Journal of Mines, Metals and Fuels, 71(5): 650-655; 2023. DOI: 10.18311/jmmf/2023/34165

Print ISSN : 0022-2755

Journal of Mines, Metals and Fuels

Contents available at: www.informaticsjournals.com/index.php/jmmf

Study of Online Pre-admission Enquiry Prediction under the Framework of k-Nearest Neighbor (k-NN) Algorithm

Bidyut Roy*, Bablu Pramanik and Manoj Kumar Srivastav

Department of Computer Science and Technology, Swami Vivekananda University, Barrackpore, West Bengal, India.
E-mail: bidyutr@gmail.com / bablu.cse@gmail.com / mksrivastav2015@gmail.com

Abstract

Admission to an appropriate college is an important process. The method of enquiry is an important part of it. As a result, students gather data, analyse it, and attempt to predict which college is best for them to attend. In general, students find it useful to research the trends in fee structures, academic staff numbers, multi-national companies tie up, and placements received by students during that academic year. K-Nearest Neighbors (k-NN) is one of the simplest learning algorithms used in supervised learning approaches. It makes assumptions about the similarity between new cases or data and available cases and places new cases in the category that most closely resembles the available categories. In this paper, the authors have tried to identify or predict the category or class of a particular dataset using the k-NN algorithm.

Keywords: *Enquiry, Admission, k-NN, predict*

1.0 Introduction

Every year, thousands of students apply to be taken to college, either physically or online. Physical enquiries lead to problems in time consumption, receiving the right information, and handling students' different types of problems. Also, students are not aware of the academic structure, fees, faculty, placement, etc. Additionally, the admissions procedure is required to be transparent to give appropriate students a chance. The "Online Pre-admission Enquiry System" aims to help applicants for admission by giving data in a quicker, more transparent, and simple-to-use format for reference and subsequent actions.

A. Procedure for the Online Pre-admission Enquiry System in Steps

When a candidate opens a specified URL, he will be able to search for their enquiry:

1. Initial Admission Process: Student online admission starts with the tab "Search." The student collects all required information and documents, checks all the check boxes, and proceeds to the next step.
2. The Advanced Admissions Process: Students can login to complete the intermediate level process and make detailed enquiries using the chat box, such as the number of academic staff, seats, tuition fees, number of students who received placements, and so on.
3. Compare their enquiries and the data obtained from various enquiries: Students can make comparative studies of different colleges based on factors like JEE-rank, fee structures, academic staff, student placement, etc.
4. Make a prediction to take admission (yes or no): Students will be able to use k-NN algorithm methods and will try to predict which college is better to join.

Data science is growing and machine learning is an important component in it. Students can make enquiries through chatbots. The artificial intelligence in chatbots is programmed to respond to voice or text commands. Students

*Author for correspondence

can collect different types of data based on their requirement in enquiry system. Machine learning helps with classification and prediction through the use of statistical methods. A natural approach to similarity-based classification is k-Nearest Neighbor (k-NN), which is a non-parametric method used for both classification and regression problems. It is a simple and powerful non-parametric algorithm that predicts the category of the test instance according to the 'k' training samples that are closer to the test instance and classifies it in the category that has the largest probability. During the admissions process, online enquiries for student admission to the college as a result, the k-Nearest Neighbor algorithm is applicable to investigate pre-admission enquiries. The algorithm relies on the assumption that similar objects are close to each other in the feature space, and k-NN performs instance-based learning, which just stores the training data at the time when predictions need to be made[2]. The most popular distance measure, such as Euclidean distance, is used in k-NN to determine the k-instances that are similar to the test instance. The value of 'k' best calculated by tuning with different 'k' values and choosing the 'k' that classifies the test instance more accurately[1].

2.0 Literature Survey

K. Taunk et al. (2019) have discussed A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. The K-nearest neighbour algorithm is discussed, which is a simple but high-accuracy algorithm that has proven effective in several cases[3].

L. Wang (2019), in his paper Research and Implementation of Machine Learning Classifier Based on k-NN, discusses the implementation of the k-NN machine learning classifier, and the results of the data test show that the classification effect is achieved. The k-NN classification algorithm is subjective because a distance scale must be defined[4].

J. Sun et al. (2018) discussed improving the algorithm's efficiency by reducing computational complexity and optimising similarity measures in their paper A Survey of the kNN Algorithm[5].

A. Joshi et al. (2015), in their paper Analysis on Student Enquiry Systems, discussed the best tools for data mining, known as WEKA (Waikato Environment Knowledge Analysis), which is used to formulate the process of analysis[6].

O.Lala et al (2020) in their paper An Improved Rapid Response Model for University Admission Enquiry System Using Chatbot have tried to develop a model using IBM Watson to design a Chatbot for rapid response to admission enquiries[7].

3.0 Data Work

A. Study of data work for students' pre-admission Enquiry using the k-NN algorithm

Implementation of k-NN algorithm methods:

Step 1: A dataset is required to apply the k-NN algorithm. So, it is required to collect training data and test instances.

Step 2: Identify the value of the parameter k as the number of nearest neighbors.

Step 3: Determine the Euclidean distance between the test instance and each of the training instances.

Step 4: Sort the distance in increasing order and choose the first k training data instances that are closest to the test instances.

Step 5: Using a majority vote, determine the class of the test instance.

B. Training the dataset and making a prediction

Inputs: training dataset T, distance metric d, test instance t, and the number of nearest neighbours k.

Output: Predicted class or category

Prediction: For test instance t,

Test instance (10000,8.7, 100,5, and 2)

```
# Determine the Euclidean distance between five vectors.
from math import sqrt
def euclidean_distance(col1, col2,col3,col4,col5):
#       distance = 0.0
#       for i in range(len(col1)-1):
        distance = ((col1-11000)**2)+((col2 - 8.7)**2)+
((col3 - 100)**2)+((col4 - 5)**2)+((col5-2)**2)
        return sqrt (distance)
```

- Example showing euclidean distance calculation:

```
distance = euclidean_distance (13256,7.7,82,0,2)
print (distance)
distance = euclidean_distance (13256,5.2,90,2,0)
print (distance)
output:
2256.0775695884217
2256.0277591377285
```

Problem for k-NN Algorithm: There is an online pre-admission enquiry by different students, we have a dataset that contains information related to searching for pre-admission from different educational organization. This dataset contains different information related to students' JEE

rank, Fees of that course, No. of academic staff, No. of placement founds above 10 lakhs, No. of MNC TIE up with college and admission taken by the students (yes/no). The dataset contains students' JEE rank, Fees of that course, No. of academic staff, No. of placement founds above 10 lakhs, No. of MNC TIE up with college are considered for the independent variable and the admission taken (yes/no) is for the dependent variable[14]. Table 1 is the dataset:

Sorting the euclidean distance in ascending order and selecting first 7 nearest training data instances to the test instance. For k=7, the selected nearest neighbors are shown in Table 2:

Table 1: Student Enquiry Details

Name	JEE Rank	Serial no of search college	Fees (Lacs)	No of Academic staff	No. of placement founds above 10 lakhs	No. of MNC TIE up with college	Admission taken (yes/no)	Euclidean Distance
Sraman Dutta Bar	13256	1	7.7	82	0	2	No	2256.07757
		2	5.2	90	2	0	No	2256.027759
		3	16	120	10	4	Yes	2256.106888
Gourab Biswas	4780	4	14	90	8	4	No	6220.011342
		5	16	120	10	4	Yes	6220.038769
Sohini Sen	4423	6	9.3	100	8	1	No	6577.000788
		7	8.4	120	7	2	No	6577.03072
		8	16	120	10	4	No	6577.036665
Prajita Bera	9965	9	16	120	10	4	Yes	1035.232964
Chanchal Mandal	18235	10	4.7	88	0	0	No	7235.013061
		11	7.9	90	6	2	No	7235.007024
		12	3.2	90	1	0	No	7235.010384
		13	4.8	80	3	0	No	7235.029247
Bineet Pradhan	4751	14	8.4	110	7	2	No	6249.008329
		15	16	120	10	4	No	6249.038589
		16	6.7	100	2	1	No	6249.00112
Tithi Das	18452	17	3.8	86	1	0	No	7452.016104
		18	8.4	120	7	2	No	7452.027113
		19	16	120	10	4	No	7452.03236
Shruti Das	15687	20	16	120	10	4	Yes	4687.051449
Debolina Jana	18750	21	7.8	80	3	2	No	7750.026117
		22	4.2	75	2	0	No	7750.042468
		23	16	120	10	4	No	7750.031115
		24	9.2	85	6	1	No	7750.014661
Sourav Barik	12214	25	4.7	60	1	0	No	1214.673619
		26	5.8	90	3	1	Yes	1214.046708
Debarshi Mohan Jana	18478	27	4.8	96	1	0	No	7478.003424
		28	4.2	80	2	0	No	7478.028968
		29	16	120	10	4	No	7478.032247
		30	3.8	90	1	0	No	7478.009629

Sathi Jana	16511	31	4.7	82	0	0	No	15511.01189
		32	5.7	90	2	1	No	5511.010797
		33	4.7	90	5	0	No	5511.010887
Saheli Rana	6471	34	8.4	92	7	2	Yes	4529.007517
		35	6.7	92	2	1	No	4529.008611
		36	7.8	84	0	2	No	4529.031112
		37	6.8	82	0	1	No	4529.039038
Ipshita Jana	13854	38	5.8	90	2	0	No	2854.02127
		39	4.7	60	0	1	Yes	2854.287652
		40	16	120	10	4	No	2854.084492
Anulipi Das	15960	41	5.2	90	3	0	No	6960.008639
		42	5.2	85	0	0	No	6960.019127
		43	4.7	60	0	1	Yes	4960.165521
Ahana Manna	14014	44	16	120	10	4	Yes	3014.080007
Debarupa Mondal	7589	45	14	120	15	5	Yes	3411.078728
Rima Barman	15411	46	12	115	9	3	No	4411.028666
		47	7.2	85	2	2	No	4411.02678
		48	16	120	10	4	No	4411.054669
		49	3.8	90	1	0	No	4411.016324
Shilpa Bar	14102	50	4.7	83	1	0	No	3102.052385
		51	3.8	90	3	0	No	3102.021278
		52	14	90	8	4	Yes	3102.022742
Aritra Bag	11257	53	16	120	10	4	Yes	257.9366007

Table 2: Euclidean distance sorted instance

Instance	Euclidean Distance	Result
53	257.9366007	Yes
9	1035.232964	Yes
26	1214.046708	Yes
25	1214.673619	No
2	2256.027759	No
1	2256.07757	No
3	2256.106888	Yes

The class for the predicted test instance is 'Yes' due To majority of "yes".

To test the accuracy of above problem, Euclidean distance value 5603.4518 above for prediction is taken as 'No' and below 'Yes' (Table 3).

Table 3: Actual and Predicted table

Euclidean distance	Admission Taken (Yes/No)	Prediction
6220.011342	No	No
6577.000788	No	No
6577.03072	No	No
6577.036665	No	No
7235.013061	No	No
7235.007024	No	No
7235.010384	No	No
7235.029247	No	No
6249.008329	No	No
6249.038589	No	No
6249.00112	No	No
7452.016104	No	No
7452.027113	No	No
7452.03236	No	No

C. Confusion Matrix

A classification model's performance is frequently represented as a confusion matrix. The matrix (Table 4) shows

7750.026117	No	No
7750.042468	No	No
7750.031115	No	No
7750.014661	No	No
7478.003424	No	No
7478.028968	No	No
7478.032247	No	No
7478.009629	No	No
15511.01189	No	No
5511.010797	No	No
5511.010887	No	No
6960.008639	No	No
6960.019127	No	No
6220.038769	Yes	No
2256.07757	No	Yes
2256.027759	No	Yes
1214.673619	No	Yes
4529.008611	No	Yes
4529.031112	No	Yes
4529.039038	No	Yes
2854.02127	No	Yes
2854.084492	No	Yes
4411.028666	No	Yes
4411.02678	No	Yes
4411.054669	No	Yes
4411.016324	No	Yes
3102.052385	No	Yes
3102.021278	No	Yes
2256.106888	Yes	Yes
1035.232964	Yes	Yes
4687.051449	Yes	Yes
1214.046708	Yes	Yes
4529.007517	Yes	Yes
2854.287652	Yes	Yes
4960.165521	Yes	Yes
3014.080007	Yes	Yes
3411.078728	Yes	Yes
3102.022742	Yes	Yes
257.9366007	Yes	Yes

Table 4: Confusion matrix Table

Actual value	Predicted value 01		
	TN	FP	0(No)1(Yes)
	FN	TP	
Actual value	Predicted value		
0	0	TN (True Negatives)	
0	1	FP (False Positive)	
1	0	FN (False Negative)	
1	1	TP (True Positive)	

TN=27, FP=14, FN=1, TP=11, N=53

Accuracy = (TP+TN)/N = (11+27)/53 = 0.716981

Table 5: Calculation of different measure[12]

Measure	Value	Derivations
Sensitivity or true Positive Rate (TPR)	0.9167	TPR = TP/(TP+FN)
Specificity (SPC) or true Negative rate (TNR)	0.6585	SPC = TN/(FP+TN)
Precision or positive predictive value (PPV)	0.4400	PPV = TP/(TP+FP)
Negative Predictive value (NPV)	0.9643	NPV = TN/(TN+FN)
False positive rate (FPR)	0.3415	FPR = FP/(FP+TN)
False discovery rate (FDR)	0.5600	FDR = FP/(FP+TP)
False negative rate (FNR) or miss rate	0.0833	FNR = FN/(FN+TP)
Accuracy (ACC)	0.7170	ACC = (TP+TN)/(P+N)
F1 Score (F1)	0.5946	F1 = 2TP/(2TP+FP+FN)

the number of correctly and incorrectly classified examples compared to the actual result (target value) of the test data. Advantage of using the confusion matrix as an evaluation tool is that it allows for more detailed analysis (e.g., when a model mix two of its classes) than a simple subset of correctly classified examples (accuracy). The data set is unbalanced (i.e., there are large numerical differences between different classes). The binary classifier, which is the simplest types of classifier, simply has two classes: positive and negative, yes and no, etc. A confusion matrix[12] is a matrix that compares actual and predicted values and is represented in tabular in Table 4:

4.0 Advantage

This online portal will help students make decisions before admission to the institute. The online analysis will help the student identify the academic structure, fees, faculty, and placement of the students from institute. This procedure will make admission decisions more transparent and faster.

5.0 Conclusions and future scope

The basic approach of this paper is based on the k-NN algorithm, which allows us to select the trends of the appropriate college and take admission. Students will be able to predict which college is better for them if an appropriate dataset based on the k-NN algorithm is available from the online portal. In the future, there will be more scope to study k-NN algorithm methods if it is possible to include all records related to students' applications on the online pre-admission enquiry portal. In future, the following modules are recommended for inclusion in the portal.

- Use of weighted k-NN algorithm to get a more accurate result.
- Use the Decision Tree feature in the application to make the decision to enrol in the institute.
- Need to develop a pre-enquiry admission management system.
- Maintaining a central database to access the data.

6.0 References

1. S.Sridhar and M.Vijayalakshmi, (2021): Machine Learning. Oxford Press.
2. A. Srinivasaraghavan and V. Joseph, (2019): Machine learning. Willey Publication.
3. K. Taunk, S. De, S. Verma, and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019), IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8.
4. L.Wang, "Research and Implementation of Machine Learning Classifier Based on k-NN", IOP Conference Series: Materials Science and Engineering 2019 IOP Conf. Ser.: Mater. Sci. Eng. 677 052038.
5. J. Sun, W.Du, and N. Shi, "A Survey of kNN Algorithm" Information Engineering and Applied Computing pp-1-10 (2018).
6. A. Joshi and M. Singh, Analysis on Student Admission Enquiry System IJSRD - *International Journal for Scientific Research & Development*| Vol.3, Issue 10, pp-967-969, 2015.
7. Lala O. Gbenga, O.T. Oluwafunto, and A. H. Oluwatobi, (2020): "An Improved Rapid Response Model for University Admission Enquiry System Using Chatbot" *International Journal of Computer* (IJC) vol.38(1), pp-123–131 June.
8. R.K. Sarawale and G.V. Karbhari, (2019): Machine Learning, Niarali Prakashan.
9. "Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python by Rudolph Russell - PDF Drive," www.pdfdrive.com. https://www.pdfdrive.com/machine-learning-step-by-step-guide-to-implement-machine-learning-algorithms-with-python-e158324853.html (accessed Dec. 22, 2022).
10. k-NN: K-Nearest Neighbors (no date) PDF Drive. Available at: https://www.pdfdrive.com/knn-k-nearest-neighbors-e49415195.html (Accessed: December 22, 2022).
11. M.K.Srivastav, (2017): "Study of Correlation Theory with Different Views and Methods among Variables in Mathematics", *International Journal of Mathematics and Statistics Invention* (IJMSI), Volume 5 Issue 2 PP-21-23, February.
12. "Confusion Matrix - Online Calculator," onlineconfusionmatrix.com. https://onlineconfusionmatrix.com/
13. M.K.Srivastav, (2018): "Management Information System in Fully Connected Network Topology". *International Journal on Future Revolution in Computer Science & Amp;* Communication Engineering, vol- 4(2), pp- 346–349.
14. JavaTpoint, "K-Nearest Neighbor (KNN) Algorithm for Machine Learning - Javatpoint," www.javatpoint.com. https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning