Application of Data Mining in Optimization of Hotel's Food and Beverage Costs

I Wayan Surya Pramana[1*], Putu Risanti Iswardani[2], Ni Wayan Sri Aryani[3]

[1,2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3]Department of Electrical and Computer Engineering, Udayana University
[*]guzsurya.pramana@gmail.com

Abstract - Optimized costs could increase hotel revenue. However, based on observations, there are various methods that can be used in cost optimization, this indicating the possibility that there are other methods that can be used for this purpose. This study aims to propose application of data mining using the K-Nearest Neighbor (KNN) method to optimize costs by classifying feasibility of addition of raw materials for food and beverages based on data such as number of requests, supplies, usage, and purchases. Data used in this study is raw materials data for hotel food and beverage during January and February 2019 which amount to 152 data. Furthermore, data cleaning process applied to eliminate incomplete and duplicated data. This process produces 99 data that has been clean. Based on results of application and testing of the KNN method using confusion matrix, it is known that the value of k = 3 gives the best classification accuracy results of 80%. Then the classification results are represented in the form of graphs that are used as a basis for consideration of cost control. Based on this study, it was concluded that data mining using KNN method can be used in optimization of Hotel's Food and Beverage Costs.

***Index Terms*—Costs Optimization, Data Mining, K-Nearest Neighbor, Confusion Matrix.**

I. Introduction

OPTIMIZING food and beverage costs is a concern for every company that presents food and beverage, including hotels. Hotel revenue can be increased if unnecessary expenses can be anticipated. In research [1]-[5], [14] which discusses cost optimization, it is known that there are various methods that can be used in cost optimization in a process. This indicates possibility that there are other methods can be used for cost optimization. Due that, it's necessary to know bases used to determine the method in cost optimization.

To begin, from research [1] - [5] authors found that cost optimization can be done if the data such as number of requests [3], [4] stocks [1], [5], usage [2], and purchases data [1], [5] is known. Author concludes that the main points are in the data. Furthermore, research [6], [7] gave an idea about method in optimization hotel's food and beverage costs using data mining by classifying the feasibility of adding of raw materials for food and beverage based on data such as number of requests, supplies, usage, purchases, hotel occupancy. Classification method used is K-Nearest Neighbor (KNN) method which according to research [8], [9] found that the KNN method is better than other data mining methods through performance comparison testing in a case. Based on that, purpose of this study is to apply and test the application of data mining in optimization of hotel's food and beverage costs by classifying the feasibility of adding of raw materials for food and beverage using the KNN method.

The next section explains literature review. then in section III describes research methodology. Next, in section IV describes application dan testing of KNN method. Then based on these results, conclusions are given in section V

II. Literature Review

To find out methods that can be used in optimization of hotel's food and beverage costs, authors conducted a study on the literature with a discussion of cost optimization. As mentioned above, several previous studies have shown that there are various methods that have been used for cost optimization. For example, study of Azizah & Oesman [1] discussed how to optimize the cost of product distribution using the Saving Matrix and Generalized Assignment approach, then Pratiwie et al. [2] discussed how to optimize

the cost of meeting nutritional needs using genetic algorithms, Ali et al. [3] discussed how to optimize the cost of road construction maintenance projects using the stepping stone method, and study of priyo and paridi [4] discussed how to optimize the cost and time of construction projects in building construction using the Time Cost Trade Off method. This indicates possibility that there are other methods that can be used in cost optimization.

Furthermore, research leads to the purpose to know the bases used to determine method in cost optimization. Based on research [3], [4] it is known that optimization of costs can be done if the number of requests is known, in the study explained that the number of requests data can show the urgency of an item, so that the supply of goods can be monitored and costs can be allocated optimally . Then from research [1], [5] it is known that data on the amount of stocks and purchases must also be known to be able to optimize costs. By knowing the amount of stock, it can be estimated that the purchase of goods is needed or not, so that excessive expenditure can be reduced. Furthermore, from the research [2] it is known that in addition to the data mentioned earlier, data on the amount of usage must also be known. This is because if the amount of usage is known then the movement of an item can be identified, whether the item includes goods with fast moving or slow moving. So that, procurement of goods can be planned better. Based on this, the authors conclude that the main point to be able to determine the method in cost optimization is in the data.

Based on these findings, authors then studied other literature to find other methods that can be used for cost optimization, where research [6] [7] gave an idea to applied data mining in optimization of hotels food and beverage costs by classifying the feasibility of adding of raw material for food and beverage based on data such as number of requests, supplies, usage, and purchases. Furthermore, classification method used in this study is the K-Nearest Neighbor (KNN) method. This method was chosen based on research [8] which conducted a performance comparison test between KNN and Support Vector Machines (SVM) for classification of diabetes. The study found that performance of KNN was better than SVM. Next is research [9] which conducted a performance comparison test between KNN and Naïve Bayes for classification of work status data. The study found that performance of the KNN was better than Naïve Bayes.

III. METHODOLOGY

A. Research Location

The research took place at the Mercure Bali Nusa Dua Hotel, in the accounting department cost control section

B. Data collection methods

--Interview, conducted to Mr. Hery Suryawan as person in charge of cost control section of Mercure Bali Nusa Dua Hotel. This method obtained raw material data for hotel food and beverage during January and February 2019, furthermore this data will be used as data sets.

--Observation, conducted in daily cost monitoring process. This method obtained procedure for monitoring costs including checking number of sales, supplies, usage, and number of purchases

--Literature review, conducted in the literature related to data mining, especially the KNN method. This method obtained theory, how to applied and how to test KNN method

C. Data Preparation

Data preparation is carried out in three stages as follows [10] - [12];

--Data Selection, data obtained from previous stage is selected according to the research needs.

--Data cleaning, at this stage, the selected data is checked to ensure that data is unique (not duplicated) and complete (all data has value)

--Creation of New Data, cleaned data (unique and complete), arranged in new table for the application of the KNN method

D. Method Application and Testing

Application and testing of KNN method is applied to training data obtained from data sets, namely raw material data for hotel's food and beverage during January and February 2019. KNN method is applied and tested through the following stages [11]:

--Determination of K value, K value used in this study is K = 3, K = 4 and K = 5

--Calculation of distance between data evaluated with training data, calculation of distance is done using the following formula [12]:

$$Distance = \sqrt{\sum_{i=1}^{n} \left(X_{Training}^{i} - X_{Testing}\right)^2} \ldots\ldots\ldots (1)$$

With;

$X_{Training}^{i}$ = Training data of-i,
$X_{Testing}$ = Data to be evaluated
i = Training Data
n = Total of Training Data

--Determining the closest neighbors as much as K value, after the distance between all training data and data to be evaluated has been obtained, then the closest neighbors as much as K value is determined

--Accuracy testing with confusion matrix, in confusion matrix, accuracy can be calculated by knowing the value of TP (actual data is positive, classification data is positive), TN (actual data is negative, classification data is negative), FP (actual data is negative, classification data is positive), and FN (actual data is positive, classification data is negative) [13].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \ldots\ldots (2)$$

E. Results Analysis

At this stage method results are represented in graphical form to determine the cost of raw materials between food and beverages during the January and February 2019 periods which must be controlled so that strategies for cost optimization can be proposed

IV. Application And Testing

A. Data Selection

Data used in this study are data from stock report, purchases, sales and occupancy during January and February 2019 with total amount of 152 data. Details can be seen as follow

No.	Item	Actual	Purchased	Sold	Occ
1	Aperitif	Rp922,912.00	Rp1,089,039.00	Rp166,127.00	83
2	Arcoroc	Rp0.00	Rp0.00	Rp0.00	42
3	Beef Import	Rp2,034,000.00	Rp6,894,250.00	Rp4,860,250.00	79
4	Beef Local	Rp2,757,300.00	Rp6,444,950.00	Rp3,687,650.00	48
5	Beer Import	Rp912,334.00	Rp5,267,267.00	Rp4,354,933.00	94
6	Beer Local	Rp2,158,709.00	Rp22,834,203.00	Rp20,675,494.00	86
7	Bendorf	Rp0.00	Rp0.00	Rp0.00	63
8	Butter	Rp1,634,000.00	Rp14,263,663.00	Rp12,629,663.00	87
9	Candies	Rp0.00	Rp0.00	Rp0.00	75
10	Canned Food	Rp1,484,361.00	Rp5,733,255.00	Rp4,248,894.00	79
…	…	…	…	…	…
74	Wine Rose	Rp139,067.00	Rp1,161,538.00	Rp1,022,471.00	90
75	Wine Special	Rp389,541.00	Rp895,944.00	Rp506,403.00	83
76	Wine White	Rp1,501,229.00	Rp6,109,944.00	Rp4,608,715.00	56

Fig. 1. Datasets January 2019

No.	Item	Actual	Purchased	Sold	Occ
1	Aperitif	Rp490,009.00	Rp922,912.00	Rp432,903.00	53
2	Arcoroc	Rp0.00	Rp0.00	Rp0.00	89
3	Beef Import	Rp357,000.00	Rp2,916,000.00	Rp2,559,000.00	92
4	Beef Local	Rp2,656,310.00	Rp7,934,160.00	Rp5,277,850.00	92
5	Beer Import	Rp2,094,990.00	Rp7,419,170.00	Rp5,324,180.00	95
6	Beer Local	Rp2,548,056.00	Rp15,018,335.00	Rp12,470,279.00	90
7	Bendorf	Rp0.00	Rp9,400,000.00	Rp9,400,000.00	83
8	Butter	Rp2,030,500.00	Rp7,615,000.00	Rp5,584,500.00	56
9	Candies	Rp0.00	Rp0.00	Rp0.00	67
10	Canned Food	Rp779,659.00	Rp4,043,190.00	Rp3,263,531.00	63
…	…	…	…	…	…
74	Wine Rose	Rp74,500.00	Rp288,067.00	Rp213,567.00	92
75	Wine Special	Rp370,064.00	Rp389,541.00	Rp19,477.00	95
76	Wine White	Rp1,043,646.00	Rp6,210,031.00	Rp5,166,385.00	90

Fig. 2. Datasets February 2019

B. Data Cleaning

Data cleaning is performed to ensure that datasets is complete and no duplicated data. This process reduces the amount of dataset to 99 data

No.	Item	Actual	Purchased	Sold	Occ
1	Aperitif	Rp922,912.00	Rp1,089,039.00	Rp166,127.00	83
2	Beef Import	Rp2,034,000.00	Rp6,894,250.00	Rp4,860,250.00	42
3	Beef Local	Rp2,757,300.00	Rp6,444,950.00	Rp3,687,650.00	79
4	Beer Import	Rp912,334.00	Rp5,267,267.00	Rp4,354,933.00	48
5	Beer Local	Rp2,158,709.00	Rp22,834,203.00	Rp20,675,494.00	94
6	Butter	Rp1,634,000.00	Rp14,263,663.00	Rp12,629,663.00	86
7	Canned Food	Rp1,484,361.00	Rp5,733,255.00	Rp4,248,894.00	63
8	Cereals and Crackers	Rp678,068.00	Rp5,311,765.00	Rp4,633,697.00	87
9	Cheese	Rp5,262,225.00	Rp21,206,784.00	Rp15,944,559.00	75
10	Chicken	Rp1,097,861.00	Rp28,177,872.00	Rp27,080,011.00	79
…	…	…	…	…	…
48	Wine Rose	Rp139,067.00	Rp1,161,538.00	Rp1,022,471.00	90
49	Wine Special	Rp389,541.00	Rp895,944.00	Rp506,403.00	83
50	Wine White	Rp1,501,229.00	Rp6,109,944.00	Rp4,608,715.00	56

Fig. 3. Cleaned Datasets January 2019

No.	Item	Actual	Purchased	Sold	Occ
1	Aperitif	Rp490,009.00	Rp922,912.00	Rp432,903.00	53
2	Beef Import	Rp357,000.00	Rp2,916,000.00	Rp2,559,000.00	89
3	Beef Local	Rp2,656,310.00	Rp7,934,160.00	Rp5,277,850.00	92
4	Beer Import	Rp2,094,990.00	Rp7,419,170.00	Rp5,324,180.00	92
5	Beer Local	Rp2,548,056.00	Rp15,018,335.00	Rp12,470,279.00	95
6	Butter	Rp2,030,500.00	Rp7,615,000.00	Rp5,584,500.00	90
7	Canned Food	Rp779,659.00	Rp4,043,190.00	Rp3,263,531.00	83
8	Cereals and Crackers	Rp457,110.00	Rp3,380,381.00	Rp2,923,271.00	56
9	Cheese	Rp2,773,385.00	Rp15,268,255.00	Rp12,494,870.00	67
10	Chicken	Rp531,819.00	Rp20,072,861.00	Rp19,541,042.00	63
…	…	…	…	…	…
47	Wine Rose	Rp74,500.00	Rp288,067.00	Rp213,567.00	92
48	Wine Special	Rp370,064.00	Rp389,541.00	Rp19,477.00	95
49	Wine White	Rp1,043,646.00	Rp6,210,031.00	Rp5,166,385.00	90

Fig. 4. Cleaned Datasets February 2019

C. Creation of New Data

Cleaned data is arranged into a new table with the addition of the "Status" field which indicates the feasibility of adding of the item. The status given to each data is determined by cost control section of Mercure Bali Nusa Dua Hotel

No.	Item	Actual	Purchased	Sold	Occ	Status
1	Aperitif	Rp922,912.00	Rp1,089,039.00	Rp166,127.00	83	Not Feasible
2	Beef Import	Rp2,034,000.00	Rp6,894,250.00	Rp4,860,250.00	42	Not Feasible
3	Beef Local	Rp2,757,300.00	Rp6,444,950.00	Rp3,687,650.00	79	Not Feasible
4	Beer Import	Rp912,334.00	Rp5,267,267.00	Rp4,354,933.00	48	Feasible
5	Beer Local	Rp2,158,709.00	Rp22,834,203.00	Rp20,675,494.00	94	Feasible
6	Butter	Rp1,634,000.00	Rp14,263,663.00	Rp12,629,663.00	86	Feasible
7	Canned Food	Rp1,484,361.00	Rp5,733,255.00	Rp4,248,894.00	63	Not Feasible
8	Cereals and Crackers	Rp678,068.00	Rp5,311,765.00	Rp4,633,697.00	87	Feasible
9	Cheese	Rp5,262,225.00	Rp21,206,784.00	Rp15,944,559.00	75	Not Feasible
10	Chicken	Rp1,097,861.00	Rp28,177,872.00	Rp27,080,011.00	79	Feasible
…	…	…	…	…	…	…
48	Wine Rose	Rp139,067.00	Rp1,161,538.00	Rp1,022,471.00	90	Feasible
49	Wine Special	Rp389,541.00	Rp895,944.00	Rp506,403.00	83	Feasible
50	Wine White	Rp1,501,229.00	Rp6,109,944.00	Rp4,608,715.00	56	Not Feasible

Fig. 5. New Datasets January 2019

No.	Item	Actual	Purchased	Sold	Occ	Status
1	Aperitif	Rp490,009.00	Rp922,912.00	Rp432,903.00	53	Not Feasible
2	Beef Import	Rp357,000.00	Rp2,916,000.00	Rp2,559,000.00	89	Feasible
3	Beef Local	Rp2,656,310.00	Rp7,934,160.00	Rp5,277,850.00	92	Feasible
4	Beer Import	Rp2,094,990.00	Rp7,419,170.00	Rp5,324,180.00	92	Feasible
5	Beer Local	Rp2,548,056.00	Rp15,018,335.00	Rp12,470,279.00	95	Feasible
6	Butter	Rp2,030,500.00	Rp7,615,000.00	Rp5,584,500.00	90	Feasible
7	Canned Food	Rp779,659.00	Rp4,043,190.00	Rp3,263,531.00	83	Feasible
8	Cereals and Crackers	Rp457,110.00	Rp3,380,381.00	Rp2,923,271.00	56	Feasible
9	Cheese	Rp2,773,385.00	Rp15,268,255.00	Rp12,494,870.00	67	Feasible
10	Chicken	Rp531,819.00	Rp20,072,861.00	Rp19,541,042.00	63	Feasible
…	…	…	…	…	…	…
47	Wine Rose	Rp74,500.00	Rp288,067.00	Rp213,567.00	92	Feasible
48	Wine Special	Rp370,064.00	Rp389,541.00	Rp19,477.00	95	Not Feasible
49	Wine White	Rp1,043,646.00	Rp6,210,031.00	Rp5,166,385.00	90	Feasible

Fig. 6. New Datasets February 2019

D. Application and Testing of KNN Method

To find the most optimal K value, KNN method was applied to the January datasets which were divided into 30 training data and 20 testing data. As mentioned above, the K value used in this study is K = 3, K = 5, and K = 7. Method application obtained results as follows

No.	Item	Status	Status (M)
1	Testing Data 1	Feasible	Feasible
2	Testing Data 2	Not Feasible	Not Feasible
3	Testing Data 3	Feasible	Feasible
4	Testing Data 4	Feasible	Feasible
5	Testing Data 5	Feasible	Not Feasible
6	Testing Data 6	Feasible	Feasible
7	Testing Data 7	Feasible	Feasible
..	…	…	…
19	Testing Data 19	Feasible	Feasible
20	Testing Data 20	Not Feasible	Not Feasible

Fig. 7. KNN Application Result with K = 3

No.	Item	Status	Status (M)
1	Testing Data 1	Feasible	Feasible
2	Testing Data 2	Not Feasible	Not Feasible
3	Testing Data 3	Feasible	Feasible
4	Testing Data 4	Feasible	Feasible
5	Testing Data 5	Feasible	Not Feasible
6	Testing Data 6	Feasible	Feasible
7	Testing Data 7	Feasible	Not Feasible
..	...	...	...
19	Testing Data 19	Feasible	Not Feasible
20	Testing Data 20	Not Feasible	Feasible

Fig. 8. KNN Application Result with K = 5

No.	Item	Status	Status (M)
1	Testing Data 1	Feasible	Feasible
2	Testing Data 2	Not Feasible	Not Feasible
3	Testing Data 3	Feasible	Feasible
4	Testing Data 4	Feasible	Feasible
5	Testing Data 5	Feasible	Feasible
6	Testing Data 6	Feasible	Feasible
7	Testing Data 7	Feasible	Not Feasible
..	...	...	...
19	Testing Data 19	Feasible	Not Feasible
20	Testing Data 20	Not Feasible	Feasible

Fig. 9. KNN Application Result with K = 7

Furthermore, classification accuracy is measured using confusion matrix and obtained results as follows

No.	K Value	Accuracy
1	K = 3	80 %
2	K = 5	65 %
3	K = 7	75 %

Fig. 10. Classification Accuracy Results

Based on these results, it is known that K = 3 gives the best classification accuracy results which is 80%. So that K = 3 is stated to be the most optimal K value among those three options. Furthermore, K = 3 is used for application of KNN method in February data

E. Results Analysis

After KNN method is applied to all datasets, then data that classified as not feasible is represented as graphic as follows:

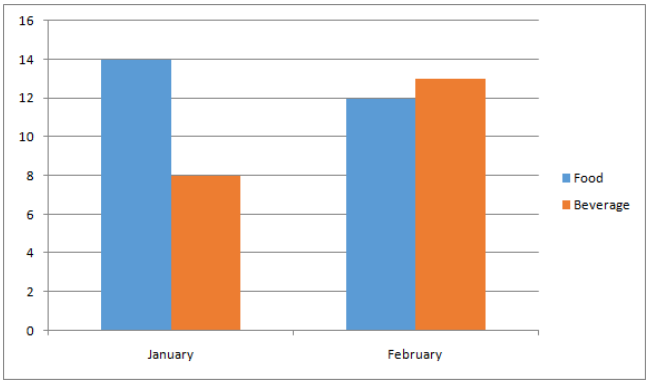

Fig. 11. Classified Not Feasible Data

Graph shows that data with the biggest not feasible classification on January data is on food raw material which is 14 data from a total of 20 classification data and data with the biggest not feasible classification on February data is on beverage raw material which is 13 data from 25 classification data. So that, to be able to optimize costs it is recommended that cost control to be focused on raw materials for both food and beverages that have the biggest not feasible classification data based on results of KNN method application

V. Conclusion

Based on above description, it can be concluded that KNN method can be used for classify feasibility of adding of raw material for food and beverage with classification accuracy of 80%. To optimize costs, cost control should be focused on raw materials for both food and beverages that have the biggest not feasible classification based on results of KNN method application.

Acknowledgment

Thank you to Mercure Bali Nusa Dua Hotel for provide us a place and data to conduct this research. Thank you to Mr. Hery Suryawan for his insightful contribution and valuable time.

References

[1] U., N., Azizah, and T., I., Oesman, "Optimalisasi Biaya Distribusi Produk PT. MADUBARU dengan Pendekatan Metode Saving Matrix dan Generalized Assigment," *Jurnal Rekavasi*, vol. 3, no. 2, pp. 102–110, Des. 2015.

[2] W., I., Pratiwi, W., F., Mahmudy, and C., Dewi, "Implementasi Algoritma Genetika pada Optimasi Biaya Pemenuhan Kebutuhan Gizi," *Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya*, vol. 4, no. 6, 2014.

[3] N., P., H., Ali, H. Tarore, D., R., O., Walangitan, and M., Sibi "Aplikasi Metode Stepping-Stone Untuk Optimasi Perencanaan Biaya pada Suatu Proyek Konstruksi (Studi Kasus : Proyek Pemeliharaan Ruas Jalan di Senduk, Tinoor, dan Ratahan)," *Jurnal Sipil Statik*, vol. 1, no. 8, pp. 571-578, Jul. 2013.

[4] M., Priyo, dan M., R., A., Paridi, "Studi Optimasi Waktu dan Biaya dengan Metode Time Cost Trade Off pada Proyek Konstruksi Pembangunan Gedung Olah Raga (GOR)," *Semesta Teknika*, vol. 21, no.1, pp. 72-84, Mei, 2018.

[5] Indroprasto, and E., Suryani, "Analisis Pengendalian Persedian Produk dengan Metode EOQ Menggunakan Algoritma Genetika untuk Mengefisiensikan Biaya Persediaan," *Jurnal Teknik ITS*, vol. 1, Sep. 2012.

[6] D., A., Kurniawan, and D., Kriestanto, "Penerapan Naïve Bayes untuk Prediksi Kelayakan Kredit," *Jurnal Informatika dan Komputer (JIKO)*, vol. 1, no. 1, Feb. 2016.

[7] H., Leidiyana, "Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *Jurnal Penelitian Ilmu Komputer*, vol. 1, no. 1, pp. 65-76, 2013.

[8] S., Aulia, S., Hadiyoso., and D., N., Ramadan, "Analisis Perbandingan KNN dan SVM untuk Klasifikasi Penyakit Diabetes Retinopati Berdasarkan Citra Eksudat dan Mikroaneurisma," *Jurnal ELKOMIKA*, vol. 3, no. 1, Jan-Jun. 2015.

[9] R., E., Putri, Suparti, and R., Rahmawati, "Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012," *Jurnal Gaussian*, vol. 3, no. 4, pp. 831-838, 2014.

[10] M., A., Banjarsari, H., I., Budiman., and A., Farmadi, "Penerapan K-Optimal pada Algoritma KNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer FMIPA UNLAM Berdasarkan IP sampai dengan Semester 4," *Kumpulan Jurnal Ilmu Komputer (KLIK)*, vol. 2, no. 2, Sep. 2015.

[11] Suwirmayanti, N.,L.,G.,P., "Implementation of K-Nearest Neighbor for Car Selection Recommendation System," *Techno.COM*, vol. 16, no. 2, pp. 120–131, Mei. 2017.

[12] Yustanti, W., "Algoritma K-Nearest Neighbor untuk Memprediksi Harga Jual Tanah," *Jurnal Matematika, Statistika, dan Komputasi*, vol. 9, no. 1, pp. 57–68, Jul. 2012.

[13] T., Rosandy, "Perbandingan Metode Naïve Bayes Classifier dengan Metode Decision Tree (C4.5) untuk Menganalisa Kelancaran Pembiayaan (Studi Kasus: KSPPS/BMT AL-FADHILA)," Jurnal ITM Darmajaya, vol. 2, no. 1, Mei. 2016.

[14] D., Ardiada, P., A., Ariawan, and M., Sudarma, "Evaluation of Supporting Work Quality using K-Means Algoritm," *International Journal of Engineering and Emerging Technology*, vol. 3, no. 1, Jan-Jun. 2018.