

Aplication of KNN

by Tony Yulianto

Submission date: 25-Jul-2023 02:15PM (UTC+0700)

Submission ID: 2136508015

File name: fkkn.pdf (2.23M)

Word count: 2112

Character count: 10283

PAPER • OPEN ACCESS

Application of FKNN on Positioning of Potential Salt in Coastal South Beach of Madura

To cite this article: Tony Yulianto *et al* 2018 *J. Phys.: Conf. Ser.* **974** 012010

[View the article online](#) for updates and enhancements.

You may also like

- Detection of GHRIA1 gene polymorphism and its association with body weight of Madura cattle in Indonesian Beef Cattle Research Station
Hartati, N H Krisna, F Firdaus et al.
- Influencing factors of madura cattle farmers' courage to face risks of breeding
P R Novanda and M Khaliq
- Model For Marketing Strategy Decision Used On Multicriteria Decision Making: A Case Study In Batik Madura Industry
I D Anna, I Cahyadi and A Yakin

Application of FKNN on Positioning of Potential Salt in Coastal South Beach of Madura

Tony Yulianto^{1*}, Rica Amalia², Kuzairi³

^{1*, 2, 3} Department of Mathematics, Mathematics and Science Faculty, Islamic University of Madura, Indonesia

Email: ^{1*}toniyulianto65@gmail.com

Abstract. The sea holds a lot of potential to be utilized, including salt which is one of the most important needs in everyday life. In terms of quality domestic salt production is still not meet the health requirements, especially salt produced from salt farmers. So, there is an effort to optimize salt production by improving the quality of sea water, especially in terms of salinity. Madura is one of the areas in Indonesia that can be developed as a center of national salt, especially the southern region of Madura. That opportunities can be developed considering that until now Indonesia is still imported salt, whereas the potential of salt in Madura is very large. Therefore, in this study used FKNN method to simplify the process of classification for decision making about the position of salt content in the beach of Madura that has the potential to be a quality salt. From all the south beach of Madura, 12 data were taken (each district had 3 coastal data) with each data is tested as many as 8 test. Then from 12 data is taken 11 data for the training and the rest in testing using FKNN. So that, the results are obtained that almost all beaches on the south beach of Madura are in the category of Kw III, except for 2 beaches that are Pang Pajung beach (located at latitude -7,207,387 and Coordinates longitude 113,006943) located in the category Kw I and Gading beach (located at latitude -7,185039 coordinates and longitude coordinates 112,915988) located in the category Kw II and the two beaches are located in Bangkalan district.

2 Introduction

Indonesia as an archipelagic country with a long coastline of 81,000 kilometers is a coastal and marine area that has a huge variety of biological and non-biological resources. With the oceans constituting 70% of the total area of the archipelago, the sea has a lot of potential to be utilized, among others are salt (Yulianto, Komariyah, & Ulfaniyah, 2017).

Salt is one of the most important needs in everyday life. The making of salt is mostly done traditionally by smallholders besides by industrial salt companies. In terms of quality of domestic salt production is still not meet health requirements, especially salt produced from salt farmers, because the quality of salt is generally below the quality of II according to SNI / SII specifications No.140-76 (Hasan, 2011).

Madura is one of the areas in Indonesia that can be developed as a national salt center. That opportunities can be developed considering that until now Indonesia is still imported salt, whereas the potential of salt in Madura is very large. Of the 2.4 million tons of national salt demand, currently only 1.4 million tons are supplied. Of that amount, Madura's salt contribution is 600,000 ton. This data shows that almost half of the national salt demand is supplied by Madura, optimize salt production by



improving seawater quality especially in terms of its salt content. The phenomenon of most of the traditional salt processing on the coast of Madura does not see the proportion of salt content in the region (Hasan, 2011). The research uses the south coast of Madura, because the southern coast becomes the main crossing point for frequent transportation by many vehicles, thus becoming the main access in industry.

The background of the use of the method among others is based on previous research stated that k-nearest neighbor fuzzy method (FKNN) proved to consider the ambiguous nature of neighbors, where ambiguous neighbors do not have an important role in the classification process, this is due to the membership value of each class giving an instance to a class (Putri, Cholissodin, & Setiawan, 2015). Therefore, in this study, the authors will use the FKNN method in the classification process for decision making about the position of salt content in the coast of Madura that has the potential to be a quality salt.

2. Preliminaries

This section presents briefly the basic concept related to salt and FKNN method and implementation of FKNN.

2.1. Definition of Salt

Salt is a crystal-shaped white solid object that is a collection of compounds with the largest portion of Sodium Chloride (> 80%) as well as other compounds, such as Magnesium Chloride, Magnesium Sulphate, and Calcium Chloride. Natural salt sources derived from seawater, salt lake water, in-ground deposits, salt mines, groundwater sources. Components - these components have an important role for the human body, so it takes salt consumption with the right size to support human health (Hasan, 2011).

According to Desrosier (1988), there are three main sources of salt, namely (Mohi, 2014):

1. Solar salt is salt obtained by evaporation from brine either from the sea or from the saline lake of the land.
2. A salt or salt mine is a salt normally declared as a salt rock, derived from mining operating as deep as a thousand feet or more below the earth's surface.
3. The salt obtained from the evaporation of the sun contains chemical waste and salt-tolerant halophilic microbes. Mineral salt or salt sources are generally free from contamination of these organisms.

2.2. Quality of Salt Based on SNI Salt

Salt can be divided into several categories such as; very good category, good and medium. It is said to be very good if it contains NaCl > 95%, good NaCl content 90-95%, and moderate NaCl between 80-90% but the preferred is the salt content above 95% (Mayasari & Lukman, 2004).

Salt quality based on NaCl content can be mapped (see table 1).

Table 1 Salt Quality based on NaCl (Usman, 2017)

No.	Substance	Ks
Quality I	NaCl > 98%	Maximum water content of 4%
Quality II	94,4% < NaCl < 98%	Maximum water content of 5%
Quality III	NaCl < 94%	Water content of > 5%

Every salt producer is required to obtain Indonesian National Certificate (SNI) before trading (see table 2). According to the Indonesian National Certificate (SNI) no. 01-3556 of the year 1994 and the Decree No. 77/1995 the salt used must contain iodine of 30-80 ppm (Hartati, Supriyo, & Zainuri, 2014).

Table 2 Standard of Salt SNI (Hartati, Supriyo, & Zainuri, 2014)

No.	Parameter	Unity	SNI Standard
1.	Flavors	-	Salty
2.	Smell	-	Normal
3.	Color	-	White
4.	NaCl (kitchen salt)	w/w	Min 94,73
5.	Water (H ₂ O)	w/w	Max 7
6.	Fe ₂ O ₃ (iron oxide/sludge/sand)	mg/kg	Max 100
7.	Ca (calcium) atau Mg (magnesium)	w/w	Max 1,00
8.	SO ₄ (sulphate)	w/w	Max 1,00
9.	Impurity	w/w	Max 2,00
10.	Iodium as KIO ₃	w/w	30-80
11.	Pb (lead)	mg/kg	Max 10
12.	Cu (copper)	mg/kg	Max 10
13.	Hg (mercury)	mg/kg	Max 0.1
14.	As (arsen)	mg/kg	Max 0.1

2.3. Fuzzy Set

The firm set (crisp) A is defined by the items in the set. In the set of crisp, the membership value of an item x in a set A ($\mu_A(x)$) has two possibilities, namely (Yulianto, Komariyah, & Ulfaniyah, 2017):

1. One (1), which means that an item becomes a member in a set.
2. Zero (0), which means that an item is not a member in a set.

The Fuzzy set is based on the idea of extending the range of characteristic functions such that the function will include real numbers at intervals $[0,1]$. The membership value indicates that an item in the universe of speech is not only 0 or 1, but also the value that lies between them. In other words, the truth value of an item is not just right or false. Value 0 indicates false, value 1 indicates true, and there are still values that lie between right and wrong (Yulianto, Komariyah, & Ulfaniyah, 2017).

2.4. Data Mining

Data mining is a term used to describe the discovery of knowledge in a database. Data Mining is a process that uses statistical techniques, mathematics, artificial intelligence and machine learning to identify useful information and knowledge assembled from various large databases (Wijaya, Mardji, & Furqon, 2015).

In this study used normalization min-max. Normalization of min-max is calculated (see equation 1). According to Wijaya et al (2015) normalization min-max has the advantage of maintaining relationships in the data. And has a function that unites units of various attributes (Wijaya, Mardji, & Furqon, 2015):

$$V' = \frac{V - \min_i A}{\max_i A - \min_i A} \tag{1}$$

with:

- V' : The result of normalization whose value ranges between 0 and 1
- V : The value of attribute A to be normalized
- $\min_i A$: The minimum value of an attribute, A
- $\max_i A$: The maximum value of an attribute, A

2.5. Fuzzy K-Nearest Neighbor

Fuzzy K-Nearest Neighbor (FKNN) is a Fuzzy set theory that generalizes the classical K-NN theory by defining the membership value of a data in each class. The formula used to obtain the results of FKNN with the form (Selly, Ridok, & Muflikhah, 2013):

$$u(x, y_i) = \frac{\sum_{j=1}^K w(x_j, y_i)^* d(x, x_j)^{\frac{-2}{(m-1)}}}{\sum_{j=1}^K d(x, x_j)^{\frac{-2}{(m-1)}}} \quad (2)$$

with $u(x, y_i)$ is the membership value of x class c . K is the number of nearest neighbors used. $w(x_j, y_i)^*$ is the value of neighboring member's membership in the neighboring K in the y_i class, the value 1 if the training data x_k belongs to class c_i or 0 if not belong to class c_i , for $d(x, x_k)$ is the distance from data x to x_k data in the nearest neighbor's k , m is the weight exponent of magnitude $m > 1$ (Selly, Ridok, & Muflikhah, 2013).

The membership value of a data in a class is greatly influenced by the distance of that data to its nearest neighbor, the closer to its neighbor or the greater the membership value of that data to its neighboring class, and vice versa. The distance is measured by N dimension (feature) data. The measurement of distance (inapplication) of two data used in FKNN (see equation 2) is generalized by (Putri, Cholissodin, & Setiawan, 2015):

$$d(x_i, x_j) = \left(\sum_{l=1}^N |x_{il} - x_{jl}|^p \right)^{\frac{1}{p}} \quad (3)$$

with N is the dimension (number of features) of data. While p is the distance determinant used, if $p = 1$ then the distance used is Manhattan, if $p = 2$ fed the distance used is Euclidean, if $p = \infty$ then the distance used is Chebychev. Although FKNN uses membership values to declare membership data on each class, but to provide final output, FKNN must still provide the final class of predicted results. For this purpose, FKNN selects the class with the largest membership value in the data (Putri, Cholissodin, & Setiawan, 2015).

Prediction Algorithm with FKNN (Putri, Cholissodin, & Setiawan, 2015):

1. Normalize the data using the largest and least values of data on each feature (see equation 1).
2. Find the nearest neighbor's K for the test data x using equation (3).
3. Calculate the membership value $u(x, y_i)$ using equation (2) for each i with $1 \leq i \leq C$.
4. Take the largest value $v = u(x, y_i)$ for all $1 \leq i \leq C$, with C is the number of classes.
5. Give the class label v to the data u in x that is y_i .

2.6. Data Collection

The data collected from all districts in Madura located on the south coast of Madura as much as 12 data per each point (beach) with sampling per each district there are 3 samples. We just use 12 data not more, because we have no much cost to do more and variable test tools also require a large cost. Because of that reason, we can only retrieve the data is not too much and we limit the data retrieval south coast in each district only as much as 3. The data collected in the form of seawater on the coast by taking a bottle of drinking water and then tested in the lab for the test level of each concentration. The concentrations of the test were 8 concentrations: Mg (magnesium), Cu (copper), SO4 (sulfate), As (arsenic), Hg (mercury), Pb (lead), NaCl (kitchen salt), and Ca (calcium).

2.7. Salt Quality Criteria

Based on coastal water data from laboratory test result, to be able to determine the quality of salt contained in sea water because it has not done the processing process in this study does not directly use the existing SNI salt, but using the approach approach value almost close to the actual SNI Salt. So the salt quality criterion data is determined first to facilitate the calculation process using the FKNN method which then can be obtained coastal location containing adequate salt quality which is almost equal to the best salt quality on the market. The SNI Salt is used for salt that has been done processing and distillation of seawater in several stages, so in this data do not use SNI Salt directly. The salt quality criterion data to the selected concentration (see table 3).

Table 3 Data on Salt Quality Criteria

Quality	NaCl	Mg	Ca	SO ₄	Pb	Cu	Hg	As
Kw I	10	300	200	0	0	0	0	0
Kw II	27	400	300	1	5	5	0,05	0,25
Kw III	30	1000	400	2	10	10	0,1	0,5

2.8. FKNN Process

As for the FKNN algorithm, so we take 11 data from 12 data to do the training test. So the first step by entering the value of m . In this research is inserted value $m = 5$, then after that do normalization of data by using equation (1), for example:

- At Mg test in 12 beaches it is known that the smallest value of data is 270 and the biggest value is 1200 so that the normalization result of Mg Patoroman beach (south coast of Bangkalan) is worth $\frac{330-270}{1200-270} = \frac{60}{930} = 0,064516129$.
- On the NaCl test at 12 beaches it is known that the smallest value of data is 4 and the largest value is 30 so that the normalization result of NaCl beach Patoroman (south coast of Bangkalan) is worth $\frac{29-4}{30-4} = \frac{25}{26} = 0,9615384615$, and so on.
- For weighted values also performed normalization of data also, for example on the test Mg on Kw I obtained $\frac{10-10}{30-10} = \frac{0}{20} = 0$, on Kw II obtained value $\frac{27-10}{30-10} = \frac{17}{20} = 0,85$ and on Kw III we get value $\frac{30-10}{30-10} = \frac{20}{20} = 1$, and also done for other weight values.

Then look for the distance using equation (3) and the value $w(x_k, y_i)^*$ from equation (2). From the result of normalization then do a distance search such as used $p = 2$ (Euclidean distance) and $w(x_k, y_i)^*$. Example:

- For the Mg test when y_1 (on Kw I) on Patoroman beach the distance value:

$$d_{\text{Patoroman}} = \sqrt{(330 - 330)^2 + (330 - 1200)^2 + (330 - 960)^2 + \dots} \quad (4)$$

Equation (4) is calculated continuously to 12 points then obtained value $d_{\text{Patoroman}}$.

- Whereas to obtain $w(x_k, y_i)^*$ from equation (2) for test Mg y_1 (at Kw I) on Patoroman beach is: $w^* = 0$.

The result of the value u of equation (2) on the south coast of Madura is obtained by combining the values of w^* and d from all points. After finish, we can do testing for the remaining data in the same way.

3. Result And Discussions

The simulation result after testing on all the data using matlab for south beach of Madura using FKNN method obtained the largest u result (Rank u) (see table 4).

Table 4 Ranking Results u Using FKNN On the South Beach

Beach Name	Latitude Coordinate	Longitude Coordinate	Rank u
A. Bangkalan District			
1. Patoroman Beach	-7,214059	113,040249	3
2. Pang Pajung Beach	-7,207387	113,006943	1
3. Gading Beach	-7,185039	112,915988	2
B. Sampang District			
1. Taddan Beach	-7,215940	113,283964	3
2. Camplong Beach	-7,218862	113,319411	3
3. Tanjung Beach	-7,217936	113,394611	3
C. Pamekasan District			
1. Branta Beach	-7,223800	113,449749	3
2. Jumiang Beach	-7,236570	113,544467	3
3. Talang Siring Beach	-7,141684	113,593849	3
D. Sumenep District			
1. Prenduan Beach	-7,110188	113,676640	3
2. Aing Panas Beach	-7,110020	113,689512	3
3. Kalianget Beach	-7,050465	113,932520	3

The result (see table 4) shows that almost all beaches on the south beach of Madura are in the category of Kw III, except for 2 beaches that are Pang Pajung beach (located at latitude -7,207,387 coordinates and longitude coordinates 113,006943) located in the Kw I and Gading beach (located at latitude -7,185039 coordinates and longitude coordinates 112,915988) located in the Kw II category and both beaches are located in Bangkalan district. From the data we can get as much 83.33% of beaches are in cluster 3 (Kw III), 8.33% of beaches are in cluster 1 (Kw I) and 8.33% another of beaches are in cluster 2 (Kw II).

4. Conclusions

The conclusion of this research (see table 4) shows that almost all beaches on the south beach of Madura are in the category of Kw III, except for 2 beaches that are Pang Pajung beach (located at latitude -7,207,387 coordinates and longitude coordinates 113,006943) located in the Kw I and Gading beach (located at latitude -7,185039 coordinates and longitude coordinates 112,915988) located in category Kw II and second beach located in Bangkalan district. Because the data is still small, so the future can be done quite a lot to the validity of the data can be processed better.

5. Acknowledgments

Praise the presence to Allah SWT and the author also thanked Ristek Dikti who has allocated research funding for later can be utilized in this research. In addition, the authors also thank the Dean of Mathematics and head of Mathematics department of UIM who has provided support in the form of initial assistance to conduct research, as well as dean of FKIP UIM who has assisted in the provision of laboratory for testing the concentration of sea water data in this study so that this study to run smoothly.

References

- [1] Hartati, R., Supriyo, E., & Zainuri, M. (2014, April). Yodisasi Garam Rakyat Dengan Sistem Screw Injection. *GEMA TEKNOLOGI*, 17(4), 160-163.
- [2] Hasan, T. I. (2011). Identifikasi Sosial Ekonomi Dan Ketenagakerjaan Petani Garam Di Kabupaten Bireuen. *Sains Riset*, 1(2).
- [3] Mayasari, V. A., & Lukman, R. (2004). *Studi Peningkatan Mutu Garam Dengan Pencucian*. Surabaya: Fakultas Teknologi Industri Jurusan Teknik Kimia Institut Teknologi Sepuluh Nopember.
- [4] Mohi, R. A. (2014). Analisis Potensi Pengembangan Tambak Garam Di Desa Siduwonge Kecamatan Randangan Kabupaten Pohuwato. Dalam *Tesis* (hal. 12-20). Gorontalo: Universitas Negeri Gorontalo.
- [5] Putri, I. R., Cholissodin, I., & Setiawan, B. D. (2015). Optimasi Metode Adaptive Fuzzy K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Klasifikasi Status Sosial Ekonomi Keluarga. *Repository Jurnal Mahasiswa PTIIK UB*, 5(3).
- [6] Selly, Y., Ridok, A., & Muflikhah, L. (2013). Perbandingan K-Nearest Neighbor Dan Fuzzy K-Nearest Neighbor Pada Diagnosis Penyakit Diabetes Melitus. *Repository Jurnal Mahasiswa PTIIK UB*, 2(10).
- [7] Usman, M. A. (2017). Studi Eksperimen Penggunaan Air Garam Sebagai Sumber Energi Alternatif. Dalam *Skripsi* (hal. 15-18). Kendari: Jurusan Teknik Mesin, Fakultas Teknik, Universitas Halu Oleo.
- [8] Wijaya, H. T., Mardji, & Furqon, M. T. (2015). Penerapan Fuzzy K-Nearest Neighbor (FKNN) Untuk Diagnosa Penderita Liver Berdasarkan Indian Liver Patient Dataset (ILPD). *Repository Jurnal Mahasiswa PTIIK UB*, 5(9).
- [9] Yulianto, T., Komariyah, S., & Ulfaniyah, N. (2017). Application of Fuzzy Inference System by Sugeno Method on Estimating of Salt Production. *AIP Conference Proceedings* (hal. 1-7). Surabaya: American Institute of Physics. doi:10.1063/1.4994442

Application of KNN

ORIGINALITY REPORT

14%

SIMILARITY INDEX

3%

INTERNET SOURCES

14%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

- 1** I Gede Surya Rahayuda, Ni Putu Linda Santiari. "Crawling and cluster hidden web using crawler framework and fuzzy-KNN", 2017 5th International Conference on Cyber and IT Service Management (CITSM), 2017
Publication 7%

- 2** Tony Yulianto, Siti Komariyah, Nurita Ulfaniyah. "Application of fuzzy inference system by Sugeno method on estimating of salt production", AIP Publishing, 2017
Publication 4%

- 3** Hartati, Angga Ardhati Rani Hapsari, Sri Putri Rahmadani, Retno Widiyawati. "Identification of growth hormone receptor (GHR|AluI) gene polymorphism in grati-madura cattle and pamekasan-madura cattle population", IOP Conference Series: Earth and Environmental Science, 2021
Publication 3%

Exclude quotes On

Exclude bibliography On

Exclude matches < 3%