International Journal of Advanced and Applied Sciences, 9(9) 2022, Pages: 70-77

Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Centroid competitive learning approach for clustering and mapping the social vulnerability in Morocco

Ilyas Tammouch [1,*], Abdelamine Elouafi [1], Souad Eddarouich [2], Raja Touahni [1]

[1]*Faculty of Science, Telecommunications Systems and Decision Engineering Laboratory, Ibn Tofail University, Kenitra, Morocco*
[2]*Regional Educational Center, Rabat, Morocco*

ARTICLE INFO

Article history:
Received 9 March 2022
Received in revised form
8 June 2022
Accepted 10 June 2022

Keywords:
Student's social vulnerability
Centroid neural network
Choropleth map
Distance
Clustering

ABSTRACT

Over the last three decades, growing inequalities in countries have compounded the issues faced by society's most vulnerable populations. Students are facing the brunt of the effects in particular. A student's social vulnerability emerges as a result of the interaction of a variety of individual and environmental factors that accumulate over time. Poverty, material deprivation, and a lack of parental education can all have an impact on student assessment in school. Previous research has focused on the impact of psychological, cognitive, and physical functioning on children's education, ignoring students' social vulnerability and its impact on educational achievements in developing countries. This paper aims to identify vulnerable regions that need attention and intervention by clustering Moroccan students based on their social vulnerability using an unsupervised competitive learning approach "Centroid neural network," subsequently displaying the results in a choropleth map to visualize the results. For this purpose, we used the PISA dataset which contains the full set of responses from individual students focusing on specific information such as their parent's backgrounds, socioeconomic position, and school conditions. Based on our current findings, we concluded that social vulnerability has a detrimental impact on students' cognitive development. Moreover, the choropleth map shows that 'Beni Mellal-Khenifra' has the highest level of social vulnerability among all regions, besides "Marrakech-Safi" "Eddakhla-Oued Eddahab" and "Guelmim-Oued Noun" all of which have a high level of social vulnerability as well, urging academicians to incorporate resilience building into the design of policies in these regions in order to improve student's educational outcomes.

1. Introduction

Since education matters for human development, including health, economic growth, and democracy, improving education is included in policy agendas in most countries around the world. There is a growing literature describing the correlations between children's educational outcomes and students' well-being for the most part focusing on developed countries. Previous studies have investigated the impact of psychological, cognitive, and physical functioning on children's education and arrive at a similar conclusion, namely, that a student's well-being is a pivotal factor. In this study, we are going to focus on students' social vulnerability.

Arora et al. (2015) defined vulnerability as the relative state of a group of people who are more exposed to risks than their peers. The degrees of child vulnerability may be viewed as a downward spiral, with each loop leading to a point where the kid is more likely to experience a negative outcome as a result of a shock. Young people are also prone to abuse, neglect, deprivation, and violence due to extreme poverty, chronic sickness in themselves or their parents, and a lack of social support and education. Because of the high frequency of extreme poverty and chronic illnesses, it is believed that a significant share of the population lives in poverty. While Schweiger (2019) described students' vulnerability as the consequence of the interaction of a number of individual and environmental factors that fluctuate and evolve throughout time, which expose them to vulnerabilities and hazards. Infants are especially sensitive to financial difficulties since

* Corresponding Author.
Email Address: ilyas.tammouch@uit.ac.ma (I. Tammouch)
https://doi.org/10.21833/ijaas.2022.09.009
Corresponding author's ORCID profile:
https://orcid.org/0000-0003-2752-4413

they are completely dependent on their parents and expect responsive and consistent care. Young children and teenagers are especially vulnerable to familial stress and material deprivation due to the rapid rate of early brain development. Some of the factors that contribute to social vulnerability are as follows: Individual factors that contribute to a child's vulnerability include cognitive, emotional, and physical abilities, as well as personal circumstances, such as disability (children with varying abilities and needs whose individual functioning is limited by physical), mental health issues, and poverty. Background as an immigrant (factors such as parents with lower educational attainment and less financial resources in the household can impact their capacity to succeed in or complete school). Maltreatment (students who have been subjected to abuse or neglect). Environmental factors that contribute to child vulnerability can be observed in both the family and community. Such as Material Deprivation, parents' academic level where Parent's educational level heavily affects children's academic attainment, as children benefit from many opportunities to overcome shortcomings and acquire skills valued by the job market, in which parent's levels of education and income help children succeed notwithstanding the of skills and ability), furthermore, families have structural and internal characteristics that might lead to either risk or protective outcomes for children's development (Walsh, 2016).

A recent study by Martineli et al. (2018) revealed that social vulnerability, low mother educational level, low family income, the existence of more chronic adversity, and present maternal depression are characteristics related to the outcome of behavioral problems in children in family and school contexts.

With the help of the clustering algorithm technique, it is possible to discover the key characteristics that could act as barriers to Student assessment. Shovon and Haque (2012) defined data clustering as an unsupervised statistical data processing technique. It is used to group similar data into a homogeneous group in order to uncover hidden patterns and relationships, allowing for faster and more efficient decision-making. In a nutshell, cluster analysis divides a large set of data into smaller groups called clusters. Each cluster is made up of data objects that are similar to one another but not to objects from other clusters. The aim of applying clustering was to characterize students with difficulties Which can help academicians to enhance the education quality.

The major goal of this study is to investigate and quantify the impact of social vulnerability on cognitive development in students, and subsequently construct a choropleth map that depicts the level of social vulnerability in each region of Morocco using an unsupervised competitive learning algorithm “Centroid neural network” in order to identify vulnerable regions that need attention and intervention. The paper is organized as follows: Section one investigates the impact of social vulnerability on academic attainment, whereas Section two revolves around clustering students based on their social vulnerability. Section Three contains the production of a choropleth map and a discussion of the results.

2. Data and methods

2.1. Data

The Program for International Student Assessment (PISA) assesses 15-year-old students’ capacity to apply their knowledge and skills in reading, mathematics, and science to real-world obstacles in around 80 nations and education systems. PISA 2018 included alternative financial education assessments in the United States (Pholphirul and Teimtad, 2018; Brunello and Rocco, 2013).

We used PISA because it is one of the few open-source empirical data about Morocco students’ educational achievement. Secondly, PISA gathers specific information about each student, such as their parents' backgrounds, socioeconomic position, school conditions, class size, and so on. The depth of the PISA dataset enables us to investigate the relationship between educational achievement and the factors that may impact it. Lastly, unlike in previous years, PISA 2018 participants were requested to complete a questionnaire about their attitudes regarding education and their emotional state.

The impact of social vulnerability on education has received a lot of attention from Moroccan education scholars. This paper uses student-level survey data from PISA 2018 that covers 6814 Moroccan students since it may contribute to the education reform.

2.2. Studying the impact of student’s social vulnerability on their assessment

The initial phase of this research is a preliminary study of the correlation between social vulnerability features and the performance of students in the program of international student assessment (PISA) Dataset which assesses the level of academic performance to 15-year-old school. It consists of three subtests: Reading literacy, mathematical literacy, and scientific literacy, the average of the scores of each subtest results in a total score.

Pearson correlation coefficient measures how strong is the linear association between two continuous variables. Therefore we will use it to analyze the impact of social vulnerability on students is a very simple and effective method, which will make the method of clustering students and creating a map of social vulnerability more valuable.

Al-Shehri et al. (2017) have proposed that the Pearson correlation coefficient is represented as follows:

$$\rho X, Y = \frac{cov(X,Y)}{\sigma X \bullet \sigma Y} = \frac{E(X - \mu X) \bullet (Y - \mu Y)}{\sigma X \bullet \sigma Y} \quad (1)$$

where, $cov(X, Y)$ denotes the covariance of sample vectors X and Y, σX and σY denote the mean of sample vectors X and Y, respectively the correlation analysis is carried out between student's social vulnerability, and the assessment (total score) on the PISA test.

The value of 'P' can be anywhere between '-1' and '+1'. The value '0' indicates that the two variables have no relationship. A positive association between two variables is indicated by a value greater than '0,' suggesting that an increase in the value of one variable raises the value of the other. A negative relationship between two variables is shown by a value less than '0', which means that increasing the value of one decreases the value of the other.

Remark: By analyzing the Pearson correlation coefficient between variables, useful information in variables can be mined more accurately, giving the final clustering results greater Significance.

2.3. Clustering student

2.3.1. Data-preprocessing

Our data-preprocessing involved three steps:

- Step 1. Due to the disparity number of students in different Morocco regions, we performed a representative random sampling (65 students in each region) which has the closest mean to one of the populations in order to compare the results between regions.
- Step 2. Data standardization, because If PCA is applied to our raw dataset, the result for features with high variance will also be high. As a result, principal components will be skewed towards features with high variance, resulting in erroneous results.
- Step 3. Principle Component Analysis (PCA), which is widely used for dimensionality reduction. PCA can create mutually independent principal components (PCs) that contain the great majority of variance information.in our case, we selected PCs with a corresponding variance larger than 0.90. Thus, 2 PCs were chosen for our normalized data.

2.3.2. Determining the k value

The next step was to estimate the number of clusters required for the effective clustering of the data. For this objective, we adopted the Elbow method. The centroid neural network clustering works by defining the clusters in such a way that the total variation within a given cluster is minimal. The total inside cluster sum of squares, denoted by WSS, indicates how compact the cluster is, and our goal, while utilizing the clustering approach, is to make the WSS as minimal as possible. This results in the creation of efficient data clusters.

2.3.3. Centroid neural network algorithm

Clustering is a key method for multivariate data analysis, and it's used in a wide range of scientific, technical, and business applications. For decades, researchers have studied clustering strategies in depth. As a result, several algorithms have been created.

Individually beneficial traits have been offered to overcome many practical problems. Signal processing issues, data mining applications, and marketing are all issues that need to be addressed. Activities include research, image processing, and pattern recognition.

The Centroid Neural Network was proposed by Park (2000) as an unsupervised competitive learning algorithm based on the classical k-means clustering approach. At each presentation of the data vector, it discovers the cluster centroids. When data x is given to the network, the winner neuron at the epoch (k) is the neuron with the shortest distance to x. The neuron that was the loser at the epoch (k) is the neuron that was the winner at the epoch (k-1) but is not the winner at the epoch (k). The CNN only changes its weights when the state of the output neuron for the current data differs from the status of the previous epoch.

In CNN, when an input vector x is presented to the network at iteration n, Park (2009) has determined that the weight update equations for winning neuron j and loser neuron i may be written as follows:

$$W_j(n+1) = W_j(n) + \frac{1}{N_j+1}[X(n) - W_j(n)] \quad (2)$$

$$W_i(n+1) = W_i(n) + \frac{1}{N_i+1}[X(n) - W_i(n)] \quad (3)$$

where, wj(n) and wi(n) are the weight vectors of the winning and loser neurons, respectively, at iteration n.

The CNN algorithm offers numerous advantages over traditional algorithms like SOM or k-means, when utilized for clustering and unsupervised competitive learning, the method comes in handy as it does not require a specified learning gain schedule or the total number of iterations for clustering. It always converges to sub-optimal solutions, whereas traditional methods like SOM might provide inconsistent outcomes depending on the initial learning gains and the total number of iterations.

In order to improve the stability of clustering, the Centroid Neural Network (CNN) algorithm introduces the concept of reward and punishment to the winner and loser neurons Furthermore, in order to get a solution that is closer to the optimal clustering outcome, CNN starts with 2 groups and gradually raises the number of groups until the needed number of groups is obtained. in most studies CNN outperformed traditional algorithms such as SOM and Differential Competitive Learning. Variations of CNN have been proposed for different problems and applied to various areas.

In our study CNN algorithm is used to cluster students based on social vulnerability features which are correlated with student assessment into two clusters, one describes students with high social vulnerability levels, and the other describes students with low vulnerability levels.

2.4. Model performance metrics

2.4.1. Silhouette coefficient (SC)

A good cluster has a very tiny gap between samples of the same category and a very large distance between samples of different categories. Rousseeuw (1987) proposed the silhouette coefficient (SC) which can evaluate both characteristics at the same time. A model with a higher silhouette coefficient score offers better clusters.

The silhouette coefficient s for a single sample is given as:

$$\mathrm{SC} = \frac{b-a}{\max(a,b)} \quad (4)$$

where, a denotes the mean distance between a sample and all other points in the same class and b denotes the mean distance between a sample and all other points in the next closest cluster. The mean of the silhouette coefficients for each sample is used to get the silhouette coefficient for a set of samples.

2.4.2. Calinski–Harabasz index (CHI)

Kozak (2012) has proposed that a model with a higher Calinski–Harabasz score has better clusters. Where the Calinski–Harabasz score s is defined as the ratio of the between-clusters dispersion mean and the within-cluster dispersion means for a collection of data E of size n_E that has been clustered into k clusters.

$$\mathrm{CHI} = \frac{\mathrm{tr(Bk)}}{\mathrm{tr(Wk)}} \times \frac{n_E - k}{k-1} \quad (5)$$

where, $tr(Bk)$ is the trace of the dispersion matrix between clusters and $tr(Wk)$ is the trace of the dispersion matrix within the cluster defined by:

$$W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - C_q)(x - C_q)^T \quad (6)$$

$$B_k = \sum_{q=1}^{k} n_q (C_q - C_E)(C_q - C_E)^T \quad (7)$$

where, C_q represents the set of points in cluster q, C_q represents the center of cluster q, C_E represents the center of E, and n_q represents the number of points in cluster q.

2.4.3. Davies–Bouldin index (DBI)

Davies and Bouldin (1979) suggested that a lower Davies–Bouldin index corresponds to a model with higher cluster separation.

Where the index is defined as the average similarity between each cluster Ci for i=1, ..., k and its closest neighbor Cj. In the context of this index, the similarity is defined as a Rij trade-off metric.

The Davies–Bouldin index is calculated as follows:

$$\mathrm{DBI} = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} \mathrm{Rij} \quad (8)$$

2.4.4. Separation (S)

A good clustering in our study indicates that the observations from two clusters should be far from each other. This means we want the distance between the mean of observations from cluster 0 corresponding to students with high social vulnerability and cluster 1 corresponding to students with low social vulnerability to be as high as possible.

For this purpose, we propose S “the separation between cluster 1 and cluster 0.” It is calculated according to the following formula:

$$S = \frac{1}{n} \sum_{i=1}^{n} |(x_n - y_n)| \quad (9)$$

where, n is the number of features that have an impact on academic progress; x_n is mean of feature number n corresponding to students assigned to cluster 1; y_n is the mean of feature number n corresponding to students assigned to cluster 0.

2.5. Social vulnerability map

After clustering students into two clusters. We group the clustered data by region, then we count the number of students assigned to cluster 0 (high social vulnerability level) and the number of students assigned to cluster 1 (low social vulnerability level) for each region. Afterward, we calculate the difference between the two counts on a scale of -1 to 1, D” the difference” can be calculated by following this process:

1. Subtract the two numbers to get the difference. (The first value refers to the number of students in cluster 1, whereas the second value refers to the number of students in cluster 0).
2. Calculate the average of the two values.
3. Divide the difference calculated in step 1 value by the average value calculated in step 2.

For this purpose, we propose D “the difference between the student assigned to cluster 1 compared to student assigned to cluster 0.” It is calculated according to the following formula:

$$D = \frac{c1 - c0}{\left[\frac{c1 + c0}{2}\right]} \quad (10)$$

where, C1 is number of students assigned to cluster 1 and C0 is number of students assigned to cluster 0.

A difference greater than 0 indicates that the region has a low level of social vulnerability (most students were assigned to cluster 1). While a difference less than 0 indicates that this region has a significant level of social vulnerability (most students were assigned to cluster 0).

Finally, we proceed to map the results using a choropleth map, Stewart and Kennelly (2010) elucidated that a choropleth map is used to show statistical variation among map enumeration units, Subsequently, we utilized it to depict all 12 Moroccan regions together with their level of social vulnerability (the difference calculated between cluster 0 and cluster 1 for each region).

3. Results and discussion

3.1. Studying the impact of student's social vulnerability on their assessment

According to the findings in Table 1, we can see that some social vulnerability features have a substantial impact on academic progress. ESCS (Index of economic, social, and cultural status) has the highest correlation with the academic assessment with a significant p-value. Furthermore, students from wealthier homes performed much higher with reference to their total scores. In terms of resources, having more ICT (Information and Communication Technologies resources) results in pupils who have greater levels of cognitive skills. Respectively, home educational resources have a beneficial impact on student academic ability.

3.2. Clustering student

3.2.1. Data preprocessing

After normalizing the data, we selected PCs with a corresponding variance larger than 0.90. Thus, 2 PCs were chosen for our normalized data.

3.2.2. Estimating the number of clusters

After plotting the graph (Fig. 1), we decided on 2 as the cut-off threshold since, while the WSS continues to fall, it does not appear to be at a substantial enough rate to accept the added complexity of more clusters.

Table 1: Impact of social vulnerability on student's academic achievement using Pearson correlation

Variable	Label	Pisa Score
ESCS	Index of economic, social and cultural status	0.34**
ICTRES	ICT resources	0.33**
HEDRES	Home educational resources	0.30**
WEALTH	Family wealth	0.30**
EMOSUPS	Parents' emotional support	0.26**
BELONG	Sense of belonging to school	0.21**
ICTHOME	ICT available at home	0.16*
CULTPOSS	Cultural possessions at home	0.09*
DISCRIM	Discriminating school climate	-0.22**

Note: * refers to p-value where: *p<0.1, **p<0.05

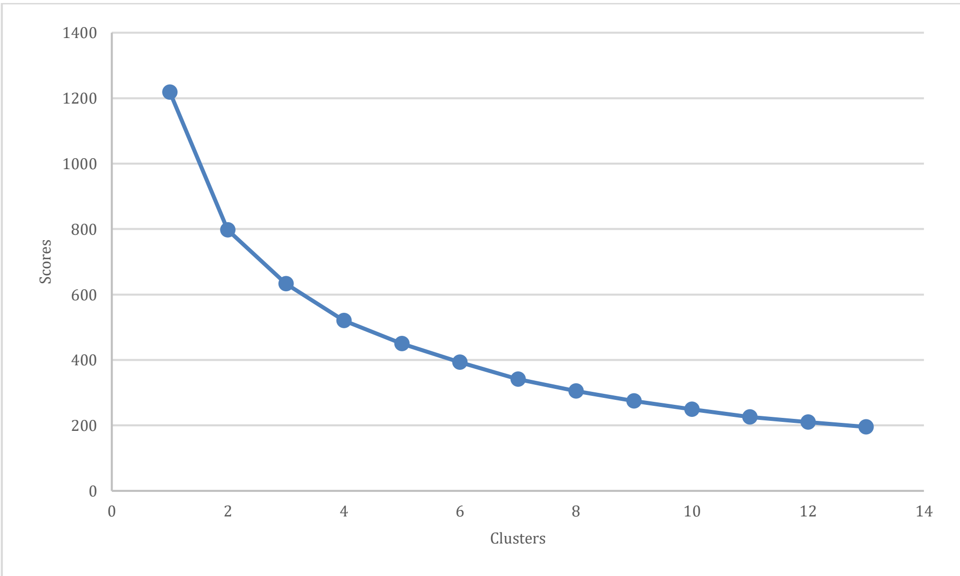

Fig. 1: Estimating the number of clusters using the elbow method

3.2.3. Clustering students

The number of clusters employed in this study was 2, as predetermined using the Elbow method. The 2 PCs are then fed as an input to the Centroid neural network, allowing us to cluster students into groups based on social vulnerability factors that have an impact on academic attainment. We want to identify which students are disfavored and put them together. This data can be used to separate all students into different groups in order to implement policies for further educational reform.

According to the radar chart displayed in Fig. 2 comparing the two clusters, it can be seen that students with low social vulnerability lie in cluster 1 (invulnerable students), while students with high social vulnerability lie in cluster 0 (vulnerable students).

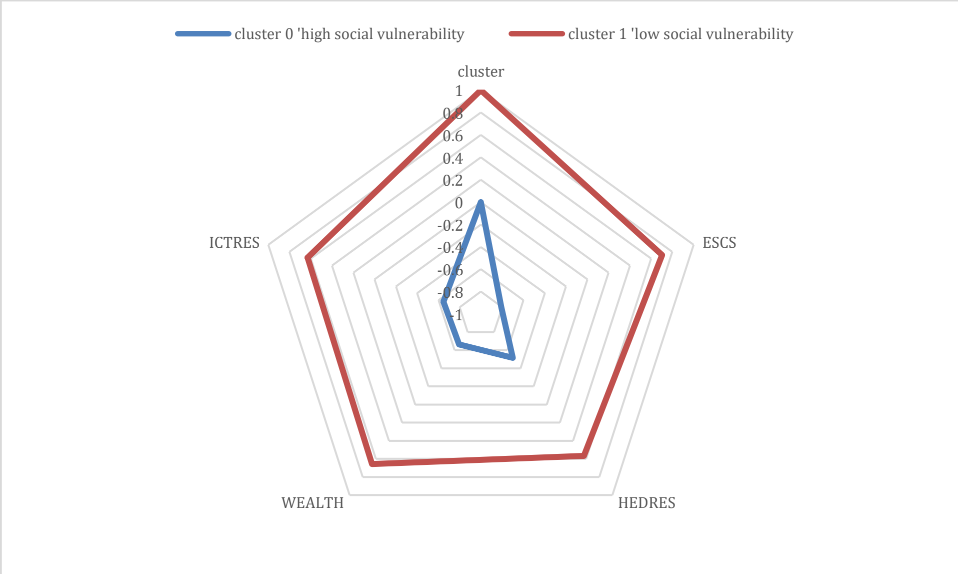

Fig. 2: Radar chart of clustering data using centroid neural network

3.3. Evaluation and comparison of three algorithms

According to results discussed in section 3.1, where we found 4 social vulnerability factors that have a substantial impact on academic achievement (ESCS, ICTRES, HEDRES, WEALTH), the Eq. 9 of separation can be written as follows:

$$S = \frac{[(\overline{E1}-\overline{E0})+(\overline{H1}-\overline{H0})+(\overline{W1}-\overline{W0})+(\overline{I1}-\overline{I0})]}{4} \quad (11)$$

where,

$\overline{E}1$: mean of ESCS corresponding to students assigned to cluster 1.
$\overline{E}0$: mean of ESCS corresponding to students assigned to cluster 0.
$\overline{H}1$: mean of HEDRES corresponding to students assigned to cluster 1.
$\overline{H}0$: mean of HEDRES corresponding to students assigned to cluster 0.
$\overline{W}1$: mean of WEALTH corresponding to students assigned to cluster 1.
$\overline{W}0$: mean of WEALTH corresponding to students assigned to cluster 0.
$\overline{I}1$: mean of ICTRES corresponding to students assigned to cluster 1.
$\overline{I}0$: mean of ICTRES corresponding to students assigned to cluster 0.

After measuring and comparing the effects of the three algorithms on the behavior of clustering students by calculating SC, CHI, and DBI, it can be seen from Table 2 which exhibits the comparative results of each model under various evaluation criteria. that the SC, CHI, and DBI of the Centroid neural network where higher than those of GMM and Spectral clustering, which shows that the centroid neural network could better gather students with similar social vulnerability traits.

Using the formula of separation illustrated in Eq. 11, Table 3 reveals the separation between clusters 0 and 1 of the three algorithms.

Furthermore, according to the S (separation) value shown in Table 3, we can see that the separation between cluster 0 (vulnerable student) and cluster 1 (invulnerable students) was slightly higher than the S value of GMM and spectral clustering which confirm that the clustering effect of Centroid neural network was the best and could better cluster the student based on their social vulnerability.

Table 2: Evaluation and comparison of three algorithms using silhouette coefficient (SC), Calinski–Harabasz Index (CHI), and Davies–Bouldin Index (DBI)

Algorithm	SC	CHI	DBI
Centroid Neural Network	0.47	1019.478	0.752390
GMM	0.46	994.781	0.750725
Spectral Clustering	0.47	1018.91	0.751250

Table 3: Separation between clusters 0 and 1

Algorithm	S
Centroid neural network	-1.298
GMM	-1.301
Spectral Clustering	-1.302

3.4. Social vulnerability map

The creation of a social vulnerability map is broken into the following steps:

- Step 1: For every region, we count the number of students assigned to cluster 0 (vulnerable students) and the number of students affected to cluster 1 (invulnerable students) as shown in Table 4. Note that each region has 65 students.

Table 4: Number of students assigned to cluster 0 and students affected to cluster 1

region	Number of vulnerable students (cluster 0)	Number of invulnerable students (cluster 1)
Tanger-Tetouan-Al Hoceima	32	33
Oriental	32	33
Fès-Meknès	31	34
Rabat-Salé-Kénitra	26	39
Béni Mellal-Khénifra	42	23
Casablanca-Settat	17	48
Marrakech-Safi	38	27
Drâa-Tafilalet	30	35
Souss-Massa	31	34
Guelmim-Oued Noun	34	31
Laayoune-Sakia El Hamra	21	44
Eddakhla-Oued Eddahab	39	26

- Step 2: We calculate D "the difference between the number of students in each cluster for every region on a scale from-1 to 1" as tabulated in Table 5 using the formula shown in Eq. 10.

Table 5: Difference between the number of students in each cluster

region	D
Tanger-Tetouan-Al Hoceima	0.03
Oriental	0.03
Fès-Meknès	0.09
Rabat-Salé-Kénitra	0.40
Béni Mellal-Khénifra	-0.58
Casablanca-Settat	0.95
Marrakech-Safi	-0.34
Drâa-Tafilalet	0.15
Souss-Massa	0.09
Guelmim-Oued Noun	-0.09
Laayoune-Sakia El Hamra	0.71
Eddakhla-Oued Eddahab	-0.40

- Step 3: The last step is to create a choropleth map displaying the social vulnerability level of all 12 Moroccan regions using the difference "D" calculated in step 2. The results displayed in Fig. 3, indicate that 'Casablanca-settat' has the lowest social vulnerability level while 'Beni Mellal-Khenifra' has the highest social vulnerability level among all regions, which will impact the overall educational achievement. It should be noted that even regions closer to a social vulnerability level of 0 must be taken into account for further education reform.

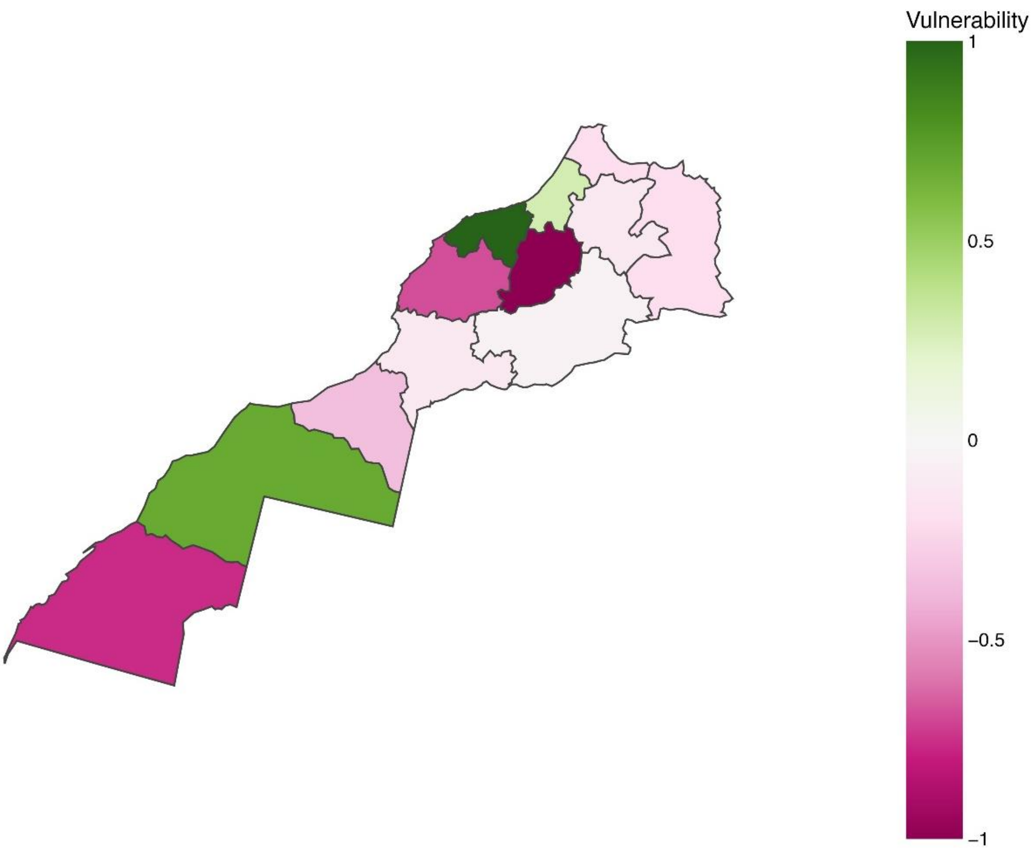

Fig. 3: Social vulnerability map

4. Conclusion

The study aims to examine and quantify the impact of social vulnerability on children's cognitive development. And clustering them based on social vulnerability features that have an impact on their achievement. The results reveal that the Index of economic, social, and cultural status, ICT resources, Home educational resources and family wealth has a detrimental impact on students' cognitive development. Furthermore, clustering students according to features that have an impact on student assessment in order to build a map illustrating regions that are vulnerable throughout the 12 Moroccan regions. For instance 'Beni Mellal-Khenifra' has the highest social vulnerability level among all regions, so academicians may seek to integrate resilience-building into the design of policies in those regions such as empowering vulnerable families by parenting support, lowering segregation between students by addressing the practical barriers to accessing certain schools such as tuition costs and availability of public transport, increasing parental employment and the quality of jobs in order to improve student's educational outcomes.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Al-Shehri H, Al-Qarni A, Al-Saati L, Batoaq A, Badukhen H, Alrashed S, and Olatunji SO (2017). Student performance prediction using support vector machine and k-nearest neighbor. In the 30th Canadian Conference on Electrical and Computer Engineering, IEEE, Ontario, Canada: 1-4. https://doi.org/10.1109/CCECE.2017.7946847

Arora SK, Shah D, Chaturvedi S, and Gupta P (2015). Defining and measuring vulnerability in young people. Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive and Social Medicine, 40(3): 193-197. https://doi.org/10.4103/0970-0218.158868 **PMid:26170545 PMCid:PMC4478662**

Brunello G and Rocco L (2013). The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores. Economics of Education Review, 32: 234-246. https://doi.org/10.1016/j.econedurev.2012.10.006

Davies DL and Bouldin DW (1979). A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2): 224-227. https://doi.org/10.1109/TPAMI.1979.4766909

Kozak M (2012). "A dendrite method for cluster analysis" by Caliński and Harabasz: A classical work that is far too often incorrectly cited. Communications in Statistics-Theory and Methods, 41(12): 2279-2280. https://doi.org/10.1080/03610926.2011.560741

Martineli AKB, Pizeta FA, and Loureiro SR (2018). Behavioral problems of school children: Impact of social vulnerability, chronic adversity, and maternal depression. Psicologia: Reflexão e Crítica, 31: 11. https://doi.org/10.1186/s41155-018-0089-9 **PMid:32026068 PMCid:PMC6967022**

Park DC (2000). Centroid neural network for unsupervised competitive learning. IEEE Transactions on Neural Networks, 11(2): 520-528. https://doi.org/10.1109/72.839021 **PMid:18249781**

Park DC (2009). Centroid neural network with weighted features. Journal of Circuits, Systems, and Computers, 18(08): 1353-1367. https://doi.org/10.1142/S021812660900571X

Pholphirul P and Teimtad S (2018). Living with parents and educational outcomes in developing countries: Empirical evidence from PISA Thailand. Journal of Population Research, 35(1): 87-105. https://doi.org/10.1007/s12546-017-9196-1

Rousseeuw PJ (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20: 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Schweiger G (2019). Ethics, poverty and children's vulnerability. Ethics and Social Welfare, 13(3): 288-301. https://doi.org/10.1080/17496535.2019.1593480 **PMid:31839805 PMCid:PMC6884154**

Shovon HI and Haque M (2012). An approach of improving students academic performance by using k means clustering algorithm and decision tree. International Journal of Advanced Computer Science and Applications, 3: 146-149. https://doi.org/10.14569/IJACSA.2012.030824

Stewart J and Kennelly PJ (2010). Illuminated choropleth maps. Annals of the Association of American Geographers, 100(3): 513-534. https://doi.org/10.1080/00045608.2010.485449

Walsh F (2016). Processos normativos da família: Diversidade e complexidade. Artmed Editora Publisher, Porto Alegre, Brazil.