Research Article

Journal of Scientific Enquiry
Vol: 2, Pages: 26-29

Quality Assessment of H.S. (10+2) Students for the Development of the Civilized Society using KNN

Rabi Sankar Pandey[1], Sanat Kumar Mahato[1,2,*], Nikita Banerjee [2], Abhaya Pada Das[2], Ekram Ansary[2], Rimpa Deshmukh[2], Sumanta Ray[2]

1 Department of Mathematics, Sidho-Kanho Birsha University, Purulia, India
2 Department of Comp. Sc., Sidho-Kanho Birsha University, Purulia, India

Received: 20.05.2022; Accepted: 16.06.2022; Published Online: 30.06.2022

Abstract:

The aim of this paper is to identify the students, those are not continuing their higher education, after passing H.S. (10+2) examination, due to several reasons. The research work has been done on primary datasets, collected from Hijuli High School (H.S.), Purulia. Here we have implemented the concept of KNN algorithm by using R programming language. If we will be able to identify the students, has the probability to discontinue their higher education after passing H.S. (10+2) examination in a very current state, by means, when they are appearing H.S. (10+2) examination and have a special concentrate on that through counselling or any other process, maybe they are continuing their higher education.

Keywords:

KNN, R, training dataset, testing dataset, H.S. (10+2) students

1. Introduction:

In the present scenario, we will find so many students, they are not continuing their higher education after qualifying H.S. (10+2) examinations. Here in this research paper, we have tried to identify the students, continuing or not continuing their higher educations in terms of some parameters related to the H.S. (10+2) students, after completing their H.S. (10+2) examination. Our basic moto is to identify the students, has the probability to discontinue their higher education, after completing their H.S. (10+2) examination. Here we have used KNN data classification tool. The concept of KNN has been implemented using R programming language. This research has been conducted using the primary datasets, collected from Hijuli High School (H.S.), Purulia.

2. Data collection:

Primary data has been collected from Hijuli High School, which is one H.S. (10+2) standard school in Purulia district. In this present research work we have required one training Dataset & one testing Dataset. Our training dataset has 76 records and testing dataset has 25 records. Here training and testing records has been selected randomly. Training datasets has the records related to H.S. (10+2) pass out students along with the output column “drop out” or “not drop out”, and testing dataset has the records related to current H.S. (10+2) students. Testing dataset does not have the final output column. The testing dataset has been furnished in Table-1.

2.1 Selected indicators of H.S. (10+2) students:

Six different parameters of H.S. (10+2) pass out students along with their final remarks has been selected for training dataset and these are i) Regularity, ii) Discipline, iii) Dress code, iv) Financial Background, v) Performance in class test, vi) Communication Skill, and vii) Remarks and same number of parameters excluding final remarks of H.S. (10+2) current year students has been selected for testing dataset and these are i) Regularity, ii) Discipline, iii) Dress code, iv) Financial Background, v) Performance in class test, vi) Communication Skill. We have prepared the primary datasets, both training and testing, by the guidance of the teachers of Hijuli High school, Purulia. Each parameters have maximum 10 marks and minimum 0 marks, in terms of this concept each and individual tuple is scaled. Say as for example, if we consider the first parameter, if one student has 100% attendance, he/she will be obtained 10 out of 10, if one has 0% attendance, he/she will be obtained 0 out of 10, if one has 50% attendance, he/she will be obtained 5 out of 10, if one has 70% attendance, he/she will be obtained 7 out of 10, and so on so forth.

3. KNN:

In Machine learning, we have supervised, unsupervised and reinforcement learning. In supervised machine learning algorithm, we will have labeled data, means well defined output will be there but in unsupervised machine learning algorithm we will only be able to find the hidden

[1] Email: sanatkmahato@gmail.com

 DOI: https://doi.org/10.54280/jse.222104

structure from the dataset. In supervised learning we have classification problems and regression problems. KNN is one example of supervised learning and closely associated with classification. So, we can conclude KNN is an approach for data classification. In a dataset, how one observation will be a member of one group that we will be able to judge by applying the concept of KNN algorithm.

Say, we have the observations x_1, x_2,..... x_n, belonging in the class c_1,c_2.....c_n respectively. Now if one new observation enters into the system say x_{new}, by applying the KNN algorithm we will be able to judge in which class x_{new} data point will retains.

Here first we have to find out the distance in between x_{new} to x_1, x_{new} to x_2 , ..., x_{new} to x_n. Let us assume the distance in between x_{new} to x_1, x_{new} to x_2 , ..., x_{new} to x_n are d_1 , d_2 , ...,d_n respectively. Now we have to arrange all the distance according to closest to farthest. Then if we assume K=3, then three observations associated to first three shorted distances are in the same class and if all the three observations belonging to c_1, then x_{new} belongs to c_1 and if two observations belong to c_1, in terms of majority votes x_{new} is also belong in c_1.

4. R Programming Language

We all know R is an open-source programming language, so that we can easily download it from internet. It is most widely used programming language for data analysis. It has so many packages. R has developed by Ross Ihaka and Robert Gentleman at the university of Auckland in mid 1990s. R comes from S language.

5. Method, Result and Discussion:

Step 1: Prepare training dataset and testing dataset as per the requirement.

Step 2: Install R-console and R-studio respectively.

Step 3: Import training and testing dataset in R as ".csv" format.

```
> a<-read.csv(file="C:/Users/COM-003/Desktop/Knn_TKdata.csv")

> View(a)

> t<-a[,-3]

> View(t)

> b<-read.csv(file="C:/Users/COM-003/Desktop/Test data.csv")

> View(b)
```

Step 4: set the working directory in R.

```
>getwd()

[1] "C:/Users/COM-003/Documents"

> setwd("C:/Users/COM-003/Desktop")

>getwd()

[1] "C:/Users/COM-003/Desktop"
```

Step 5: Install "class" and "caret" packages one by one. (These R packages are required to implement the concept of KNN algorithm)

```
>install.packages("class")

>install.packages("caret")

>library(class)

>library(caret)
```

Step 4: write R code to implement the concept of KNN on the training dataset as well as testing dataset. The R code is as follows:

```
> View(b)

> c<-b[,-1]

> f<-c[,-1]

> x<-f[,-3]

> predict<-knn(train=t[,-7],test=x[,-7],cl=t$REMARKS,k=2)

> predict
```

Step 5: the results is as follows:

```
[1] Yes Yes Yes No  Yes No  Yes No  Yes Yes Yes No
Yes No  Yes Yes No  No

[19] Yes Yes Yes Yes No  Yes Yes

Levels: No yes Yes
```

We had 25 records excluding remarks and here we established the final remarks of 25 students in terms of yes or no by using KNN. By means we will be able to identify the students, has the probability for not continuing their higher studies.

6. Conclusion:

In this research paper, we can easily identify the H.S. (10+2) pass out students, those has the probability to discontinue higher studies. Here also we have evaluated H.S. (10+2) students in a different way. We can conclude in this way that, in this research paper, we have tried to predict the students, has the probability to discontinue their higher education after passing H.S. (10+2) examination. A huge amount of work has been carried out for the assessment of H.S. (10+2) level students. But here we have assessed the students in

a different way and that will be helpful for the researchers in this domain.

Table-1: Testing Dataset

REGULARITY	DISCIPLINE	DRESS CODE	FINANCIAL BACKGROUND	PERFORMANCE IN CLASS TEST	COMMUNICATION SKILL
4.5	5	10	6	6	5
4.5	10	8	0	10	2
2.5	9	10	10	9	10
3	9	10	7	9	10
3	7	9	5	8	6
1	4	2	10	5	3
4	9	10	10	9	9
3	5	3	6	7	0
7.419355	9	9	5	6	4
9.354839	8	10	5	7	10
7.741935	9	10	5	8	10
6.774194	9	9	6	10	5
2.903226	6	6	7	10	5
9.354839	4	6	7	4	10
6.451613	8	10	3	6	10
6.451613	6	5	7	7	5
7.741935	5	0	8	4	2
5.806452	8	10	3	3	5
2.580645	8	10	2	7	5
2.903226	4	5	1	5	0
2.258065	8	10	8	2	10
3.225806	7	8	5	9	5
4.516129	2	0	8	3	0
6.774194	7	10	10	9	10
7.741935	5	9	5	4	5

References:

[1] O.Ore, *Graphs and their uses*, The L. W. Singer Company, New York (1963).

[2] F.Harary, *Graph Theory*, Addison- Wesley, Reading, Mass (1969)

[3] N. Deo, *Graph Theory With Applications to Engineering and Computer Science*, Prentice-Hall of India Private Limited, New Delhi (1997).

[4] R. J. Wilson, *Introduction to graph theory*, Longman Group Ltd., London (1978)

[5] A. T. Balaban, (Eds.), *Chemical Applications of Graph Theory*, Academic Press, London (1976).

[6] S. J. Cyvin, I. Gutman, Kekulé *Structures in Benzenoid Hydrocarbons*, Springer- Verlag, Berlin, 1988.

[7] I. Gutman, and O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, New York (1986).

[8] J. R. Dias, *Molecular Orbital Calculations Using Chemical Graph Theory*, Springer-Verlag, New York (1993).

[9] H. Hosoya, H. Kumazaki, K. Chida, M. Ohuchi, and Y. D. Gao, *Pure & Appl. Chem*., 62, 445 (1990).

[10] D. J. Klein, M. J. Cravey, and G. E. Hite, *Polycyclic Aromatic Compounds*, **2**, 163 (1991).

[11] J. Cioslowski, *Electronic Structure Calculations on Fullerenes and Their*

Derivatives, Oxford University Press, New York, 1995.

[12] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Application*, Academic Press, San Diego, CA (1979).

[13] A. Jr. Streitwieser, *Molecular Orbital Theory for Organic Chemists*, John Wiley & Sons, Inc., New York (1961).

[14] A. Graovac, I. Gutman, and N. Trinajstić, *Topological Approach to the Chemistry of Conjugated Molecules*, Springer, New York (1977).

[15] (a) A. K. Mukherjee, and K. K. Datta, *Proc. Indian Acad. Sci. (Chem. Sci.)*, **101**, 499 (1989); (b) A. K. Mukherjee, and D. K. Das, *Proc. Indian Acad. Sci. (Chem. Sci.)*, **105**, 111 (1993).

[16] (a)R. K. Mishra, B. K. Mishra, and S. Singh, *Indian J. Chem.*, **27A**, 377 (1988); (b) S. Singh, R. K. Mishra, and B. K. Mishra, *Indian J. Chem.*, **27A**, 653 (1988).

[17] A. T. Balaban, (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers, The Netherlands, 1999.

[118] R. E. Merrifield, and H. E. Simmons, *Topological Methods in Chemistry*, Wiley, New York (1989).

[19] M. A. Johnson, and G. M. Maggiora, *Concepts and Applications of Molecular Similarity*, Wiley Interscience Publication, 1990.

[20] I.T. Jolliffe, Principal Component Analysis", Springer.

[21] T.Zhang, J.Guo, N.Zhang, "Application of PCA in Quality Assessment of College Students" International Conference on Information Technology, Computer Engineering and Management Science, IEEE Computer Society, 2011, pp.134-137.

[22] S.N. Mandal, J. Pal Choudhury, S.R. Bhadra Choudhury, "Neuro-PCA-Factor Analysis in Prediction of Time Series Data," American Journal of Intelligent Systems, 2012, pp.45-52.

[23] G.Hong-bo, L.Xuan, "Research on Optimizing Quality of Higher Vocational Education Based on Principal Component Analysis of Data Mining", International Conference on Intelligence Science and Information Engineering, IEEE Computer Society, 2011, pp.534-537.

[24] S.Ray, A.C.Mondal, A.Neogi, "A Study of the Development of Principal Component Analysis using Neural Network in Assessment of Higher Education System", Proceedings of NACCS- 2012, Burdwan University, West Bengal, India, March 2012.