Sainmatika: Jurnal Ilmiah Matematika dan Ilmu Pengetahuan Alam
Volume 22 No. 2, December 2025 DOI 10.31851/sainmatika.v22i2.20427
https://jurnal.univpgri-palembang.ac.id/index.php/sainmatika

Multi-Objective KNN Algorithm Based on the Level Expenditure in the Snack Food Group for Classification of City in Indonesia

Agistha Srikandi, Eka Susanti*, Novi Rustiana Dewi, Oki Dwipurwani, Evi Yuliza, Indrawati
*e-mail: eka_susanti@mipa.unsri.ac.id

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Sriwijaya University

ABSTRACT

Classification is an analytical process aimed at grouping objects into specific categories based on the relationships between attributes. Classification can be used in planning the supply and distribution of specific products. This study aims to classify cities in Indonesia based on the level of expenditure in the snack food group by applying the Multi Objective Particle Swarm Optimization K-Nearest Neighbor (KNN-MOPSO). The results of this classification can be used to see an overview of the level of food expenditure in each district/city in Indonesia. The KNN-MOPSO algorithm is solved using Python programming. The RandomizedSearchCV module is used to determine the best *k* parameters and the Distributed Evolutionary Algorithms in Python (DEAP) module is used for the MOPSO solving stage. The ratio of training data and testing data used to 80% and 20% from 496 dataset. Based on testing data, there are 81 districts/cities in the low category and 19 districts/cities in the high category.The accuracy results obtained are 96% in very good criteria and F1 score 93.50%. Based on the data, the application of the Multiobjective KNN algorithm with the addition of the searchCV and DEAP modules can improve model performance.

Keywords: classification, KNN-MOPSO, searchCV, DEAP

INTRODUCTION

Classification is an analytical process that aims to group objects into predetermined categories based on the relationships between data attributes (Rukmana et al., 2022). Economic growth and changes in modern lifestyles have driven an increase in spending on snacks as part of daily needs (Ahmada et al., 2023). Per capita expenditure is the total cost of food consumption spent by all members of a household within a certain period of time. The period of per capita expenditure is usually measured in one week, one month, or one year (Nugroho et al., 2020). Snack foods are foods consumed between main meals (Kaluku et al., 2023). The classification results can be used as a basis for decision making for snack food providers in Indonesia, formulating more effective and efficient marketing strategies (Lina & Wati, 2023). Several algorithms that can solve the problem of classifying including Decision Tree algorithm, Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbor (KNN), and so on (Widyatmoko et al., 2022). The study (Zulfallah, 2022) applied the KNN algorithm to predicting the graduation rate of students at UIN Syarif Hidayatullah Jakarta. In the research (Ahluna et al., 2023) analyzed user sentiment on Twitter, YouTube, and Instagram regarding the issue of abolition the National Examination (UN) using the KNN algorithm. (Safitri et al., 2024) used the KNN algorithm to classify social assistance recipients in Serunai Village to ensure that the assistance was targeted appropriately.

The effectiveness of the KNN algorithm in classification analysis is greatly influenced by the accuracy,

p-ISSN 1829 586X
e-ISSN 2581-0170

precision, recall, and F1 score. In some cases, we expect the optimal performance of our model to be achieved simultaneously. The problems with several objective functions to be achieved simultaneously are called multi-objective (MO) optimization. Research by (Alamsyah & Sari, 2023) used MO problems to find optimal values for health services in hospital inpatient rooms. Research by (Jatnika & Nababan, 2022) applied MO problems to optimize investor profits and losses in optimal stock portfolios. One of the algorithms for solving MO optimization problems is Particle Swarm Optimization (PSO).

In related research, PSO has been used by (Febiani et al., 2024) to optimize the scheduling of teaching and learning activities, helping to simplify and maximize schedule preparation. Research (Wei et al., 2024) shows that the application of the PSO algorithm can significantly increase annual net income and reduce carbon emissions in power plants.

In this study, KNN algorithm was developed with two objective functions, including maximizing accuracy and F1 score for the classification problem of districts and cities based on snack food expenditure levels. The calculations were solved using Python programming with the addition of the searchCV and DEAP modules to optimize model performance. This study aims to classify cities in Indonesia based on the level of expenditure in the snack food group by applying the Multi Objective Particle Swarm Optimization K-Nearest Neighbor (KNN-MOPSO). The results of this classification can be used to see an overview of the level of food expenditure in each district/city in Indonesia.

METHOD

There are two stages in multi-objective classification. The first stage is classification using the KNN method. The second stage is optimization of the objective function using the PSO algorithm. The following provides an explanation of each stage.

1. Classification using KNN algorithm.
   The classification stage using the KNN method starts from data labeling. Dividing the dataset into training and test data with a ratio of 80%:20%. Calculating the z-score normalization value in the training and test data, determining the K values, calculating the distance between the data using the Euclidean distance formula. (Zulfallah, 2022), (Safitri et al., 2024).
2. Solve the MO problem using the PSO algorithm (Febiani et al., 2024). PSO algorithm is used to determine the optimal solution in multi-objective cases.
   There are two objective function.
   Maximum $f_1 = \frac{TP+TN}{TP+TN+FP+FN}$

   Maximum $f_2 = 2\frac{(\text{presisi})(recall)}{\text{presisi}+recall}$

   Where,

   Precision$= \frac{TP}{TP+FP}$

   Recall $= \frac{TP}{TP+FN}$

   TP : *True Possitive*
   TN : *True negative*
   FP : *False positive*
   FN : *False negative*
3. Single objektive Model
   The following is a general form of a single objective model with constraints being the weight values of each objective function (Anggi et al., 2022).
   $max\ fitness = w_1 f_1 + w_2 f_2$

   Subject to
   $$0 < w_1 < 1$$
   $$0 < w_2 < 1$$

RESULT AND DISCUSSION

The data used in this study is secondary data sourced from the Central Statistics Agency (BPS). The research data on the average per capita expenditure per rupiah per week on food and beverages from all cities/districts in Indonesia. The data in this study consisted of 514 district/city. In the calculations, the number of classified district/cities was taken as 496. Data deletion is carried out for districts/cities whose data is incomplete. This was due to incomplete data in several cities. The districts/cities that are ignored in the calculations are West Southeast Maluku, Sula Islands, Taliabu Island, Jaya Wijaya, Paniai, Puncak Jaya, Yahukimo, Bintang Mountains, Tolikara, Mamberamo Raya, Nduga, Lanny Jaya, Central Mamberamo, Yalimo, Puncak, Dogiyai, Intan Jaya, and Deiyai. The research attributes consist of sweet bread and other breads, cookies, biscuits, semprong, cakes (layered cakes, Bika Ambon, Lemper, etc.), fried foods (tofu, tempeh, bakwan, fried bananas), fried foods, green bean porridge, children's snacks, crackers/chips, cooked processed meat (sausages, nuggets, smoked meat, etc). Complete data can be seen in the link https://www.bps.go.id/id/statistics-table/2/MjEyMyMy/rata-rata-pengeluaran-perkapita-seminggu--menurut-kelompok-makanan-minuman-jadi-per-kabupaten-kota--rupiah-kapita-minggu-.html.

The next step is data labeling. Data labeling refers to the food poverty line. Low income is defined as spending less than IDR 1,824. Medium income is defined as spending equal to IDR 1,824 and high income is defined as spending more than IDR 1,824. The labeled data is given in Table 1.

Table 1. Labeled Data

Criteria	District/Cities
Low (L)	376
Medium (M)	0
High (H)	120

The mean and standard deviation were calculated from the values of all districts/cities for each category of snacks. The mean and standard deviation for each attributes in the training data are given in Table 2.

Table 2. The mean and standard deviation

Attribute	Mean	Standard Deviation
sweet bread and other breads	1904,42	891,30
Cookies, biscuits, Semprong	1297,47	596,55
cakes (layered cakes, Bika Ambon, Lemper, etc.)	2103,16	1093,35
fried foods (tofu, tempeh, bakwan, fried bananas)	2729,08	934,95
fried foods	258	217,12
green bean porridge	487,81	419,16
children's snacks, crackers/chips,	2324,88	988,61
cooked processed meat (sausages, nuggets, smoked meat, etc)	881,02	662,71

Mean and standard deviation are used for Z-Score normalization calculation of training and testing data. Classification results, the actual data and classification results are given in Table 3.

Table 3. Actual dan Predicted Data

No.	District	Actual Data	Predicted
1	Simeulue	L	L
2	South Aceh	L	L
3	Bireuen	H	H
⋮	⋮	⋮	⋮
36	Bangka	L	H
⋮	⋮	⋮	⋮

44	Rembang	H	L
45	Temanggung	L	L
⋮	⋮	⋮	⋮
99	Sarmi	L	L
100	Supiori	L	L

For 100 testing data, the confusion matrix of KNN MOPSO results is given in Table 4.

Table 4. Confusion Matrix

		Predicted		Total
		L	H	
Actual	L	79	2	81
	H	2	17	19
Total		81	19	**100**

A total of 200 iterations were performed. The accuracy value obtained was 96%, the best *k* is 3, F1 score, precision and recall were 93,50227%. By taking the weight of the first objective function as 0.7 and the weight of the second objective function as 0.3, the accuracy and F1 score results are in the very good criteria. The selection of the first objective function weight is greater than the second objective function due to the classification process prioritizing the accuracy value.

CONCLUSION

This study classifies districts and cities in Indonesia based on expenditure levels using the KNN-MOPSO algorithm. The KNN-MOPSO algorithm is used for classification with more than one objective function. The classification process uses Python programming. The addition of the searchCV module increases the effectiveness of the KNN method in obtaining the best *k* value. Based on the data used, the objective function of maximizing accuracy and maximizing F1 score can be achieved simultaneously. The addition of the DEAP module to MOPSO for solving multi-objective models can improve model performance. This can be seen from the results of accuracy and F1 score in the very good criteria. Based on testing data, there are 81 districts/cities in the low category and 19 districts/cities in the high category. Model performance assessment is not only from the accuracy and F1 score values, in future research, Precision and Recall assessments can be added as objective functions. In this study, deletion was performed for missing data. Further research can be carried out using imputation techniques for missing data. This study only performed classification and did not further analyze the impact of the classification results across various fields. The study could be continued with spatial analysis.

REFERENCES

Ahluna, F., Tutuarima, C. J., & Santoso, I. (2023). K-Nearest Neighbor Method for Sentiment Analysis Regarding the Elimination of National Exams. *Jurnal Ikraith-Informatika*, *7*(2), 1–6.

Ahmada, M. A. A. N., Imaretha, V. P., Munir, A. A. S., & Fitria, D. R. (2023). The Effect of Per Capita Expenditure and Human Development Index on Gross Regional Domestic Product in East Java. *SINDA: Comprehensive Journal of Islamic Social Studies*, *3*(3), 81–86. https://doi.org/10.28926/sinda.v3i3.1172

Alamsyah, A., & Sari, R. F. (2023). *Multi-Objective Optimization in Health Services in Hospital Inpatient Rooms*. *6*(1), 826–838.

Anggi, A., Prihandono, B., & Kiftiah, M. (2022). Solving Multi-Objective Integer Linear Programming Problems Using Weighting Methods and Variable Reduction Methods (Case Study: Anong Chip SME in Singkawang). *Epsilon: Jurnal Matematika Murni Dan Terapan*, *16*(2), 146–161.

https://doi.org/10.20527/epsilon.v16i2.6927

Febiani, A., Widodo, A. M., Anwar, N., Sekti, B. A., & Yulfitri, A. (2024). Implementation of the Particle Swarm Optimization (PSO) Algorithm for Teaching and Learning Scheduling. *IKRA-ITH Informatika : Jurnal Komputer Dan Informatika*, *8*(1), 152–161. https://doi.org/10.37817/ikraith-informatika.v8i1.3210

Jatnika, N. S., & Nababan, E. S. M. (2022). Multi-objective Optimization in the Formation of an Optimal Stock Portfolio Using Value at Risk (VAR) Measurement. *Journal of Fundamental Mathematics and Applications (JFMA)*, *5*(1), 52–66. https://doi.org/10.14710/jfma.v5i1.14662

Kaluku, K., Junieni, Mahmud, & Ruaida, N. (2023). Factors Affecting Snacking Habits on Academic Achievement and Nutritional Status (Literature Study). *Global Health Science*, *8*(2), 69–74.

Lina, I. R., & Wati, D. C. (2023). Classification of Per Capita Expenditure in Three Provinces of Sulawesi using K-Nearest Neighbor. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, *16*(1), 395–406. https://doi.org/10.36456/jstat.vol16.no1.a7193

Nugroho, D., Asmanto, P., & Adji, A. (2020). Leading Indicators of Poverty in Indonesia: Application to the Short-Term Outlook. In *The Nasional Team For The Acceleration Of Poverty Reduction (TNP2K)* (Vol. 92, Issue 11).

Rukmana, S. Z. H., Aziz, A., & Harianto, W. (2022). Optimization of the K-Nearest Neighbor (KNN) Algorithm with Normalization and Feature Selection for Liver Disease Classification. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 439–445. https://doi.org/10.36040/jati.v6i2.4722

Safitri, I., Satria, A., & Badri, R. M. (2024). Implementation of the K-Nearest Neighbor Algorithm in the Classification of the Human Development Index of South Sumatra Province in 2023. *Digital Transformation Technology (Digitech)*, *4*(2), 768–775. https://doi.org/https://doi.org/10.47709/digitech.v4i2.4614

Safitri, N., Kusnandar, D., & Martha, S. (2024). mplementation of the K-Nearest Neighbor Algorithm with Z-Score Normalization in the Classification of Social Assistance Recipients in Serunai Village. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, *13*(1), 99–106.

Wei, D., Zhang, Z., Zhang, W., Yang, Y., & Yang, Z. (2024). Optimization of multi-energy complementary power generation system configuration based on particle swarm optimization. *Energy Reports*, *12*(June), 2257–2269. https://doi.org/10.1016/j.egyr.2024.08.026

Widyatmoko, K., Sugiarto, E., Muslih, M., & Budiman, F. (2022). Optimization of the K-Nearest Neighbor Method with Particle Swarm Optimization for Batik Image Recognition with Geometric Ornaments. *Jurnal Informatika Upgris*, *8*(1), 123–127. https://doi.org/10.26877/jiu.v8i1.11705

Zulfallah, F. H. (2022). *Implementation of the KNN Algorithm in Measuring the Accuracy of Student Graduation at UIN Syarif Hidayatullah Jakarta*. Universitas Islam Negeri Syarif Hidayatulah.