.edu.tr

İbrahim Furkan İNCE
Assistant Professor, Istanbul Nişantaşı University, İstanbul, Türkiye
ORCID: https://orcid.org/0000-0003-1570-875X
ibrahim.ince@nisantasi.edu.tr

Pavel PETROV
Associate Professor, University of Economics - Varna, Bulgaria
ORCID: https://orcid.org/0000-0002-1284-2606
petrov@ue-varna.bg

Received: November 13, 2023 **Accepted:** March 08, 2024 **Published:** March 31, 2024

Suggested Citation:
Bulut, F., Dönmez, İ., İnce, İ. F., & Petrov, P. (2024). Machine learning for enhanced classroom homogeneity in primary education. *International Online Journal of Primary Education (IOJPE), 13*(1), 33-52. https://doi.org/10.55020/iojpe.1390421

Abstract

A homogeneous distribution of students in a class is accepted as a key factor for overall success in primary education. A class of students with similar attributes normally increases academic success. It is also a fact that general academic success might be lower in some classes where students have different intelligence and academic levels. In this study, a class distribution model is proposed by using some data science algorithms over a small number of students' dataset. With unsupervised and semi-supervised learning methods in machine learning and data mining, a group of students is equally distributed to classes, taking into account some criteria. This model divides a group of students into clusters by the considering students' different qualitative and quantitative characteristics. A draft study is carried out by predicting the effectiveness and efficiency of the presented approaches. In addition, some process elements such as quantitative and qualitative characteristics of a student, data acquisition style, digitalization of attributes, and creating a future prediction are also included in this study. Satisfactory and promising experimental results are received using a set of algorithms over collected datasets for classroom scenarios. As expected, a clear and concrete evaluation between balanced and unbalanced class distributions cannot be performed since these two scenarios for the class distributions cannot be applicable at the same time.

Keywords: Unsupervised and semi-supervised methods, class distribution, classroom homogeneity, ability grouping, similar academic performance.

INTRODUCTION

One of the key elements that is thought to promote overall success is the establishment of uniform classes at all educational levels, from kindergarten to university classes (Yoleri et al., 2022). Especially in primary and secondary school years, educators rather than parents have enough experienced that this kind of class grouping among students is better (Gabaldón-Estevan, 2020). As it is known, students, parents, and teachers who support the idea that classes should be homogeneous are in the majority compared to those who have other ideas. This argument is not only in Turkey but also in many countries (Bosworth, 2014).

The concept of homogeneity in a school classroom essentially depends on different criteria. When examining a class in terms of a certain set of criteria, it may have an equal distribution of students in some respects. However, in terms of other criteria, the class might be considered heterogeneous.

Especially in primary schools, an unbalanced class distribution can hinder both total and individual success from reaching expected levels. It is a fact that homogeneous distributions enhance both individual and overall achievement during the primary school years (Rivkin et al., 2015). In today's conditions, class distributions may become unbalanced and heterogeneous for several reasons, including random assignment of students to classes, preferences for certain teachers, personal discretion in class assignment, or the influence of requests from powerful individuals. In such cases, while the majority of students in the same class may experience rapid and lasting learning, another segment may struggle to reach the average (Gabaldón-Estevan, 2020; Rivkin et al., 2015). This discrepancy can lead to various psychological issues among students who fall behind.

One common issue among primary school teachers is managing classes that include students with contrasting characteristics, such as hardworking versus lazy, willing versus unwilling, distracted versus attentive, and interested versus indifferent. Especially in elementary school, the idea is to design the best possible class distribution that puts students in similar groups based on their qualifications in order to maximize each student's academic performance. In such heterogeneous classes, while one group of students may achieve high academic success, another may not meet the expected levels of success (Koray et al., 2003).

Additionally, there is a widespread belief among both students and parents in Turkey that being in a "good" class is necessary to excel in national university entrance exams, high school entrance exams, and TÜBİTAK National Science Olympiad exams. It is thought that a student surrounded by more hardworking peers will be motivated to work harder by example. Here, the term "good class" is essentially synonymous with a "homogeneous class" (Gabaldón-Estevan, 2020).

In both sports clubs and educational courses, there is a beneficial practice of classifying individuals into homogeneous groups based on similar physical characteristics, ages, and skill levels. For instance, while most athletes can withstand heavy and high-level training, others may not, highlighting the need for such classification. Similarly, in foreign language courses, forming groups based on language proficiency leads to faster learning outcomes. This approach also applies to activities requiring specific talents or even in patterns of social behavior. These scenarios underscore the importance of creating groups comprised of individuals with similar quantitative and qualitative characteristics, such as learning abilities, levels of different types of intelligence, and demographic factors. By taking into account the unique requirements and talents of each group, the goal is to increase the overall efficacy of training or educational programs. (Gabaldón-Estevan et al., 2014; Rivkin et al., 2015).

It has been observed that machine learning algorithms, including Classification, Clustering, and Association, are not widely applied in fields such as classroom organization or in enhancing overall student achievement, especially in contexts with limited data (Alpaydin, 2021; Bacos, 2020). This observation suggests a potential area for the application of these algorithms beyond their current use, highlighting the opportunity to leverage machine learning techniques in educational settings to improve outcomes.

This study explores the effects of classroom composition on student learning and academic performance, weighing the benefits of forming either homogeneous or heterogeneous classes. While achieving perfectly uniform conditions might be challenging, there is a contention that homogeneous classrooms can enhance both collective and individual student success. The underlying premise is that classes comprised of students with similar capabilities tend to outperform others, which forms the basis of this investigation. Additionally, the study aims to lay the groundwork for utilizing unsupervised learning techniques from disciplines such as Machine Learning, Data Mining, and Pattern Recognition to devise class distributions. It offers an introductory framework for class formation employing these methods, underscoring their utility across various data science fields. The study's theoretical foundation is rooted in data science, with a specific focus on classification and clustering methods, suggesting these as effective tools for improving educational outcomes.

Creating a dataset that encompasses both qualitative and quantitative student data is a task that demands considerable time and effort. Furthermore, several preparatory actions are necessary before data collection can commence. These include choosing suitable survey questions, securing ethics committee approval, and obtaining the requisite official permissions. Once these initial preparations are finalized, the methodology progresses to further stages, such as visiting educational institutions to collect data, processing the gathered data, integrating and fine-tuning an artificial prediction model, analyzing the resulting outputs, refining the model based on feedback, and assessing the system's reliability and performance. It is inevitable that the dataset produced through such processes may not be as extensive as desired. Consequently, the objective is to generate the most accurate predictions and viable hypotheses to inform educational strategies, even with limited data. The primary goal of this research is not to create a universal hypothesis for the entire country using minimal data but rather to elucidate the development of a predictive model through a scientific publication. Introducing a dataset that combines both real-world and synthetic data, and providing a foundational framework in this area serves an initial step towards future developments.

Following a thorough review of the existing literature, no original models relevant to the specific theme of this study were identified, which precludes the possibility of conducting a comparative analysis. Therefore, the framework introduced in this document should be considered a preliminary model. This initial model aims to ensure homogeneity in class distributions, thereby creating environments more conducive to academic achievement. As discussed in the introductory section, foundational concepts and hypotheses related to the topic are presented. Subsequent sections delve into relevant academic research within this domain, followed by an exploration of the necessary methodologies and steps for creating a dataset. Upon completing these analyses, the final section provides a conclusion based on the conducted evaluations.

Literature review

Discussions on the categorization of students based on various criteria have been prevalent since the 20th century, notably intensifying in the 1950s, as highlighted by (Lebedina-Manzoni, 2004). These discussions revolve around the optimal methods for segregating or organizing primary students within educational settings, employing diverse criteria to improve educational outcomes and address the needs of a diverse student body. Key perspectives in these debates include:

- Academic Performance: Grouping students according to their grades or test scores.
- Individual Abilities: Organizing students based on their assessed intellectual or skill-based capabilities.
- Cultural Backgrounds: Considering the impact of students' cultural contexts on their academic achievements and educational needs.

These criteria represent the core considerations in academic and pedagogical discussions regarding the most effective strategies for arranging students into classes or groups.

Research in this field varies, with some scholars emphasizing the importance of academic performance for student segregation, while others focus on sorting students by their abilities (Kuh et al., 2006). It highlights the significant impact of cultural backgrounds on students' academic success (Filatova, 2015). Additionally, Mulkey et al. (2005) posited that homogeneous groupings of students provide sustained academic benefits over time.

The examination of ability-based student categorization within educational research has yielded inconsistent and sometimes conflicting outcomes. Investigations into the efficacy of segregating students into groups or classes according to their abilities or skill levels have not produced uniform findings. Specifically, whereas certain investigations highlight the advantages or positive effects of ability grouping, others point out its disadvantages, challenges, or lack of significant impact, thus sparking a debate and resulting in a lack of consensus regarding the strategy's effectiveness. The disparate results from such studies can be attributed to various factors, including the number of classes involved, the implemented curriculum, available resources, students' socioeconomic backgrounds, the size of the educational institution, and the degree of homogeneity among the student population. Consequently, due to the variability in research outcomes, a comprehensive analysis of the contextual and foundational aspects of these studies is crucial.

Considering formal educational institutions, the notion of homogeneous classes can be segmented into primary, secondary, and tertiary levels. While these groups may be delineated based on success metrics, alternative classification criteria also apply (Yoleri, 2014). Additionally, Yoleri et al. (2013) explored factors influencing the academic integration and achievement of first-year primary students.

Gifted students and their families have shown a preference for homogeneous classes in terms of academic success, underscoring the importance of aligning educational environments with individual capabilities (Adams-Byers et al., 2004). This preference highlights the necessity for distinct evaluations across various subjects, particularly noting the significance of specialized assessments in mathematics (Oakes et al., 1995). Despite this, proponents of mixed class distribution argue for a more egalitarian approach, suggesting that segregating students based on cognitive abilities can detrimentally impact their social and psychological well-being (Hallam et al., 2001; Hodum, 2016).

Achieving true homogeneity within a classroom setting is challenged by a multitude of variables, including socioeconomic factors, individual student characteristics, and external environmental factors, all of which contribute to the inherent heterogeneity of any educational group (Hodum, J., 2016). This complexity implies that striving for homogeneity, while beneficial in certain aspects, must be balanced with the recognition of the value that diverse educational experiences bring to the learning environment.

The perspective among educators often leans towards favoring homogeneous groupings, with the argument that lesson difficulty should be adapted to match the collective level of the class, thereby optimizing the learning experience (Çelenk, 2008). Empirical studies have reinforced the notion that homogeneous class distributions can significantly influence class success, suggesting a correlation between group homogeneity and academic performance (Lu et al., 2015). Yet, the literature also acknowledges the dual-edged nature of homogeneous versus heterogeneous structures, each presenting distinct advantages and challenges to educational outcomes (Schullery et al., 2006).

Comparative analyses of student attitudes and perceptions in homogeneous and heterogeneous classes reveal that, while inclusive and mixed groupings are advocated as beneficial for most students, homogeneous settings can specifically cater to the needs of academically gifted students without compromising the educational experience of others in mixed settings (Shields, CM, 1995). This suggests a nuanced understanding of classroom composition and its impact on student engagement and achievement.

Studies looking into the wider effects of class composition on academic success and student socialization have been conducted in addition to ongoing research into flexible grouping as a classroom management strategy in heterogeneous classes (Rytivaara, 2011). (Gabaldon-Estevan, 2020). These studies contribute to the ongoing dialogue regarding the optimal balance between homogeneity and heterogeneity in educational settings, examining factors such as age, gender, ethnicity, and disability, and their influence on classroom interactions and learning outcomes.

Innovative approaches to group formation in active learning environments have also been scrutinized, with research indicating that heterogeneous groups may exhibit superior learning outcomes in certain contexts, such as physics courses, suggesting that diversity within academic groups can enhance the learning process (Briggs, 2020). Concurrently, the efficacy of cooperative learning methods in both homogeneous and heterogeneous class distributions has been examined, revealing that the strategic grouping of students can leverage the benefits of peer learning, although the results are varied and context-dependent (Wyman et al., 2020).

Internationally, research in Malaysia has aimed to identify suitable clustering models for the analysis of educational data. The goal is to align student academic pathways with national interests and to refine educational strategies based on student performance data (Salwana et al., 2017). This research highlights the intricate relationship between classroom composition, educational strategies, and academic outcomes, emphasizing the significance of a nuanced approach to student grouping and instructional design in meeting educational goals.

In summary, educational leveling serves dual purposes: safeguarding the interests of diligent, high-performing students and enhancing the academic success of those with lower performance levels. The practice of mixing students with high academic potential into groups of average-performing students proves ineffective. Furthermore, placing students with certain inherent traits alongside peers who achieve at higher levels can lead to a decline in their self-esteem.

Drawing on scientific research, a method was developed that employs data science clustering algorithms to organize classroom groups. This approach is based on the assumption that grouping students into homogeneous classes is beneficial.

METHOD

Research Model

In a data set where there are samples without a class label and with certain attributes, the samples can be separated into different clusters. In the literature, this process is also called clustering and is referred to under the heading of unsupervised learning. There is no supervisory mechanism based on the class label. Finding the class label in the samples and finding which class the new test samples will belong to is called classification and is also referred to as supervised or supervised learning in the literature. The defining element here is the class label. On the other hand, the process of separating a small number of labeled samples and a large number of unlabeled samples into certain classes or clusters is called semi-supervised learning (Taghizabet et al., 2023).

Almost all of the academic and practical studies with data science in the field of education have been carried out with supervised learning methods (Holmes et al., 2019). Unlike other studies, the general structure of unsupervised learning techniques and the details of how they will be used in the subject we are working on will be discussed. First, how to create a data set, what the preliminary preparations are and how to make it ready for use will be discussed.

Design of the study

The flow chart in Figure 1 summarizes and presents the artificial decision system. It starts with defining the problem, followed by data collection, data processing, model building and fitting, and running clustering methods. The ultimate goal is to create a hypothesis and gain the necessary wisdom.

The process described transforms raw data into information using statistical and machine learning algorithms. This information then becomes knowledge through data science algorithms. Finally, wisdom is derived from this knowledge, summarizing the data science and wisdom discovery stages as illustrated in Figure 2. The DIKW Hierarchy is a model that represents the transformation of data into wisdom through various stages:

1. Data : Raw or unprocessed data.
2. Information : Processed, organized, or structured data to provide context and meaning.
3. Knowledge : Processed, analyzed, understood information for decision-making.
4. Wisdom : The highest level, where knowledge there are experience and judgment to make informed decisions.

IOJPE
ISSN: 1300 – 915X
www.iojpe.org
International Online Journal of Primary Education 2024, volume 13, issue 1

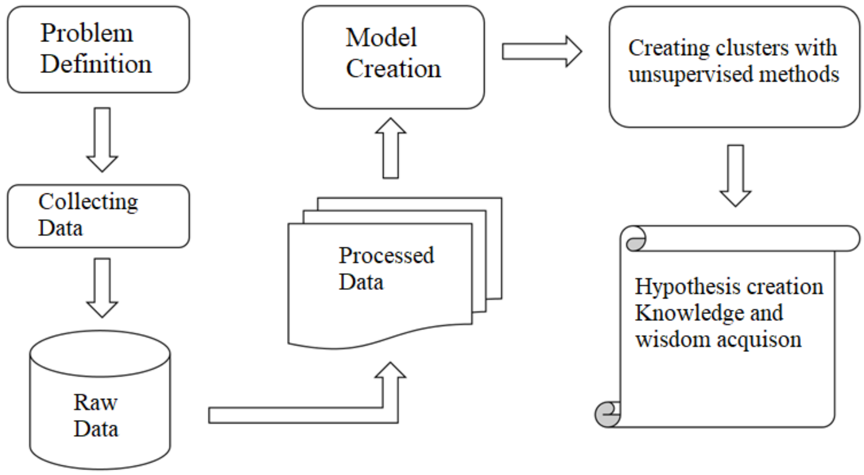
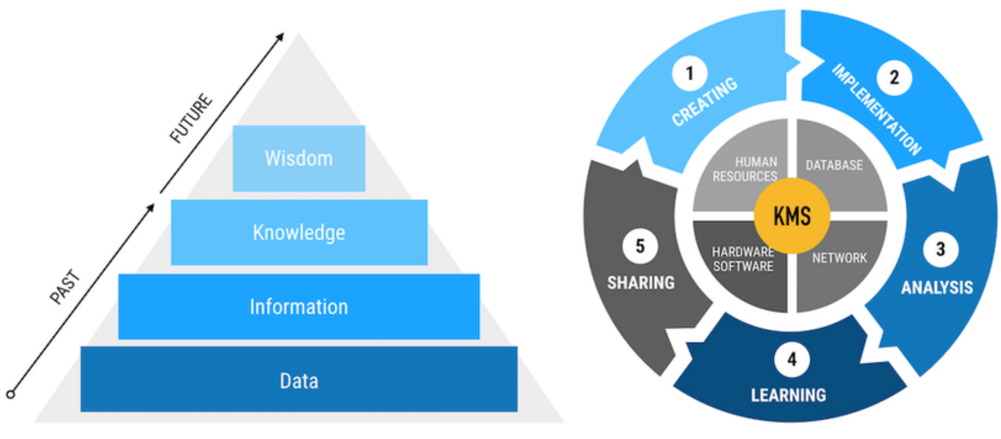


Figure 1. Flow chart of the study.

In Figure 2 (b), the Life Cycle of Data refers to the processes through which data is collected, processed, analyzed, and disposed of or archived.



(a) (b)

Figure 2. Data analytics pyramid of Data-Information-Knowledge-Wisdom (DIKW).

The process in data science normally begins with raw data, comprising numbers, characters, and symbols. From this, relational information and patterns are discerned. This leads to the formulation of solutions and hypotheses for existing problems. Ultimately, the entire process is evaluated from a broad perspective, utilizing human intelligence and understanding to derive insights and experiences that illuminate future paths.

Data collection and process management

The data obtained in this study has been made with official permissions and authorizations. A survey has been conducted with the students at the TOKİ Karşıyaka Municipality Primary School, under the permission document numbered 15585669-604.01.02-E.3979850 and dated April 8, 2016. As outlined on the website (Website, 2024), the primary instrument for data collection was a carefully designed survey, ensuring the gathering of pertinent and valuable information. The data collection involved ten first-grade students, proceeding with explicit consent from their parents and under the meticulous supervision and support of their classroom teacher. Data collection was conducted on 12 first-grade primary students in 2016. Eight boys, all seven years old, were among the students; they were all in good health and did not have any mental or physical problems.

As it is known, collecting data from primary students is a very difficult process. Therefore, to address this challenge, the collected data were resampled by using a specific algorithm (Bulut, 2020) to create a sufficient dataset comprising 24 individuals of both real and synthetic data. Therefore, this number is known to be equivalent to the population size of a typical classroom.

To evaluate the success of this study, an analysis of academic performance was conducted, comparing achievements at the conclusion of both the first and second academic terms. As a data collection tool, the survey questions specified on the website (Website, 2024) were applied to the students. The academic achievements of the students at the end of the fall and spring semesters were evaluated according to the data science criteria specified in the article in order to obtain experimental results.

Optimal attribute selection of students

Effective classroom grouping hinges on the identification of pertinent student characteristics, necessitating data collection from educators, psychologists, students, and parents. This comprehensive approach affords a detailed insight into the dynamics of classroom composition. Additionally, evaluating students' academic and extracurricular achievements, alongside personal attributes like sociability and self-confidence, offers a holistic gauge of academic success. Critically, evaluations must avoid discriminatory factors to maintain fairness in the assessment process. Within the field of artificial intelligence, the success of an application greatly depends on the quality and relevance of the selected features. A thorough examination of existing literature has pinpointed the principal factors that directly or indirectly affect educational and training outcomes:

- Turkish course success (out of 5)
- Mathematics course success (out of 5)
- Science course success (out of 5)
- English course success (out of 5)
- Visual Arts course success (out of 5)
- Music course success (out of 5)
- Physical Activities course success (out of 5)
- Does he/she have learning difficulties? (Y/N)
- Does he/she like to do sports? (Y/N)
- Does he/she have a habit of doing homework? (Y/N)
- Does he/she have a habit of reading books? (Y/N)
- Does he/she play a musical instrument? (Y/N)
- Could he/she be class president (representative)? (Y/N)
- Is he/she inquisitive? (Y/N)
- Does this person have social characteristics? (Y/N)
- Does he/she like to share? (Y/N)
- Does he/she have self-confidence? (Y/N)
- Are his/her parents together? (Y/N)

Prior to initiating data science processes, it is crucial to digitize responses for inclusion in the dataset, incorporating observational information as needed. Maintaining ethical standards in data gathering requires both getting ethics committee approval and making sure the questions respect individual rights and freedoms.

Dataset preparation procedure

Accurate quantification of attributes is crucial for data clustering. While binary questions are usually coded as 1 or 0, academic performance can be represented by integers on a scale of 4, 5, or 100, and rational numbers can also be used. Answers to multiple choice questions are coded as nominal or categorical values such as "a", "b", "c". Clustering and classification techniques make use of algorithms designed to handle these nominal values efficiently, streamlining the process without complications. All of the numbers in the created training set can be integer, real number and nominal types. All nominal data must first be converted to numeric values with the "nominal_to_binary" method. In this way, both clustering algorithms can be tested in the system, and the performance of these algorithms is increased. During data collection, unanswered questions lead to missing values, a significant issue. The literature suggests various methods to estimate and fill these gaps to maintain data integrity (Zhu et al., 2010). Numerical data should be normalized to a [0, 1] range, especially when it spans a wide scale, to improve clustering method performance (Bulut, 2016), (Bulut et al., 2016). The Min-Max and Z-Score normalization techniques are recommended to address irregular distributions and potential data issues (Jayalakshmi et al., 2011). Multidimensional datasets often become sparse as the number of features increases, which can negatively impact clustering success. To address this, increasing the sample size or using dimensionality reduction algorithms like Principal Component Analysis, Linear Discriminant Analysis, Factor Analysis, t-SNE, and others can help reduce dimensions and improve data density (Gao et al., 2021; Wang et al., 2021; Mair, 2018; Xiao et al., 2023; Groth et al., 2013). Before creating a hypothesis, it is critical to evaluate the impact of each feature on cluster results using methods such as information retrieval. Removing low-impact features can increase dataset density and improve clustering results. This approach, which involves eliminating unnecessary dimensions and outliers, leads to improved data usage for analysis. The study examined traits based on their distribution to achieve cumulative success (Zhang et al., 2003).

Analysis of data

The data analysis process is critical in transforming raw data into meaningful insights through a structured approach that includes data preprocessing, application of machine learning algorithms, and the evaluation of results.

Data preprocessing is the first step that involves cleaning and normalizing the data to ensure it is suitable for further studies. This phase lays a solid foundation for applying machine learning algorithms by addressing issues such as missing values, outliers, and data format inconsistencies.

The cleaned data is then analyzed using various machine learning algorithms facilitated by programming languages such as Python and MATLAB. This study focuses on classification, clustering, and semi-supervised learning techniques. In particular, the clustering algorithms include k-means, Hierarchical clustering, Self-Organizing Maps (SOM), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and Clustering with Voronoi Diagrams. Each algorithm is selected for its ability to uncover patterns and groupings within data that align with research objectives.

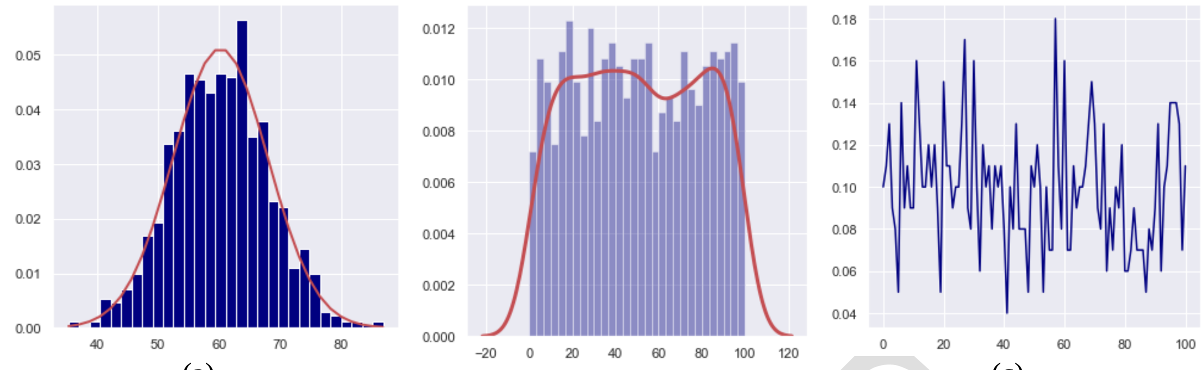
The results of these algorithms are critically evaluated to evaluate their performance and effectiveness. For example, clustering algorithms are evaluated on their ability to form consistent and meaningful clusters using metrics such as silhouette scores. The evaluation phase is crucial to interpreting the data in light of the objectives of the study and allowing meaningful conclusions to be drawn that will contribute to the broader field of data analysis and machine learning.

EXPERIMENTAL RESULTS AND FINDINGS

Normally distributed features

Many phenomena in nature show a normal distribution. As seen in Figure 3, the bar graph is in the form of a Gaussian curve, that is, a bell curve. The success of the students in a class in a lesson is explained with this graphic. It can be observed that unsuccessful and very successful students are in the minority, while those with normal success are in the majority. In Figure 3, the numbers in the *x* axis [0, 100] indicate students' academic achievement; the *y* axis shows the frequency status.

IOJPE
ISSN: 1300 – 915X
www.iojpe.org
International Online Journal of Primary Education 2024, volume 13, issue 1



(a) (b) (c)

Figure 3. (a) Histogram of the Gaussian (normal) distribution, **(b)** Histogram of the uniform distribution, **(c)** Histogram of random distribution, and the *x*-axis is the student course success average.

Uniformly distributed features

The uniform distribution gets its name from the fact that the probabilities for all outcomes are approximately the same. The probability of each outcome occurring is almost equal. Here, the mean and median values are almost the same. Because every outcome in a single distribution occurs at the same relative frequency, the resulting shape of the distribution is almost the shape of a rectangle. As can be seen in Figure 3 (b), the distribution related to student course success is uniform. The number of students in almost every success category is the same. The class with this type of distribution is heterogeneous and is undesirable for educators. In the dataset, there is a distribution suitable for a uniform distribution in some attributes.

Randomly distributed features

In this type of distribution, as seen in Figure 3 (c), it is the case that the attributes of each student exhibit a different distribution and cannot be evaluated within a certain distribution type. These features do not contribute to the success of the clustering algorithm. In other words, they behave somewhat like noise, making it difficult to find similar clusters.

Clustering Methods

Within the realms of Machine Learning, Data Mining, and Pattern Recognition, algorithms designed for classification, prediction, clustering, and association tasks are employed to achieve a uniform distribution of students and can also be utilized to ascertain grade levels. Unsupervised clustering algorithms enable grouping without adherence to predefined conditions or rules (Ian et al., 2005). The clustering process relies solely on the inherent functionality of the chosen algorithm. These unsupervised algorithms are broadly categorized into four principal domains:

1. Divisional methods: For example, k-Means (Xiao et al., 2018) and k-Medoids (Kaufman et al., 2009) methods.
2. Hierarchical methods: Agglomerative Clustering as an example (Han et al., 2023).
3. Density-based methods: For example, Density-based spatial clustering of applications with noise method (Sheridan et al., 2020).
4. Model-based methods: For example, Gaussian Mixture Model Expectation Maximization (Xie et al., 2023), Self-organizing map (SOM) or self-organizing feature map (Motegi et al., 2023) methods can be given.

These frequently used clustering methods show differences when compared to each other in terms of characteristics. These methods are used to detect hidden patterns and structures in a data set. In our study, it is predicted that density-based clustering methods will give more successful results than others will in order to create the most appropriate cluster. Because students with similar characteristics are concentrated in certain regions of the data set, it would be more appropriate to evaluate those regions as a group.

Comparison of the level of success achieved because of placing students in the most appropriate classes with the clustering process can be done in three ways:

1. With analytical criteria defined in the literature (Backer et al., 1981).
2. With the objective exam results obtained from the courses.
3. With the subjective confirmation of educators and parents working in the relevant educational institution.

***k*-means method**

This algorithm is the most efficient and accurate bundling method known. For a *k* value known as a parameter, the k-means bundling algorithm has 4 basic steps:

1. The dataset is divided into *k* subsets according to their proximity to *k* randomly determined points. Each tuple is treated as a subset. The *k* parameter here is essentially the number of classes planned to be created.
2. The average of each new bundle created according to the selected centers is calculated, and the center point is found (the average of the attributes of the objects in the bundle).
3. Each data point is included in the bundle with the nearest central point. New bundles are created.
4. Return to step 2 until the bundling of the data remains unchanged.

The advantage of this method is that the complexity is low, fast results can be reached even with large data, and the results are visual. The disadvantage is that the results depend on the initial value; it is possible to give different results in each run.

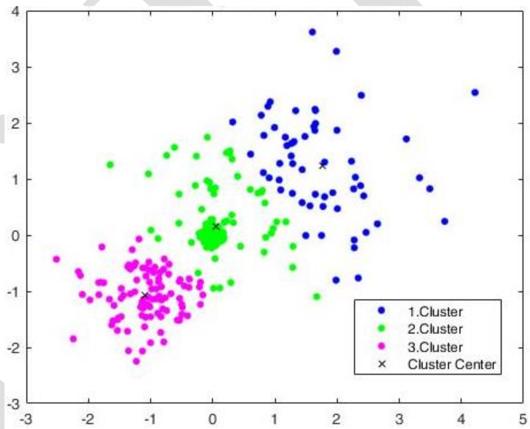


Figure 4. Example of a distance-based clustering (*k*=3).

As in Figure 4, in a data set that has only two features, in other words, *x* and *y* features, each point represents an individual. These space-scattered data are divided into clusters by the k-means method, which is a density-based method. In this clustering process, an illustrative representation of a data set with only two features can be easily performed. However, it is impossible to transfer multidimensional data with three or more attributes to a two-dimensional page layout.

In Figure 5, how the same data set, which is desired to be clustered, can be divided into clusters with different numbers has been tested with the codes written in the MATLAB environment. All MATLAB and Python codes, datasets, and results can be examined for testing purposes from the given web address (Website, 2024) and further studies can be carried out. When the simulation results were examined, a smooth and deterministic clustering result was obtained as the test points were distributed close to each other without drawing a specific pattern.

All of these tests can be repeated by deriving synthetic data sets with different correlations between the features. For such cases, the results of the algorithms found in the literature and mentioned in this study on these data sets can also be examined in further stages. We would like to point out that the clustering results on data sets created according to random distribution and a certain correlation will be different for each algorithm.

It is requested to derive *k* homogeneous clusters for this dataset,. That is, *k* is an hyper-parameter entered into the system from outside. Since some clustering methods (OPTICS, GMM EM, DBSCAN, etc.) are density-based and adaptive, the number of clusters is determined automatically (Bulut et al., 2021). Since the value of k, that is, the number of classes in the school is in a deterministic structure, other methods will not be needed.

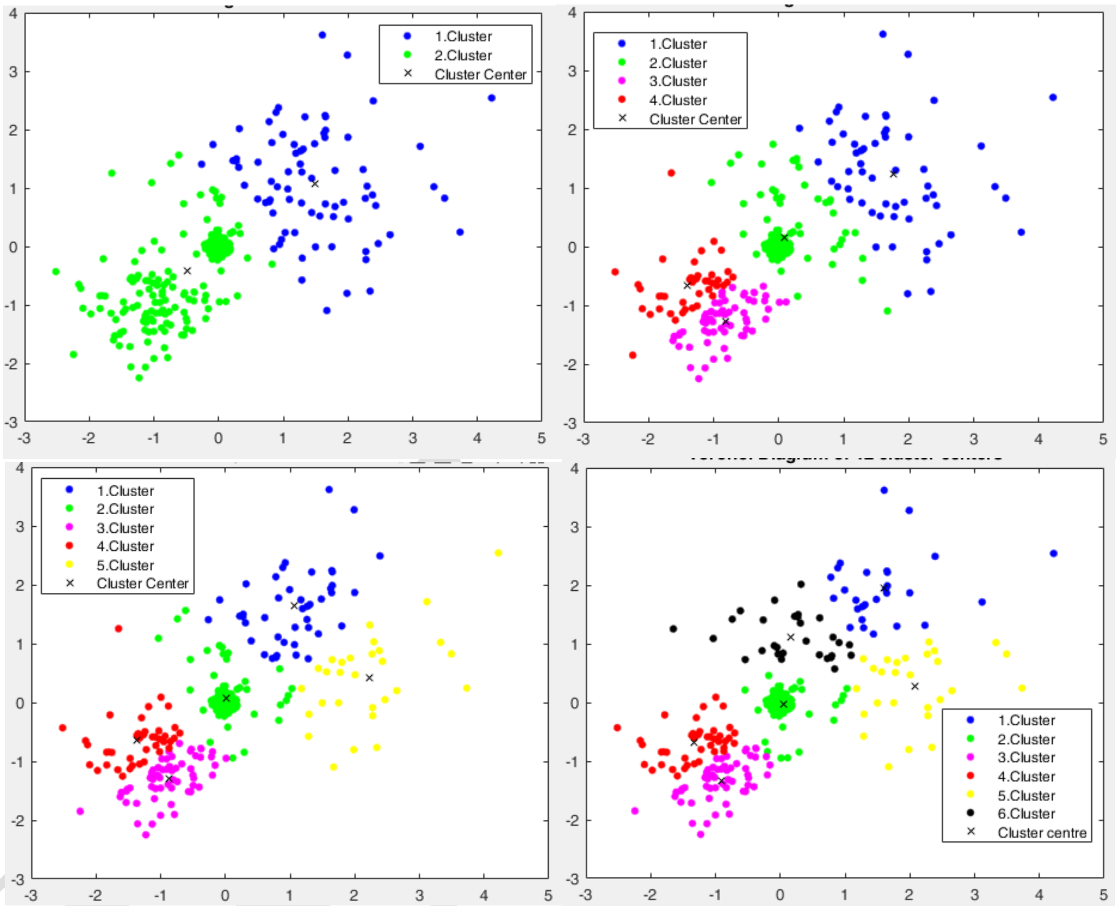


Figure 5. Clustering forms in a density-based classification example when the *k* parameter is selected as 2, 4, 5 and 6, respectively.

Hierarchical (Agglomerative) clustering method

Kaufmann and Rousseeuw introduced a hierarchical clustering method in 1990 (Kaufman et al., 2009), where each object forms a tuple, and tuples with the least distance are merged step by step. This process continues until a single tuple encompasses all objects or a desired number of tuples is reached, as depicted in Figure 6, demonstrating the bottom-up merging approach for dataset elements.

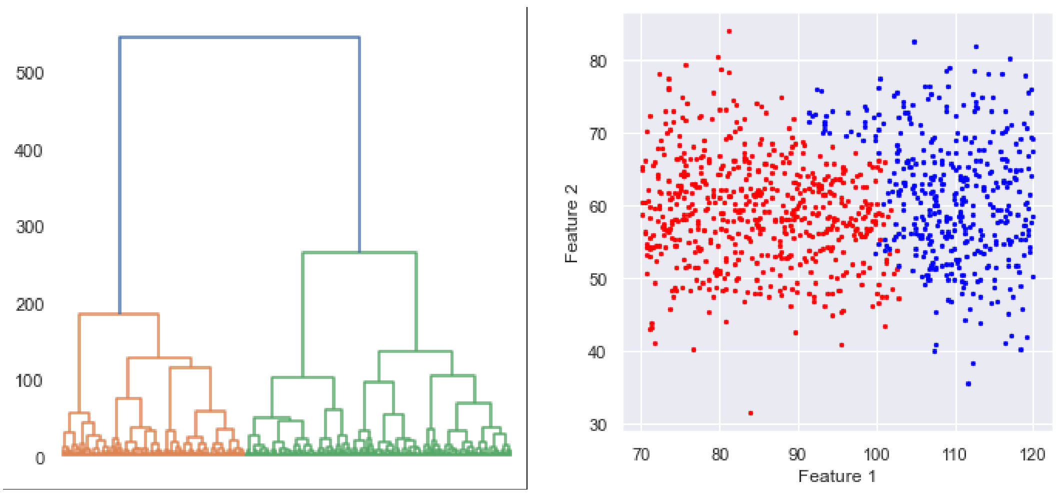


Figure 6. An example of a density-based classification.

Useful for small-scale data. Since it goes from many details to the whole, it makes comparisons one by one in large-scale data and combines similar ones, so the processing time is longer.

Density-based SOM Clustering method

SOM (self-organizing maps), which can be structurally an example of feed forward networks, behaves like the k-means algorithm for a very small number of neurons. With the increase in the number, the difference of SOM also emerges. The working steps of the algorithm are as follows:

1. The weight values in the neurons in the network are randomly assigned.
2. The input vectors are taken.
3. All values in the map are traversed.

a. The distance between the input vector and the map value being traversed is the Euclidean distance.

b. The node with the shortest distance is taken (this method is called the best matching unit).

4. All nodes adjacent to this optimal node are updated and approximated to the input vector.
5. As long as the current step is less than the time limit on the step, the process is repeated by returning to step 2.

The main advantage of using SOM is that the data is easily interpreted and understood. Reducing dimensionality and grid clustering makes it easier to observe similarities in data. Each input data is inserted into the artificial neural network one by one, and similar trained data are gathered together in a small-sized space. An example SOM scenario is given in Figure 7. As can be seen, the samples in different clusters are nested.

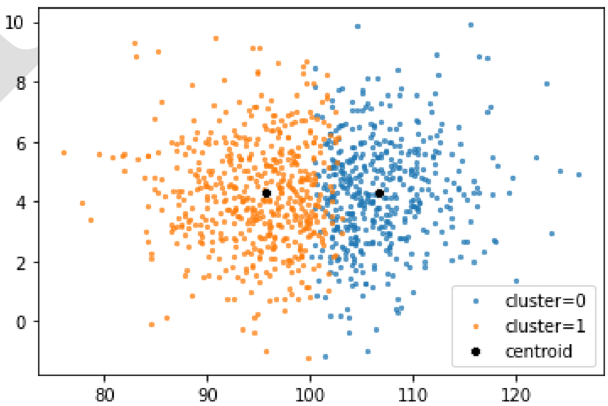


Figure 7. An example clustering scenario obtained with SOM (Self-Organizing Maps).

Model-based DBSCAN clustering method

Density-based clustering algorithms operate with two key parameters: Epsilon, which denotes the maximum radius of a neighborhood, and MinNumber of Objects, indicating the minimum quantity of objects within an Epsilon-radius neighborhood. The algorithm searches for the neighborhood of each database object within an epsilon radius. Clusters are formed around a seed object, object *p*, which has a number of objects exceeding the MinNumber of Objects in its vicinity. Core objects are those that can be directly reached. Clusters connected by density are merged together. The algorithm concludes its process when no additional objects can be added to a cluster.

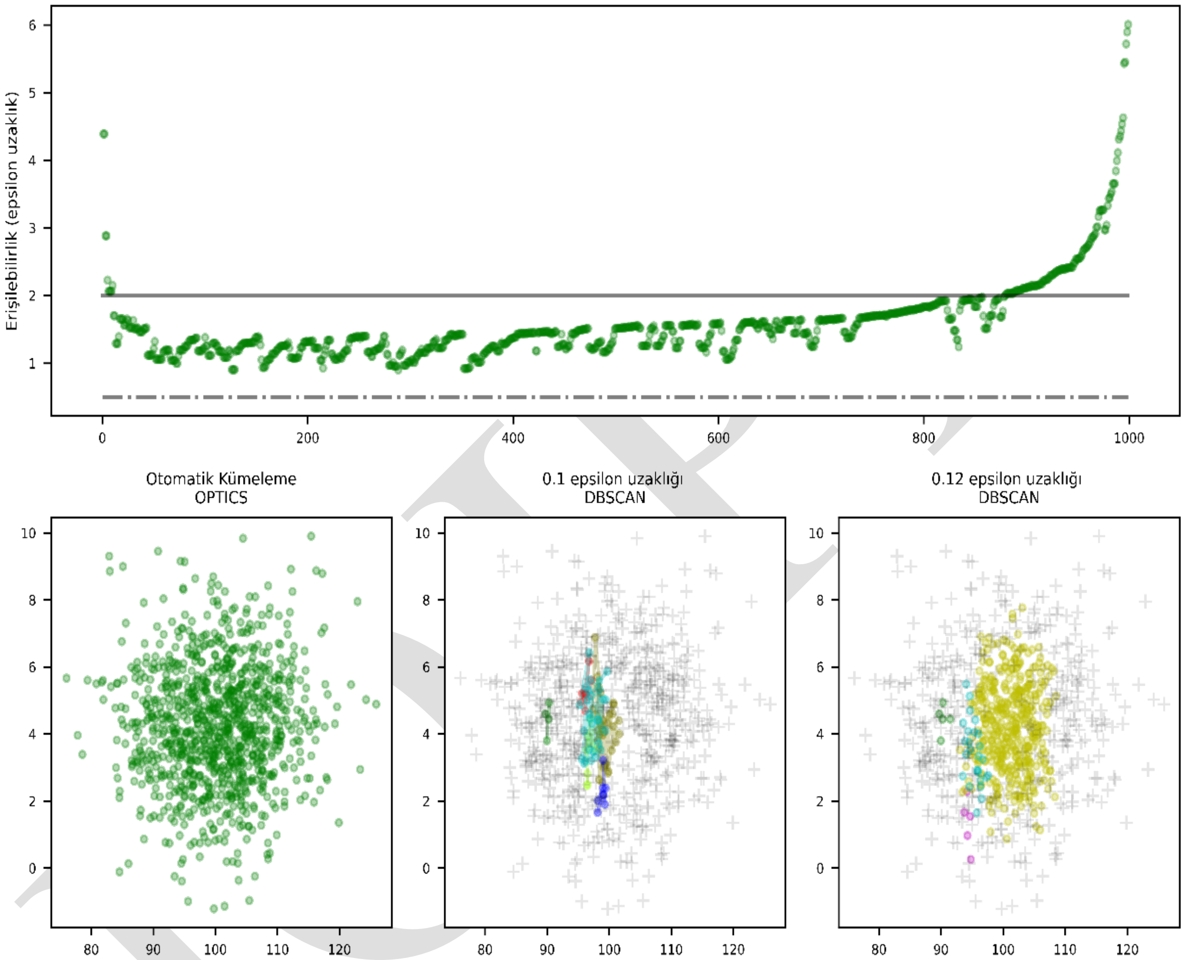


Figure 8. An example of a density-based classification by using OPTICS and DBSCAN. The *x* axis is the accessibility distance.

The methodologies of OPTICS and DBSCAN algorithms focus on achieving the most precise clustering by automatically adjusting parameters based on a predetermined step value. However, this approach may inadvertently overlook the optimal parameter setting as it progresses through the steps. As illustrated in Figure 8, identifying clusters with OPTICS proved challenging. Altering the Epsilon parameter yielded different clusters, but due to the lack of sufficient distinctiveness among these clusters, the data sizes were subsequently reduced.

IOJPE
ISSN: 1300 – 915X
www.iojpe.org
International Online Journal of Primary Education 2024, volume 13, issue 1

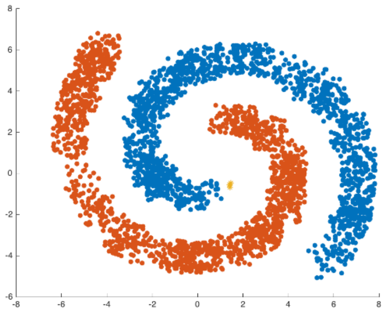


Figure 9. DBSCAN parsing nested spiral clusters in a 2D space. The *x* axis is the accessibility distance.

The biggest difference between the DBSCAN algorithm and other algorithms is that it starts in the densest regions and follows the closest points. For this reason, as seen in Figure 9, it can better detect the data extending by following it in a certain order in space (as it does not evaluate only according to location proximity).

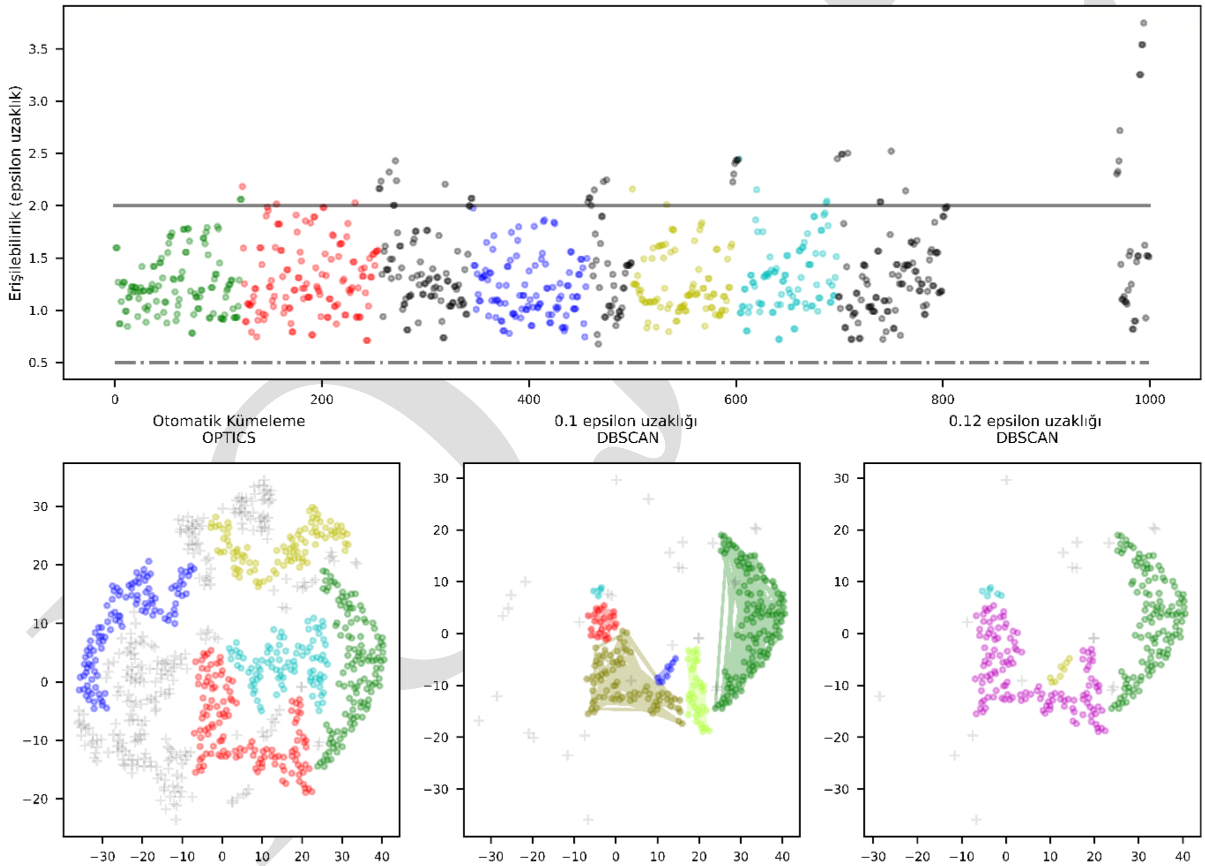


Figure 10. An example of density-based classification after t-SNE by using OPTICS and DBSCAN.

DBSCAN with data reduced in size with t-SNE

When the data size is too large, using principal component analysis (PCA) or t-distributed stochastic neighbor embedding (Xiao et al., 2023) can first reduce the size of the data and then apply DBSCAN. For data representations with more than two features, reducing the size to two or three can provide better results for clustering algorithms by both increasing visibility and reducing sparsity.

As seen in Figure 10, clusters can be formed more clearly, when DBSCAN is applied by reducing the data to two dimensions. The *x* axis is the accessibility distance.

IOJPE
ISSN: 1300 – 915X
www.iojpe.org
International Online Journal of Primary Education 2024, volume 13, issue 1

Clustering with Voronoi Diagram
100 students who will study in the first grade in a primary school are to be distributed equally to five classes. By looking at the attribute information of the students, the algorithms to be used in the project can distribute these students among five classes in a homogeneous structure. Clustering algorithms do not ask the user the number of cluster elements to be determined as a parameter. They choose the most suitable cluster members themselves, depending on their working mechanism. In this case, the number of elements in one set may be considerably higher than the other. In such cases, the data space can also be fragmented with the help of Voronoi diagrams (Feng et al., 2018).

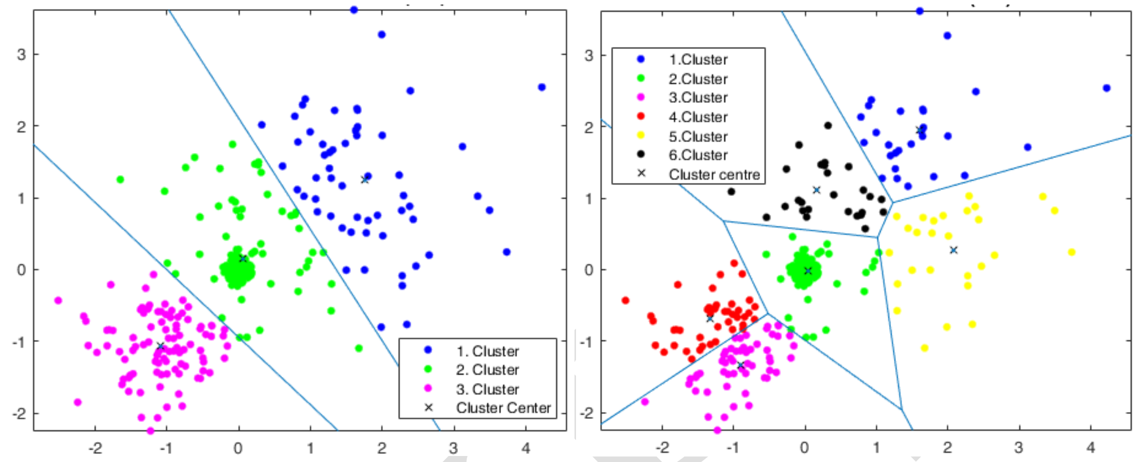


Figure 11. Determination of cluster boundaries (blue lines) with Voronoi diagram (two different cases for k=3 and k=6).

Figure 11 shows how the clustering result will be with the Voronoi diagram for k=3 and k=6 cases on the same data set. If the number of classes to be opened in the school is 3 or 6, the points between the boundary lines can be manually shifted to create an equal number of classes in another cluster.

The biggest problem that can be encountered in the studies carried out with the help of current algorithms that create the Voronoi diagram is that the obtained outputs, namely the resulting clusters, cannot be displayed in two-dimensional environments and submitted to human control. In such cases, as can be seen in the calculation below, individual examinations can be performed by producing two-dimensional Voronoi diagrams, usually as many as the 2-combinations of ($n(n$-1)/2).

Clustering/classification with Semi-Supervised Learning
Semi-Supervised Learning is the process of separating a small number of labeled samples and a large number of unlabeled samples into certain classes or clusters (Hady et al., 2013). Bee-supervised learners, on the other hand, ensure that a homogeneous class is formed by selecting other students who are suitable for the qualifications of the determined students, in case a teacher requests some students that he/she definitely wants to be in his class. Semi-supervised clusterers can be broadly subdivided as follows (Basu et al., 2002):

1. Generative models
2. Low-density separation
3. Graph-based methods
4. Heuristic approaches

After the data set is created, the four different methods mentioned can be run separately, and the educators can be expected to evaluate the results obtained. A teacher determines some of the students that he/she wants to be in his class, and the other students are distributed to the classes accordingly. This necessitates the use of semi-supervised learning algorithms.

It is an important issue whether supervised and unsupervised methods show success on a given dataset at the expected level, and different and valuable studies have been carried out in this area (Basu et al., 2002), (Maulik et al., 2002). A comparative study can also be made on the data set created by these methods.

DISCUSSIONS, FURTHER STUDIES, and CONCLUSION

The present study introduces a novel framework for the allocation of students to classes in the primary education sector, leveraging insights from a variety of methodologies evaluated against a set of established benchmarks. Each methodology brings to light unique benefits, highlighting the richness and potential inherent in diverse allocation strategies (Johnson et al., 2009). This research addresses the complexities involved in achieving an equitable and fair distribution of students across elementary classes, acknowledging the limitations of purely algorithmic solutions in addressing the nuanced demands of such allocations. Advocating for a more comprehensive approach, the study underscores the importance of incorporating expert opinions, evaluations by educators, input from parents, and a detailed analysis of student characteristics as integral to devising an efficacious class distribution system (Oakes, 2005). By amalgamating these factors, the research proposes an all-encompassing framework for student placement that is sensitive to both academic and social considerations, while also being responsive to the individual needs and capabilities of students in a primary educational context. The envisioned system is aimed at creating a more equitable and conducive learning environment, thereby facilitating improved educational outcomes and fostering student growth (Slavin, 1996).

Building upon the innovative approach to student class assignment in primary education proposed in the study, several recommendations for practitioners and further research avenues emerge. These suggestions aim to refine the comprehensive system for student placement, ensuring it effectively meets the diverse needs of students while enhancing educational outcomes.

There are some suggestions for Practitioners and researchers:

- Implementing Pilot Programs: Schools should consider piloting the proposed class assignment system under a controlled regulation. This would allow educators to identify practical challenges and assess the system's impact on student learning and social integration.
- Professional Development: Offer training programs for teachers and administrators on the holistic approach to student assignment. This includes understanding the importance of incorporating expert opinions, educator evaluations, and parental feedback.
- Continuous Feedback Mechanism: Establish mechanisms for ongoing feedback from students, parents, and educators to continually refine the class assignment process. This also ensures the system remains adaptive to changing student needs and educational goals.
- Comparative Research: Undertake comparative studies to evaluate the effectiveness of the proposed system against traditional class assignment methods. This could involve quantitative measures of student achievement and qualitative assessments of student and teacher satisfaction.
- Technology Integration: Investigate the potential for integrating advanced technologies, such as machine learning algorithms and artificial intelligence, to enhance the efficiency and effectiveness of the student assignment process while maintaining a holistic approach.

By addressing these recommendations and further studies, practitioners and researchers can contribute to the development of more equitable, effective, and responsive educational environments. This collaborative effort will pave the way for optimizing class assignments in primary education, ultimately fostering a more balanced and beneficial educational experience for students.

Conclusion

The innovative model presented for allocating students to classes increases the capabilities of artificial intelligence to significantly develop both the performance of individual students and the overall effectiveness of classrooms. By focusing on the use of Unsupervised and Semi-Supervised Learning algorithms, which are particularly adept at handling smaller datasets, the model aims to increase total educational success. Unlike traditional approaches in the literature, this study does not attempt into the creation of new technological tools or the development of novel algorithms. Rather, it applies established computer science algorithms to a new domain, specifically to the realm of primary education, thus demonstrating the versatility and potential of these algorithms when applied outside their usual contexts. The creation of this blueprint model presents a pioneering step towards integrating computer science principles into educational strategies by offering a class distribution style. Additionally, this initiative enables a way for interdisciplinary collaboration, inviting experts from various fields to contribute to the enhancement of educational outcomes through technological innovation.

This presented research provides a preliminary exploration into harnessing the power of computer science for optimizing classroom allocations in primary schools. It is notable for its creative use of existing methods in literature and its contribution to academic discourse by highlighting the availability of innovation across different fields. By merging educational goals with data science methods, this research not only introduces a novel solution to a common problem but also motivates further studies and development efforts aimed at enhancing the quality of primary education through artificial intelligence.

Availability

All of the official permission documents, survey questions, generated datasets, classification, and clustering codes written in MATLAB and Python programming languages for this research study can be obtained and downloaded from the following website (sites.google.com/site/bulutfaruk/study-of-clustering-on-education) for further studies. The specified web address offers access to various details and documents related to the research, including official permissions. It also features project details, the work area, survey questions, and sample petitions used in the official permit application. These documents, often confidential, are made available to aid potential researchers in understanding the necessary preparatory steps before initiating their projects.

Acknowledgments

We would like to express our endless thanks and gratitude to the Governorship of Izmir (*İzmir Valiliği*) and the Çiğli National Education Directorate (*Çiğli İlçe Milli Eğitim Müdürlüğü*) for allowing this study, and to the administrators, teachers, and students of TOKİ Karşıyaka Municipality Primary School (*TOKİ Karşıyaka Belediyesi İlkokulu*) who helped us in collecting data and carrying out the necessary experimental studies.

Ethics and Conflict of Interest

The survey studies were carried out with the official authorization under document number 15585669-604.01.02-E.3979850, dated April 08, 2016, acquired from the Çiğli National Education Directorate. The data collection involved ten first-grade students, proceeding with explicit consent from their parents and under the meticulous supervision and support of their classroom teacher. Since the official permissions and data used in the article study were obtained before the year 2020, there was no obligation to obtain an official permission from the Ethics Committee. The authors declare that they have no conflict of interest.

Author Contributions

The authors' contributions to the research article are almost equal.

IOJPE

ISSN: 1300 – 915X
www.iojpe.org

International Online Journal of Primary Education 2024, volume 13, issue 1

Corresponding Author
Correspondence to Faruk BULUT, farukbulut@esenyurt.edu.tr

REFERENCES

Adams-Byers, J., Whitsell, S. S., & Moon, S. M. (2004). Gifted students' perceptions of the academic and social/emotional effects of homogeneous and heterogeneous grouping. *Gifted Child Quarterly*, *48*(1), 7-20.

Alpaydin, E. (2021). *Introduction to machine learning*. Adaptive Computation and Machine Learning series, MIT press.

Backer, E., & Jain, A. K. (1981). A clustering performance measure based on fuzzy set decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1), 66-75.

Basu, S., Banerjee, A., & Mooney, R. (2002). Semi-supervised clustering by seeding. *In Proceedings of 19th International Conference on Machine Learning* (ICML-2002).

Bellinger, G., Castro, D., & Mills, A. (2004). Data, information, knowledge, and wisdom. https://homepages.dcc.ufmg.br/~amendes/SistemasInformacaoTP/TextosBasicos/Data-Information-Knowledge.pdf

Bosworth, R. (2014). Class size, class composition, and the distribution of student achievement. *Education Economics*, *22*(2), 141-165.

Briggs, M. (2020). Comparing academically homogeneous and heterogeneous groups in an active learning physics class. *Journal of College Science Teaching*, *49*(6), 76-82.

Bulut, F. (2016). Çok katmanli algilayicilar ile doğru meslek tercihi [Right career choice using multi-layer perceptron]. *Anadolu University Journal of Science and Technology A-Applied Sciences and Engineering*, *17*(1), 97-109.

Bulut, F., & Amasyali, M. F. (2016). Katı kümeleme ve yeni bir geçiş fonksiyonuyla uzman karışımlarında sınıflandırma [Classification in mixture of experts using hard clustering and a new gate function]. *Journal of the Faculty of Engineering and Architecture of Gazi University*, *31*(4), 1017-1025.

Bulut, F. (2020). The minimum ratio of preserving the dataset similarity in resampling: (1− 1/e). *International Journal of Information Technology*, *12*(1), 231-244.

Bulut, F., Bektaş, M., & Yavuz, A. (2021). Efficient path planning of drone swarms over clustered human crowds in social events. *International Journal of Intelligent Unmanned Systems*.

Çelenk, S. (2008). İlköğretim okulları birinci sınıf öğrencilerinin ilkokuma ve yazma öğretimine hazırlık düzeyleri [Level of readiness for reading and writing education among first grade students of primary schools]. *Abant İzzet Baysal University Journal of the Faculty of Education*, *8*(1), 83-90.

Feng, X., & Murray, A. T. (2018). Allocation using a heterogeneous space Voronoi diagram. *Journal of Geographical Systems*, *20*(3), 207-226.

Filatova, O. A. (2015). Cultural attributes of students to make student-centered approach successful. *International Journal of Languages, Literature and Linguistics*, *1*(1), 20-24.

Gabaldon-Estevan, D. (2020). Heterogeneity versus homogeneity in schools: A review of the educational value of classroom interaction. *Education Sciences*, *10*(11), 335. https://doi.org/10.3390/educsci10110335

Gao, Y., Lin, T., Zhang, Y., Luo, S., & Nie, F. (2021). Robust principal component analysis based on discriminant information. *IEEE Transactions on Knowledge and Data Engineering*.

Groth, D., Hartmann, S., Klie, S., & Selbig, J. (2013). Principal components analysis. *Computational Toxicology: Volume II*, 527-547.

Hady, M. F. A., & Schwenker, F. (2013). Semi-supervised learning. *Handbook on Neural Information Processing*, 215-239.

Hallam, S., & Ireson, J. (2001). Ability grouping in education. *Ability Grouping in Education*, 1-240.

Han, X., Zhu, Y., Ting, K. M., & Li, G. (2023). The impact of isolation kernel on agglomerative hierarchical clustering algorithms. *Pattern Recognition*, 139, 109517.

Hodum, J. (2016). *Ability grouping for academic growth in the elementary school*. Union University.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education. Boston: Center for Curriculum Redesign*.

Ian, HW, & Eibe, F. (2005). *Data mining: Practical machine learning tools and techniques*.

Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical normalization and back propagation for classification. *International Journal of Computer Theory and Engineering*, *3*(1), 1793-8201.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365-379. https://doi.org/10.3102/0013189X09339057

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Koray, Ö., & Tatar, A. G. N. (2003). Primary school students' misconceptions about mass and weight and the distribution of these misconceptions according to the 6th, 7th and 8th grade levels. *Pamukkale University Journal of Education Faculty*, *13*(13), 187-198.

Kuh, G. D., Kinzie, J. L., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature* (Vol. 8). Washington, DC: National Postsecondary Education Cooperative.

Lebedina-Manzoni, M. (2004). To what students attribute their academic success and failure. *Education*, *124*(4), 699-709.

Lu, F., & Anderson, M. L. (2015). Peer effects in microenvironments: The benefits of homogeneous classroom groups. *Journal of Labor Economics*, *33*(1), 91-122.

Mair, P. (2018). Factor analysis. In *Modern Psychometrics with R* (pp. 17-61). Springer, Cham.

Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. IEEE *Transactions on pattern analysis and machine intelligence*, *24*(12), 1650-1654.

Motegi, R., & Seki, Y. (2023). SMLSOM: The shrinking maximum likelihood self-organizing map. *Computational Statistics & Data Analysis*, 182, 107714.

Mulkey, L. M., Catsambis, S., Steelman, L.C., & Crain, R. L. (2005). The long-term effects of ability grouping in mathematics: A national investigation. *Social Psychology of Education*, *8*(2), 137-177.

Oakes, P. J., Haslam, S. A., Morrison, B., & Grace, D. (1995). Becoming an in-group: Reexamining the impact of familiarity on perceptions of group homogeneity. *Social Psychology Quarterly*, 52-60.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). Yale University Press.

Puntambekar, S. (2022). Distributed scaffolding: scaffolding students in classroom environments. *Educational Psychology Review*, *34*(1), 451-472.

Rivkin, S. G., & Schiman, J. C. (2015). Instruction time, classroom quality, and academic achievement. *The Economic Journal*, *125*(588), F425-F448.

Rytivaara, A. (2011). Flexible grouping as a means for classroom management in a heterogeneous classroom. *European Educational Research Journal*, *10*(1), 118-128.

Salwana, E., Hamid, S., & Yasin, N. M. (2017). Student academic streaming using clustering technique. *Malaysian Journal of Computer Science*, *30*(4), 286-299.

Schullery, N. M., & Schullery, S. E. (2006). Are heterogeneous or homogeneous groups more beneficial to students?. *Journal of Management Education*, *30*(4), 542-556.

Sheridan, K., Puranik, T. G., Mangortey, E., Pinon-Fischer, O. J., Kirby, M., & Mavris, D. N. (2020). An application of DBSCAN clustering for flight anomaly detection during the approach phase. *In AIAA Scitech 2020 Forum* (p. 1851).

Shields, C. M. (1995). A comparison study of student attitudes and perceptions in homogeneous and heterogeneous classrooms. *Roeper Review*, *17*(4), 234-238.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*(1), 43-69. https://doi.org/10.1006/ceps.1996.0004

Taghizabet, A., Tanha, J., Amini, A., & Mohammadzadeh, J. (2023). A semi-supervised clustering approach using labeled data. *Scientia Iranica*, *30*(1), 104-115.

Wang, Q., Wang, F., Ren, F., Li, Z., & Nie, F. (2021). An effective clustering optimization method for unsupervised linear discriminant analysis. *IEEE Transactions on Knowledge and Data Engineering*.

Website, https://sites.google.com/site/bulutfaruk/study-of-clustering-on-education, Received: 28/Feb/2024

Wyman, P. J., & Watson, S. B. (2020). Academic achievement with cooperative learning using homogeneous and heterogeneous groups. *School Science and Mathematics*, *120*(6), 356-363.

Xiao, B., Wang, Z., Liu, Q., & Liu, X. (2018). SMK-means: an improved mini batch k-means algorithm based on mapreduce with big data. *Comput. mater. Continua*, *56*(3), 365-379.

Xiao, C., Hong, S., & Huang, W. (2023). Optimizing graph layout by t-SNE perplexity estimation. *International Journal of Data Science and Analytics*, *15*(2), 159-171.

Xie, Y., Wu, D., & Qiang, Z. (2023). An improved mixture model of gaussian processes and its classification expectation–maximization algorithm. *Mathematics*, *11*(10), 2251.

Yoleri, S. (2013). The effects behavior problems in preschool children have on their school adjustment. *Education*, *134*(2), 218-226.

Yoleri, S. (2014). The effects of age, gender, and temperament traits on school adjustment for preschool children. *e-International Journal of Educational Research*, *5*(2), 54-66.

Yoleri, S. (2015). Preschool children's school adjustment: indicators of behavior problems, gender, and peer victimisation. *Education* 3-13, *43*(6), 630-640.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied artificial intelligence*, *17*(5-6), 375-381.

Zhu, X., Zhang, S., Jin, Z., Zhang, Z., & Xu, Z. (2010). Missing value estimation for mixed-attribute data sets. *IEEE Transactions on Knowledge and Data Engineering*, *23*(1), 110-121.

About the authors:

Faruk Bulut

Faruk Bulut holds the position of Associate Professor in the Computer Engineering Department at Istanbul Esenyurt University. His research interests include Machine Learning, Image Processing, Optimization Methods, Decision Support Systems, and Graph Theory. Throughout his career, Bulut has served in various capacities, including educator, instructor, engineer, project manager, and consultant, contributing to numerous projects within both private and government sectors.

İlknur Dönmez

İlknur Dönmez works as a Senior Researcher at TÜBİTAK Informatics and Information Security Research Center (BİLGEM). She received her doctorate degree from Istanbul Technical University with her thesis titled "Grammatic and Semantic Analysis of Turkish Sentences Based on Predicates". In addition to his academic achievements, Dönmez also has some industrial experience in areas such as Machine Learning, Data Mining, Big Data and Natural Language Processing.

Ibrahim Furkan Ince

Ibrahim Furkan Ince is an Assistant Professor in the Department of Computer Engineering at Nişantaşı University, Istanbul. He received his doctorate from Kyungsung University in Korea, where his thesis, titled “A Study on Gaze Enhanced User Interfaces: Committing Eye Tracking,” explored innovative approaches in user interface design. His academic research focuses on areas such as Image Processing, Computer Vision, Machine Learning, and Human-Computer Interaction.

Pavel Petrov

Pavel Petrov is an Associate Professor in the Department of Informatics at the University of Economics – Varna, Bulgaria. His research delves into Information Systems, Web Technologies, Data Processing, Big Data, Fuzzy Linear Hybrid Automata, and Real-Time Applications, contributing valuable insights into these evolving fields.