Comparison of K-means and DBSCAN For Prediction Determination of Down Syndrome Using Prenatal Test Data

Rulla Selfiana [1, a)] and Endah Sudarmilah [1, b)] and Devi Afriyantari Puspa Putri[1,c)]

Author Affiliations
Informatics Engineering Department, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

Author Emails
[a)] Corresponding author: l202173005@student.ums.ac.id
[b)] Endah.Sudarmilah@ums.ac.id
[c)] deviapputri@ums.ac.id

Abstract. Down syndrome is a condition in which a person has an excess of chromosomes. Babies are born with 46 chromosomes but Down syndrome babies have an extra copy of one of these chromosomes, chromosome 21. To predict a baby with Down syndrome, pregnant women can perform a series of prenatal tests. Clustering for the data in here is used to solve the problem that often occurs when determining whether Down syndrome occurs to group it into high or low risk because prenatal test data are invasive and most of the data will be immediately merged into one, therefore with accurate clustering predictions are needed so that patients can take the next step for the next test. In this study, we need k-means clustering and DBSCAN to compare which of the two unsupervised algorithms can predict the best clustering. The steps taken by involving prenatal test data attributes which will then be applied to the code of the two algorithms afterward are looking at the cluster level and grouping distance. The result required is a clustering without noise that has a small distance between the clusters that are carried out. From the results obtained, k-means that clustering is better than DBSCAN where the prediction level of clustering is better for determining Down syndrome even with a slightly high noise level which can only be removed by DBSCAN.

Keywords: K-means, DBSCAN, Down Syndrome, Unsupervised Learning

INTRODUCTION

Down syndrome is a condition in which a person has an extra chromosome. Chromosomes are small “packages” of genes in the body. They determine how a baby’s body forms and functions as it grows during pregnancy and after birth. Typically, a baby is born with 46 chromosomes. Babies with Down syndrome have an extra copy of one of these chromosomes, chromosome 21. A medical term for having an extra copy of a chromosome is ‘trisomy.’ Down syndrome is also referred to as Trisomy 21. This extra copy changes how the baby’s body and brain develop, which can cause both mental and physical challenges for the baby. the physical of a baby have syndrome down are have a flat nose, large tongue, broad forehead, small ears, and mild to moderate mental retardation. The individual is also susceptible to heart and autoimmune diseases and has a short life expectancy rate. Down Syndrome is closely related to mental retardation, which is a brain development disorder characterized by IQ scores below the average normal person and poor ability to perform daily skills. People with Down syndrome experience many growth disorders, have an IQ of 25-75 (mean <40) it can be said that all children with Down syndrome have different mental retardation (1).

There are two categories of tests that can be performed prenatally, namely screening tests and diagnostic tests. Prenatal examination estimates the possibility of the fetus having Down syndrome. This test cannot provide certainty but it provides a definite diagnosis with near-perfect accuracy. The dual marker test is done in the first trimester of pregnancy. This test is a blood test that quantifies the levels of two hormones B-HCG (serum-free Beta-Human Chorionic Gonadotropin) and PAPP-A (Pregnancy-associated plasma protein A). This technique is non-invasive in nature. Its reports are usually combined with Maternal Age, NT (Nuchal Translucency), and CRL(2).

Mastering unsupervised learning is sometimes not the right choice in machine learning, but unsupervised learning is really needed to discover the inherent structure in learning that uses unlabeled or unclassified data. In unsupervised learning, the inputs are segregated based on features and the prediction is based on which cluster it belongs to. And in machine learning, It is easier to get unlabeled data from a computer than labeled data, which needs manual intervention. Clustering is a way to group a set of data points in a way that similar data points are grouped together. Therefore, clustering algorithms look for similarities or dissimilarities among data points. Clustering is an unsupervised learning method so there is no label associated with data points. The algorithm tries to find the underlying structure of the data. K-means for input to the algorithm for predictive analysis, it stands for the number of groupings that the algorithm must extract from a dataset, expressed algebraically as a k. A k-means algorithm divides a given dataset into k clusters. K-means for input to the algorithm for predictive analysis, it stands for the number of groupings that the algorithm must extract from a dataset, expressed algebraically as a k, performance is usually not as competitive as those of the other sophisticated clustering techniques because slight variations in the data could lead to high variance. Furthermore, clusters are assumed to be spherical and evenly sized, something which may reduce the accuracy of the k-means clustering Python results. And Density-based spatial clustering of applications with noise, or DBSCAN, is a popular clustering algorithm used as a replacement for k-means in predictive analytics. To run it doesn't require input for the number of clusters but it does need to tune two other parameters.

Hence this report will present the results of the prediction the down syndrome using k-means and DBSCAN clustering using prenatal data test and compare two algorithm which the best for prediction down syndrome. After the first prenatal test prediction patient needs to take further tests or not. This method is used so that patients with low probabilities do not need to continue with expensive invasive amniocentesis and CVS tests. Amniocentesis is performed by taking a sample of amniotic fluid which is then tested to analyze the chromosomes of the fetus. In the second trimester (weeks 14-20 of gestation), this method is the most commonly used invasive technique because it is safer and easier (compared to amniocentesis in the first trimester and CVS), reliable, and accurate from a cytogenetic point of view and relatively inexpensive. than other screening methods. Complications of amniocentesis ranged from 0.5-2.2%. Amniocentesis and CVS are quite reliable but provide a miscarriage risk of about 0.5-1%. CVS is performed by taking a sample of cells from the placenta. The sample will be tested for fetal chromosomes. This technique is performed at the ninth to fourteenth week of pregnancy. This theoretical advantage of cfDNA testing prior to any invasive testing over immediate use of amniocentesis or CVS for karyotyping has not been tested in a randomized trial and should be balanced against the lower sensitivity and specificity of cfDNA in the detection of trisomy 21(3).

LITERATUR REVIEW

Machine learning algorithms pose an adaptive alternative to develop better risk assessment models using the existing clinical variables. Data analysis proposes an alternative mathematical modification to improve existing prenatal testing of Down syndrome at the first trimester, both in screening accuracy and also lowering the overall costs of the screening program(4). So that predict in the biological can using clustering, it can be used to derive plant and animal taxonomies, categorize genes with similar functionality, and gain insight into structures inherent in populations(5). In the last research by Tianfu et al assess visual parameterization was less time-consuming than those extending DBSCAN with statistical indicators. Since clustering results should meet different needs in real-life applications, the proposed visualization allowed users to obtain suitable clustering results graphically(6). According to Neocleous et al assessed that computational methods can be used to determine the risk of aneuploidy (addition of all or part of a chromosome), euploidy (acquisition of one or more complete sets of chromosomes), and other chromosomal abnormalities. The database used consisted of 51,208 singleton pregnancies undergoing first-trimester screening. Attributes that can be used as a reference in helping to conclude the risk of aneuploidy include Fetal CRL in mm, serum-free B-HGG in MoM, PAPP-A in MoM, Nuchal Translucency (CT) in mm, Nasal Bone (present or absent), tricuspid flow and ductus venous flow of the mother's current year's pregnancy and previous loss of Down syndrome. The author uses a computational intelligence approach for risk estimation. Data sources are separated into training and test sets. The training set is used to build various support vector machines, ANN, and the K-nearest neighbor model (K-NN). The test set validates the model. The best accuracy is obtained by ANN which correctly identifies all cases of Down syndrome, i.e. no false negatives are generated. The unique false-negative rate was 96.1% assigned to other euploidies(7).

Uzun et al. described predictive markers for Down syndrome which included maternal age, genetic history of Trisomy 21, biochemical markers B-HCG, PAPP-A, nuchal penetration test, and others. The data collection for Down syndrome obtained from Trakya University and George Washington University. The data set is small and therefore suffers from unequal distribution issues. This problem is solved by using oversampling and under-sampling techniques as well as threshold optimization techniques. Bayesian network and Naïve Bayes algorithm are used for predictive modeling of Down syndrome as a probabilistic classifier which can then be accepted as a trisomy 21 classifier(8). Research from Irving Cordova there some theory or DBSCAN that One important drawback of distributed algorithms is that, as their parallelism increases, so does the cost of communication between the computing nodes. (9). By learning from datasets, machine learning can find more interactions between variables and outcomes than conventional statistical methods. The six machine learning methods are decision tree, Naive Bayes, random forest, support vector machine, artificial neural network, and deep neural network (deep learning). most suitable machine learning. In this paper, we assess the random forest model for its potential to improve DS risk assessment. The results show that the Machine Learning method is comparable to the standard method, and can be proposed as an alternative model for the prediction of DS in the second-trimester antenatal screening(10). Silvana et al in a study that mentions using naïve Bayes as an expert system implementation with a sample of data sources that have been obtained, there are 25 selected data sources which are divided into two groups as Prior data and testing data. Based on the tests conducted on experts and non-experts, it was found that this system follows the logic of thinking in general. The result is a higher level of expert data accuracy(11).According to Qin, et al in their research which aims to form a convolutional neural network model that maps unlimited 2D RGB images as inputs and outputs that may have Down syndrome. The network is also compared with advanced methods to describe its performance. The experimental results showed that Down's syndrome was detected with 95.87% accuracy and 93.18% recall. The results show that our method has great potential to support automatic identification of Down syndrome(12). Clustering, a fundamental activity in unsupervised learning, is notoriously difficult when the feature space is high-dimensional. Fortunately, in many realistic scenarios, only a handful of features may be relevant in distinguishing clusters. This has motivated the development of sparse clustering techniques that typically rely on k-means within outer algorithms of high computational complexity. Current techniques also require careful tuning of shrinkage parameters, further limiting their scalability this research raised the case study of down syndrome in the literature(13). The segmentation output using the k-means clustering algorithm filter can be incorporated with the k-means clustering algorithm to get an accurate segmentation result while suppressing noise. K-means clustering segmentation is not over-segmented. It gives better accuracy and considerable processing speed compared with the existing methods(14).

METHOD

In this paper method comparison, k-means between DBSCAN for prediction a down syndrome using prenatal test data with clustering them. DBSCAN clustering is applied then it follows the same principal of clustered the new incremented data just like incremental k-means clustering except that it is able to handle the noisy data or outliers properly(15). so in the research will be using method shown in Figure 1.

FIGURE 1. Research method flowchart

Problem Identification

The research stage starts from the problem identification process which is carried out by conducting research on Down syndrome and how accurate the data from prenatal tests is to say someone has Down syndrome. In this case, problem identification is needed to search for literature studies related to this topic, the problems raised by the author. In this section search about what is problem do have in a medical world for find what they need.

Literature Study

In the literature study is to search for the k-means and DBSCAN methods that will be used as data predictions for children with Down syndrome by using clustering. As well as in the literature study, the author looks for several reference sources for previous research that has been done regarding Down syndrome. Some related things about k-means and DBSCAN will also be searched in this step. understand both algorithms to be used in a prediction with clustering.

Data Collection

The data that will be used are patient data from prenatal tests and the data is data that has previously been classified and filtered so that only a few data parameters from a prenatal data series are used such as Maternal Age, CRL (mm), NT (mm), PAPP-A ((MoM) and Free -hCG (MoM) as an attribute on clustering. The dataset of patients from the patient prenatal test in down syndrome in Kaggle. The data will be collected first and they will do reprocessing data forget the real data with the attribute for implementation in the k-means and DBSCAN code

Code Implementation

K-Means clustering intends to partition n objects into k clusters in which each object belongs to the cluster with the nearest mean. This method produces exactly k different clusters of the greatest possible distinction. The best number of clusters K leading to the greatest separation (distance) is not known as a priori and must be computed from the data. The objective of k-Means clustering is to minimize total intra-cluster variance, or, the squared error function:

$$J = \sum_{j=1}^{K} \sum_{i=1}^{n} ||xi^{(j)} - cj||^2$$

TABLE 1. Symbol used in the paper

Symbol	Explanation
J	Objective function
K	Number of cluster
n	Number of cases
xi	Case i
cj	Centroid for cluster j

We will using the detailed in understanding at a glance how to apply the code we used in k-means clustering, by providing steps for each action in the algorithm code and it is called pseudocode and for the step that is Decide the number of clusters, this number is called K and the number of clusters is equal to the number of centroids. Based on the value of K, generate the coordinates for K random centroids. For every point, calculate the Euclidean distance between the point and each of the centroids. Assign the point to its nearest centroid. The points assigned to the same centroid form a cluster. Once clusters are formed, calculate a new centroid for each cluster by taking the cluster mean. Cluster mean is the mean of the x and y coordinates of all the points belonging to the cluster. Repeat steps 2, 3, and 4 until the centroids cannot move any further. In other words, repeat these steps until convergence. The approach k-means follows to solve the problem is called Expectation-Maximization. Pseudocode in the image shown in Figure 2 will be useful as an illustration of the steps taken too.

```
## K-Means Clustering 1. Choose the number of clusters(K) and
obtain the data points
2. Place the centroids c_1, c_2, ..... c_k randomly
3. Repeat steps 4 and 5 until convergence or until the end of
a fixed number of iterations
4. for each data point x_i:
       - find the nearest centroid(c_1, c_2 .. c_k)
       - assign the point to that cluster
5. for each cluster j = 1..k
       - new centroid = mean of all points assigned to that
cluster
6. End
```

Figure 2. Pseudocode of K-means

Clustering from DBSCAN provides more benefits in grouping data with a density even though there is noise, the two have different ways of doing machine learning algorithms.

DBSCAN does it with initial steps explained with pseudocode shown in figure 3 will explained the step like finding all the neighbor points within eps and identifying the core points or visiting with more than MinPts neighbors. For each core point if it is not already assigned to a cluster, create a new cluster. Find recursively all its density connected points and assign them to the same cluster as the core point. A point a and b are said to be density connected if there exists a point c which has a sufficient number of points in its neighbors and both the points a and b are within the eps distance. This is a chaining process. So, if b is neighbor of c, c is neighbor of d, d is neighbor of e, which in turn is neighbor of a implies that b is neighbor of a. Iterate through the remaining unvisited points in the dataset. Those points that do not belong to any cluster are noise.

```
DBSCAN(DB, distFunc, eps, minPts) {
    C := 0                                                  /* Cluster counter */
    for each point P in database DB {
        if label(P) ≠ undefined then continue               /* Previously processed in inner loop */
        Neighbors N := RangeQuery(DB, distFunc, P, eps)     /* Find neighbors */
        if |N| < minPts then {                              /* Density check */
            label(P) := Noise                               /* Label as Noise */
            continue
        }
        C := C + 1                                          /* next cluster label */
        label(P) := C                                       /* Label initial point */
        SeedSet S := N \ {P}                                /* Neighbors to expand */
        for each point Q in S {                             /* Process every seed point Q */
            if label(Q) = Noise then label(Q) := C          /* Change Noise to border point */
            if label(Q) ≠ undefined then continue           /* Previously processed (e.g., border point) */
            label(Q) := C                                   /* Label neighbor */
            Neighbors N := RangeQuery(DB, distFunc, Q, eps) /* Find neighbors */
            if |N| ≥ minPts then {                          /* Density check (if Q is a core point) */
                S := S ∪ N                                  /* Add new neighbors to seed set */
            }
        }
    }
}
```

Figure 3. Pseudocode of DBSCAN

Testing and Result

In testing and results is to test whether the code used works well. The nine attributes used as markers that could help deduce the risk: Fetal CRL in mm, serum-free B-HCG in MoM, PAPP-A in MoM (Method of Medians), Nuchal Translucency in mm, maternal age in years and previous pregnancy Down's Syndrome History. The total prenatal test data we using for the testing is 27 data and in one data there are already five attributes. The testing will combination using of python for data clustering, k-means and DBSCAN used will produce Down syndrome predictions from prenatal tests. In this case the results that will be displayed are in the form of a cluster of attributes from the data that has been processed. K-means and DBSCAN work to see if Down syndrome can occur in groups of each attribute. The attribute prenatal test data shown in Table 2.

Table 2. Prenatal Test data used in paper

MA (years)	CRL (mm)	NT (mm)	PAPP-A (MoM)	Free β-hCG (MoM)
39	74.9	7.9	0.51	1.34
36	66.0	2.4	0.22	6.18
32	68.0	3.4	0.71	5.16
32	68.2	3.0	0.48	3.03
37	65.0	1.9	0.41	4.85
25	54.0	5.8	0.14	1.79
23	77.3	4.7	0.29	1.43
35	59.0	1.9	0.23	1.86
29	47.0	4.0	2.56	1.91

33	61.8	1.2	0.56	1.10
42	45.0	4.1	0.22	0.26
36	61.9	2.0	0.07	0.09
24	56.4	3.5	0.15	0.35
42	57.0	1.5	0.39	0.12
23	47.0	1.2	0.54	0.52
34	56.5	8.0	0.52	0.42
31	51.9	1.9	0.35	0.12
21	60.9	3.9	0.56	0.10
23	57.0	1.5	0.54	0.54
30	45.0	0.56	0.56	0.35
36	45.9	1.2	0.35	0.54
32	68.0	3.4	0.15	6.84
32	68.2	0.71	5.16	3.03
21	60.8	3.4	0.57	0.14
27	45.0	3.5	0.55	0.10
30	45.0	0.57	0.39	0.10

The prenatal test data above is used in the form of a csv file which will then be imported and read the data. In this case, clustering takes data from each patient and each attribute to make it a determination whether a person is at risk for Down syndrome.

RESULT AND DISCUSSION

The purpose of this research where we do a comparison of two algorithms. These two algorithms are each tested with the same number of datasets. This study uses a dataset from prenatal test of patients who are not use labeled input and output data. Predict the Down Syndrome from the dataset we use unsupervised learning to cluster with choosing the attribute for the clustering. The formula k-means can be reduced to a constant multiplied by $(n*n)$. The final resulting function since the constant will be removed is ($n2$)(16). Cluster analysis could be divided into hierarchical clustering and non-hierarchical clustering techniques(17). And this research include on hierarchal clustering. The endpoint is a set of clusters, where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other. DBSCAN spatial clustering algorithm to find optimal cluster. The clusters are validated and verified using precision and recall metrics(18). Dataset cluster can be seen in Figure 4.

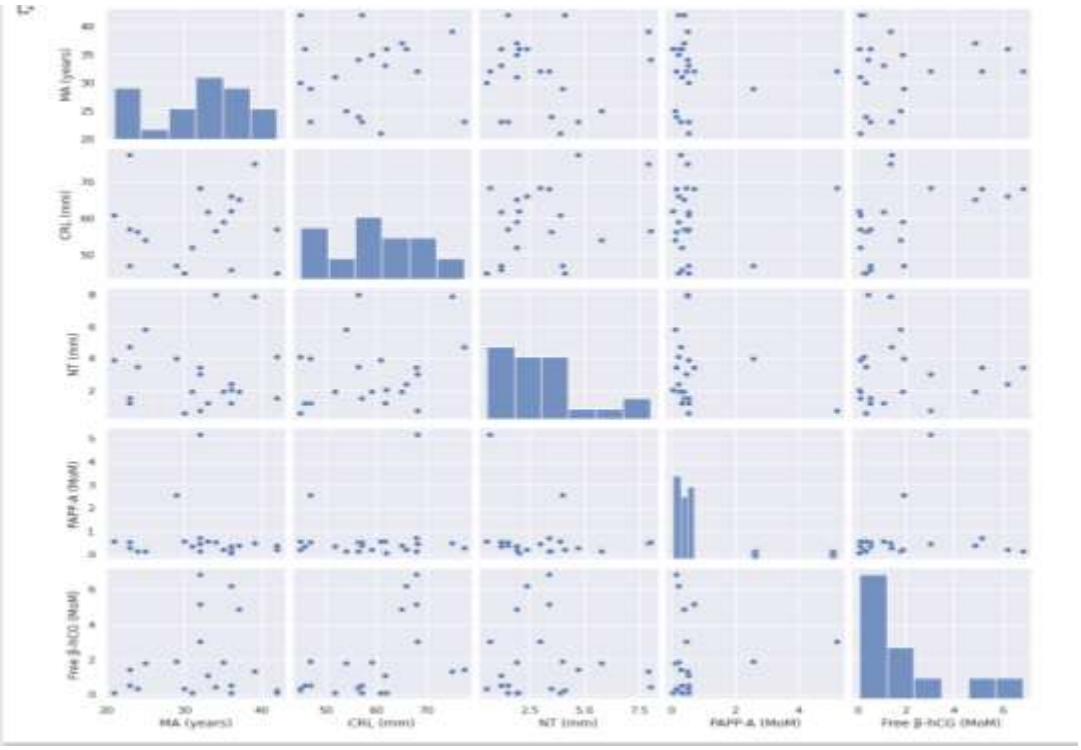

Figure 4. Cluster all data

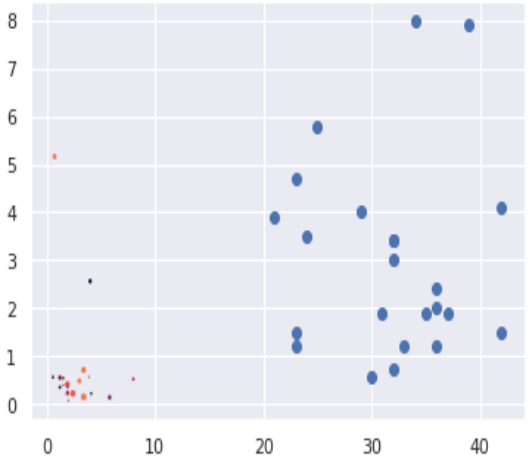

Figure 5. K-means Clustering

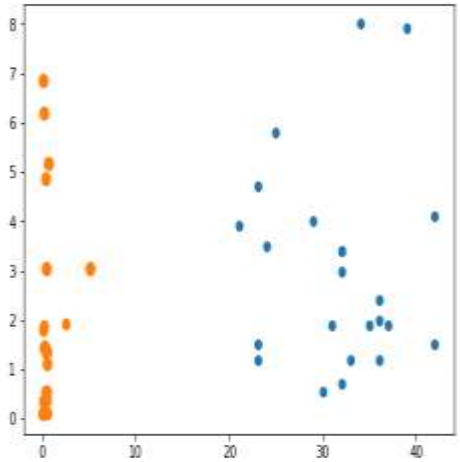

Figure 6. DBSCAN Clustering

From Figures 5 and 6, it can be seen that the clustering as a predictive analysis of Down syndrome was successful. Although the two have quite significant differences, that k-means can't let the noise get denser and only DBSCAN manages to do it. At least from the two algorithms, the results also show that the level of prediction of Down syndrome from the parameters used is very clear that the two colors come from the high risk of patient having a Down syndrome an low risk having a down syndrome. It can be seen from the data above the cluster that was carried out for the entire data that k-means clustering is more efficient for large datasets. DBSCAN Clustering cannot efficiently handle high dimensional datasets. Because if we made in comparison between the two of them so, we can see in the Table 3 that from the cluster the density for to determine the risk of Down syndrome and the origin of the cluster numbers is of two colors and shows different groupings.

Table 3. Different type of clustering

Clustering Technique	Cluster 1	Cluster 2
K-means	5	23
DBSCAN	7	20

On the cluster 1 and 2 of k-means give the distance more and DBSCAN can reduce the distance. It means k-means Gets difficult in high dimensional spaces as the distance between the points increases and Euclidean distance diverges (converges to a constant value). DBSCAN Works well for noisy datasets then can identity Outliers easily. The result shown that the algorithm have each has advantages and disadvantages. In terms of reading both have the truth that down syndrome has occurred. The results, it can be concluded that the two hierarchical clustering methods used, namely k-means clustering and DBSCAN are real results that in prenatal examinations there are still several more parameters that will be needed in predicting whether a person has Down syndrome. While clustering is useful for data analysis and as a preprocessing step for a number of learning tasks, we are interested in the specific pre-processing task of using clustering to gain more information about the data to improve prediction accuracy(19). There are some questions about whether clustering can be used correctly on an accurate prediction and It is not clear if clustering could help, though there is some evidence(20). As one as of the evidence of clustering prediction clustering in this study. Although it can be said that it is not very accurate, clustering can help in improving predictions by non-manually and the results show that the level of comparison between the two algorithms is a mean that is close to the level of clustering of better predictions although it has a noise level that cannot be compressed.

CONCLUSION

K-means and DBSCAN are still unsupervised algorithms that can perform clustering to predict Down syndrome. The grouping of the two algorithms requires attributes of the data that have been reprocessed. The future of k-means clustering can be improved by eliminating noise. and it is recommended to continue to use the K method to predict, because even though DBSCAN will be a successful cluster it is very different in terms of density within the cluster and detecting cluster noise and DBSCAN cannot accept the number of clusters in a large data attribute density level

ACKNOWLEDGMENTS

The authors would like to Informatics Department of University Muhammadiyah Surakarta for support this research paper publication

REFERENCES

1. Sundari SS, Agustin YH, Silmi H. Sistem Pakar Diagnosa Tingkat Retardasi Down Syndrom Pada Anak Menggunakan Metode Certaining Factor. Semin Nas Sist Inf Dan Tek Inform. 2019;289–300.
2. Ramanathan S, Sangeetha M, Talwai S, Natarajan S. Probabilistic Determination of Down's Syndrome Using Machine Learning Techniques. 2018 Int Conf Adv Comput Commun Informatics, ICACCI 2018. 2018;126–32.
3. Malan V, Bussières L, Winer N, Jais JP, Baptiste A, Le Lorc'h M, et al. Effect of cell-free DNA screening vs direct invasive diagnosis on miscarriage rates in women with pregnancies at high risk of trisomy 21 a randomized clinical trial. JAMA - J Am Med Assoc. 2018;320(6):557–65.
4. Koivu A, Korpimäki T, Kivelä P, Pahikkala T, Sairanen M. Evaluation of machine learning algorithms for improved risk assessment for Down's syndrome. Comput Biol Med [Internet]. 2018;98:1–7. Available from:

https://doi.org/10.1016/j.compbiomed.2018.05.004
5. Bansal A, Sharma M, Goel S. Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining. Int J Comput Appl. 2017;157(6):35–40.
6. Wang T, Ren C, Luo Y, Tian J. NS-DBSCAN: A density-based clustering algorithm in network space. ISPRS Int J Geo-Information. 2019;8(5).
7. Neocleous AC, Nicolaides KH, Schizas CN. First Trimester Noninvasive Prenatal Diagnosis: A Computational Intelligence Approach. IEEE J Biomed Heal Informatics. 2016;20(5):1427–38.
8. Kaya H. Olas ı l ı ksal S ı n ı fland ı r ı c ı lar ile Do ğ um Öncesinde Trizomi 21 Risk Hesaplamas ı Prenatal Risk Assessment of Trisomy 21 by Probabilistic Classifiers. 2013;13–6.
9. Cordova I. DBSCAN on Resilient Distributed Datasets. 2015;531–40.
10. He F, Lin B, Mou K, Jin L, Liu J. A machine learning model for the prediction of down syndrome in second trimester antenatal screening. Clin Chim Acta [Internet]. 2021;521:206–11. Available from: https://doi.org/10.1016/j.cca.2021.07.015
11. Luhpdwk K, Ri H. / Hduqlqj 7Hfkqltxhv Iru ’ Ldjqrvlv Ri. 2017;872–8.
12. Qin B, Liang L, Wu J, Quan Q, Wang Z, Li D. Automatic identification of down syndrome using facial images with deep convolutional neural network. Diagnostics. 2020;10(7).
13. Zhang Z, Lange K, Xu J. Simple and scalable sparse k-means clustering via feature ranking. Adv Neural Inf Process Syst. 2020;2020-Decem(NeurIPS):1–13.
14. Soumya DS, Arya V. Chromosome segmentation using K-means clustering. Int J Sci Eng Res. 2013;4(9):937–40.
15. Chakraborty S, Nagwani NK, Dey L. Performance Comparison of Incremental Kmeans and Incremental DBSCAN Algorithms. Int J Comput Appl. 2011;27(11):14–8.
16. Haraty RA, Dimishkieh M, Masud M. An enhanced k-means clustering algorithm for pattern discovery in healthcare data. Int J Distrib Sens Networks. 2015;2015.
17. Oyelade OJ, Oladipupo OO, Obagbuwa IC. Application of k-Means Clustering algorithm for prediction of Students Academic Performance. 2010;7:292–5. Available from: http://arxiv.org/abs/1002.2425
18. Perumal M, Velumani B. Design and development of a Spatial DBSCAN Clustering framework for location prediction- An optimization approach. Proc 3rd Int Conf Commun Electron Syst ICCES 2018. 2018;(Icces):942–7.
19. Trivedi S, Pardos ZA, Heffernan NT. The Utility of Clustering in Prediction Tasks. 2011;(September):1–11.
20. Biswas G, Trivedi S, Pardos ZA, Heffernan NT. Clustering Students to Generate an Ensemble to Improve Standard Test Score Clustering Students to Generate an Ensemble to Improve Standard Test Score Predictions. 2011;(June).