Jurnal Teknologi Informasi dan Pendidikan
Volume 15, No. 2, September 2022
https://doi.org/10.24036/tip.v15i2

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Comparison of Machine Learning Algorithms for Classification of Ultraviolet Index

Alfin Syarifuddin Syahab[1*]✉, Rianto[2]
[1]Magister of Information Technology, Universitas Teknologi Yogyakarta, Indonesia
*Corresponding Author: 6220211009.alfin@student.uty.ac.id

Article Information

Article history:

No. 692
Rec. February 03, 2023
Rev. February 10, 2023
Acc. March 25, 2023
Pub. March 29, 2023
Page. 132– 146

Keywords:

- *Ultraviolet*
- *Recommender System*
- *Collaborative filtering*
- *K-nearest Neighbor*
- *Support Vector Machine*

ABSTRACT

Indonesia is a tropical climate country which has the potential for high intensity of sunlight exposure. Several types of exposure that receives are ultraviolet rays. According to the National Agency for Meteorology, Climatology and Geophysics, it provides information regarding the impact on human activities in an ultraviolet index which has a risk scale. The aim of this research was creating a recommender system based on the ultraviolet index category in providing daily activity advice to users. The methods used the K-nearest Neighbor algorithm and Support Vector Machine with a Collaborative Filtering Model-Based approach that could recommend items based on the results of a model trained to identify input data patterns. The stages carried out in this study included data collection, data pre-processing, data division into test data and train data, dataset testing, analysis of the results of models that had been trained in the accuracy values using the algorithm tested. The results of the confusion matrix calculation produced test evaluations in accuracy values, precision values, and recall values. The comparison of result had the highest performance in K-nearest Neighbor with an accuracy value of 99.69%, a precision value of 99.00%, and a recall value of 96.20%. In research used the Support Vector Machine showed the lowest performance with an accuracy value of 97.91%, a precision value of 93.20%, and a recall value of 86.40%.

1. INTRODUCTION

Indonesia is a tropical climate country that has the potential for exposure to sunlight with high intensity. Several types of exposure from sunlight that reaches the earth's surface are ultraviolet rays. UV rays are classified into UV A, UV B and UV C which are distinguished based on the difference in wavelength. Ultraviolet radiation that has an effect on the skin is UV B radiation and has the strongest effect in causing photodamage to the skin [1]. Sun exposure can be beneficial in increasing vitamin D production based on medical research. However, excessive UV exposure can cause serious health disease such as sunburn, skin damage, and skin cancer [2]. According to the Meteorology, Climatology and Geophysics Agency (BMKG), the UV index has a risk scale from solar ultraviolet radiation. Based on the risk, the UV index is categorized into five, namely low has scale from 0 up to 2, moderate has scale from 3 up to 5, high has scale from 6 up to 7, very high has scale from 8 up to 10, and extreme has scale 11 and higher [3]. The UVA component of the solar radiation spectrum, with wavelengths ranging from 315 to 400 nm, can be measured outside using the SUV-A radiometer. Since UV A may travel through clouds and glass windows, it can penetrate the second layer of skin and contribute to various types of solar damage that cause wrinkles and premature aging of the skin. UV A typically makes up 95% of the UV radiation that reaches the earth's surface from clear sky. [4]. The UVB component of the sun radiation spectrum, with wavelengths ranging from 280 to 315 nm, can be measured outside with the SUV-B radiometer according to its design and optimization. Most sunburns are intimately linked to skin cancer and DNA damage in skin cells, and unprotected skin can become burned in just 15 minutes from UV B, which makes up 5% of the UV radiation that reaches the earth's surface from the clear sky. [5].

Warna Skala	UV index	Kategori	Himbauan
Hijau	0 - 2	"Low" (risiko bahaya rendah)	• tingkat bahaya rendah bagi orang banyak. • kenakan kacamata hitam pada hari yang cerah. • gunakan cairan pelembab tabir surya SPF 30+ bagi kulit sensitif. • permukaan yang cerah, seperti pasir, air, dan salju, akan meningkatkan paparan UV.
Kuning	3 - 5	"Moderate" (risiko bahaya sedang)	• tingkat bahaya sedang bagi orang yang terpapar matahari tanpa pelindung • tetap di tempat teduh pada saat matahari terik siang hari. • kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan. • oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat. • permukaan yang cerah, seperti pasir, air, dan salju, akan meningkatkan paparan UV.
Oranye	6 - 7	"High" (risiko bahaya tinggi)	• tingkat bahaya tinggi bagi orang yang terpapar matahari tanpa pelindung, diperlukan pelindung untuk menghindari kerusakan mata dan kulit. • kurangi waktu di bawah paparan matahari antara pukul 10 pagi hingga pukul 4 sore. • tetap di tempat teduh pada saat matahari terik siang hari. • kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan. • oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat. • permukaan yang cerah, seperti pasir, air, dan salju, akan meningkatkan paparan UV.
Merah	8 - 10	"Very high" (risiko bahaya sangat tinggi)	• tingkat bahaya tinggi bagi orang yang terpapar matahari tanpa pelindung, diperlukan tindakan pencegahan ekstra karena kulit dan mata dapat rusak rusak dan terbakar dengan cepat. • minimalkan waktu di bawah paparan matahari antara pukul 10 pagi hingga pukul 4 sore. • tetap di tempat teduh pada saat matahari terik siang hari. • kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan. • oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat. • permukaan yang cerah, seperti pasir, air, dan salju, akan meningkatkan paparan UV.
Ungu	>11	"Extreme" (risiko bahaya sangat ekstrem)	• tingkat bahaya ekstrem bagi orang yang terpapar matahari tanpa pelindung, diperlukan semua tindakan pencegahan karena kulit dan mata dapat rusak rusak dan terbakar dalam hitungan menit. • hindari paparan matahari antara pukul 10 pagi hingga pukul 4 sore. • tetap di tempat teduh pada saat matahari terik siang hari. • kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan. • oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat. • permukaan yang cerah, seperti pasir, air, dan salju, akan meningkatkan paparan UV.

Figure 1. Appeal based on UV index level

BMKG provides an appeal to the public regarding the impact of the level of UV radiation exposed to the earth's surface with a UV index value as shown in Figure 1. The UV index is a number without units to explain the level of exposure to ultraviolet radiation related to human health. By knowing the UV index, we can monitor the level of ultraviolet light that is beneficial and that can provide harm [6]. From the parameterization of the UV index and appeals, it will become material in making a recommendation system.

Collaborative Filtering is a recommender system technique that utilizes rating information from multiple users to predict item ratings for certain users. The recommender system has been implemented. The hospital applies collaborative filtering to classify illnesses based on symptoms and advise patients on the best doctor [7]. The outputs of models that have been trained to find patterns in the input data are used to make recommendations for products using the model-based approach to collaborative filtering [8]. Collaborative Filtering that combines with machine learning is less time consuming and reliable [9]. In recent years, machine learning has dramatically changed the structure of recommendation systems and presented more opportunities to improve recommender system performance. Recent developments in machine learning-based recommender systems have received significant attention in achieving high recommendation quality. The K-nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms could make a school graduate recommender system produce an accuracy performance of 0.606 [10].

Machine learning is a technique of learning that derives from data and produces future predictions. Machine learning is divided into three groups, according to recent research: Reinforcement learning, supervised learning, and unsupervised learning [11]. Labeling is used to carry out the category of supervised learning. The two categories of supervised learning are classification and regression. Predicting the class of labeled data is the goal of classification. Regression seeks to forecast the description of the regression relationship as a floating-point number in the form of a number in a predetermined range from the data [12].

This study uses a classification technique to classify UV index level categories to provide recommendations in the form of appeals to daily activities from the impact of UV rays. A classification is a work product that has a data object value that may be entered into one of the many classes that are available [13]. The classification technique is used to predict the class on the label by classifying data based on the training set on the data table which consists of several attributes and one label [14]. Linear Regression, Naive Bayes Classifier, Perceptron, Support Vector Machine, Quadratic Classifiers, Decision Trees, Random Forests, and more algorithms are used in supervised machine learning [15]. Another algorithm for classification is the K-nearest neighbor [16]. The K-nearest neighbor technique and the Support Vector Machine were both used in this study's tests. The K-nearest neighbor (KNN) algorithm is an algorithm that projects learning data into a dimensional space that represents the features of the data. KNN is included in the instance-based learning group. The KNN method is a lazy technique used in classifying closely spaced data [17].

As a kernel-based machine learning model for classification and regression applications, Vapnik introduced the Support Vector Machine (SVM) algorithm. Linear classifier is the fundamental SVM principle. In contrast, kernel technique can be added to a workspace with a high-density workspace for troubleshooting non-linear issues [18]. SVM

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

has attracted the attention of the data mining, pattern recognition, and machine learning communities in recent years because of the remarkable generalizability with optimal solutions and discriminatory power. [19].

In the literature review there is literature on calculating the UV index for health appeals and there is also literature on the use of classification algorithms in machine learning in determining the classification of objects. Utilizing observations of solar UV radiation from satellite products used to create UV index climatology at local noon, the UV index study was conducted. A tendency of a considerable rise in the UV index between 2004 and 2020 was found as a result of research using OMI satellite products gathered on the campus of King Abdul Aziz University in Saudi Arabia to estimate changes in UV exposure throughout the 2004-2020 period [2]. Subsequent research is the measurement of the UV index using an analog UV index sensor measurement that converts UV radiation into an analog voltage.

Distribution of UV index based on time span was obtained. Cloud cover causes a decrease in the UV index by four levels and an attenuation of the UV Index by some materials. Based on the data collection sample, it was obtained that the safe time for sunbathing was in the range before 10.00 WIB and after 14.00 WIB [3]. KNN, Naive Bayes, and Decision Tree techniques are used in research on classifying water quality. The KNN approach is the most accurate method for classifying data, with an accuracy rate of 86.88% compared to Decision Tree's 80.84% and Naive Bayes' 63.60% [14]. In this study, a UV index recommender system was created to provide daily activity advice to users. The recommender system is created using a model-based collaborative filtering method assisted by a machine learning classification algorithm in the algorithm of K-nearest Neighbor and Support Vector Machine which produced higher accuracy performance.

2. RESEARCH METHOD

In the methodology section, it is explained about the process of the recommender system stages starting from the preparation of the dataset, testing the algorithm used, and the performance results from the test with the accuracy value obtained. The stages of the process are outlined in the form of a flowchart so that it can be easily understood.

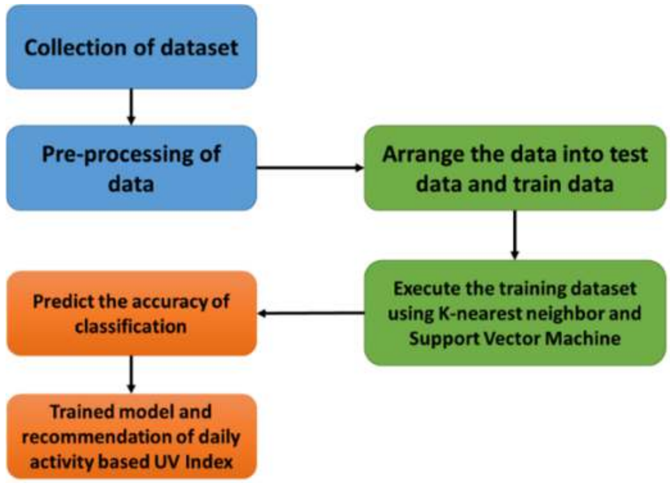

Figure 2. Flowchart on algorithm testing

Figure 2 shown a flowchart in building a machine learning-based recommendation system that requires several steps, including; collection of data, pre-processing of data, compiling data into test data and train data, training dataset testing using KNN and SVM, analysis of the accuracy value of the classification results using the algorithm tested, and the last is the result in the form of a model that has been trained along with recommendations on activities daily based on UV index level. The testing of the algorithms executed using the RapidMiner Studio. It is the analysis tool for data mining that is a standalone piece of software. Additionally, RapidMiner Studio serves as an environment for machine learning, data mining, text mining, and predictive analytics [20]. Data processing is carried out quantitatively as data values in the form of counts or numbers where each dataset has a numerical value associated with data obtained from measurement sensors. This data is quantifiable information that can be used for mathematical calculations and statistical analysis, so that decisions in real cases can be made.

The data collection of UVA and UVB was obtained from in-situ measurements using the UVA and UVB radiometer instrumentation from the Global Atmosperic Watch Observation Station in Palu City, Central Sulawesi as shown in Figure 3.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Figure 3. Measurement of UV A and UV B radiation using a radiometer in Palu City, Central Sulawesi

The UV-A and UV-B data used are daily data at local time intervals from 06.00 – 18.00 in January, February and March 2021 with measurements every one minute. This data can be used as material for data processing which will become a dataset in training data and testing data.

	A	B	C
1	Local Date Time	UV A	UV B
2	2022-01-01 09:00:00	29.8	0.8
3	2022-01-01 09:01:00	33.2	0.9
4	2022-01-01 09:02:00	38.3	1
5	2022-01-01 09:03:00	31.1	0.9
6	2022-01-01 09:04:00	23.8	0.8
7	2022-01-01 09:05:00	24.2	0.8

Figure 4. Raw data of UV A and UV B radiation measurements

Raw data obtained from measurement equipment collected as many as 2155 lines as shown in Figure 4. The raw data shows local time and UVA and UV radiation values in units of W/m^2. The data collected will be used as material for further calculations in determining the UV index as material for building a recommendation system.

Data pre-processing is a crucial step in the data processing process that converts raw data, sometimes referred to as unstructured data, gathered from numerous sources into information that may be used for further processing. The calculation of the weighting factor is the method used to derive the UV index. The UV index value is obtained from the UV erythermal value of the measurement results of the tool which is calculated using the weighing factor for UV-A and UV-B according to the CIE spectral action function. Since the weighing factor is wavelength dependent, UV radiation is received higher on surfaces with longer wavelengths. Figure 5 shows UV radiation classified between 100 nm and 400 nm into three spectral bands such as UV-A with a wavelength of 320-400 nm, UV-B, 280-320 nm. UV-A radiation accounts for a total of 95% of the sun's UV radiation that reaches the earth's surface [21]. UV-A penetrates more into surfaces than UV-B. However, energy is

Jurnal Teknologi Informasi dan Pendidikan
Volume 15, No. 2, September 2022
https://doi.org/10.24036/tip.v15i2

inversely proportional to wavelength. UV-A has lower energy than UV-B. As a result, the weighing factor for UV-B is higher than for UV-A.

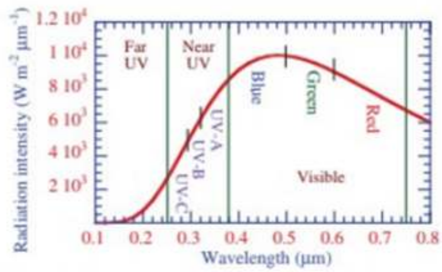

Figure 5. Effect of radiation intensity on wavelength

The radiometer operates in the spectral radiation range for both types of UV. However, the data reported includes only the integrated intensity at which UV irradiation was measured. Since the chosen wavelength can be arbitrary, we provide some assumptions which serve as our reference for the calculations. For UV-A, although the total intensity received at the surface is lower, the wavelength of 325 nm was chosen to compromise the higher contribution of UV radiation to the calculations. For UV-B, the chosen wavelength was 305 nm, which is near the UV-A spectral range with a lower weighing factor compared to the shorter wavelengths but arrives at the surface in a larger proportion. The weighing factors for the UV spectra at 305 nm and 325 nm are 0.22 and 0.0029, respectively. These factors are associated with the calculation of erythermal UV intensity. Equation 1 is the formula showing how the UV index is calculated [22].

$$UV\ Index = \frac{erythermal\ UV\ A + erythermal\ UV\ B}{0.025} \quad (1)$$

The two erythermal UVs are both expressed in W/m2. The denominator of 0.025 W/m2 is the standard increment value that corresponds to how much total UV is potentially damaging to living tissue, or in other words an increase on one scale of the UV index is equivalent to 25 m^2/W exposure to UV radiation.

	A	B	C	D	E	F	G
1	UV A	UV B	UV A erythermal	UV B erythermal	UV Index	Kategori	Rekomendasi
2	0.4	0	0.00116	0	0.0464	Low	bahaya rendah bagi orang banyak, kenakan kacamata hitam pada hari yang cerah, gunakan cairan pelembab tabir surya SPF 30+ bagi kulit sensitif, permukaan yang cerah, seperti pasir, air, dan salju, akan meningkatkan paparan UV.
3	11.5	0.2	0.03335	0.044	3.094	Moderate	bahaya sedang bagi orang yang terpapar matahari tanpa pelindung, tetap di tempat teduh pada saat matahari terik siang hari, kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan, oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat.
4	21.5	0.4	0.06235	0.088	6.014	High	bahaya tinggi bagi orang yang terpapar matahari tanpa pelindung, diperlukan pelindung untuk menghindari kerusakan mata dan kulit, kurangi waktu di bawah paparan matahari antara pukul 10 pagi hingga pukul 4 sore, tetap di tempat teduh pada saat matahari terik siang hari, kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan, oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat.
5	27.6	0.6	0.08004	0.132	8.4816	Very High	bahaya sangat tinggi bagi orang yang terpapar matahari tanpa pelindung, diperlukan tindakan pencegahan ekstra karena kulit dan mata dapat rusak rusak dan terbakar dengan cepat. minimalkan waktu di bawah paparan matahari antara pukul 10 pagi hingga pukul 4 sore, tetap di tempat teduh pada saat matahari terik siang hari, kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan, oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat.
6	39.4	1.1	0.11426	0.242	14.2504	Extreme	bahaya ekstrem bagi orang yang terpapar matahari tanpa pelindung, diperlukan semua tindakan pencegahan karena kulit dan mata dapat rusak rusak dan terbakar dalam hitungan menit. Hindari paparan matahari antara pukul 10 pagi hingga pukul 4 sore. Tetap di tempat teduh pada saat matahari terik siang hari. kenakan pakaian pelindung matahari, topi lebar, dan kacamata hitam yang menghalangi sinar UV, pada saat berada di luar ruangan. oleskan cairan pelembab tabir surya SPF 30+ setiap 2 jam bahkan pada hari berawan, setelah berenang atau berkeringat.

Figure 6 Pre-processing results of the recommender system dataset based on the UV index

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

In this test the data will be prepared to carry out the classification stage. Classification can be done using a supervised learning algorithm. The classification algorithm divides the data into two categories: train data and test data. While test data is used to assess and evaluate the performance of the model discovered at the testing stage, train data is used to instruct the algorithm on how to develop a suitable model. As was done in reference to earlier research journals, the distribution of the number of train data and test data in this study uses a percentage of 70:30 [23]. In tests that use a dataset of 30% testing data and 70% training data, the accuracy value can reach 90.33% [24]. The total UV index data used as train data in this study was 1689. The total UV index data used as test data in this study was 646.

3. RESULTS AND DISCUSSION

In this section, the research findings are discussed while also providing a thorough discussion. Results shown in tables, graphs, figures, and other formats. the division of the discussion into several sections.

3.1. Testing Data based on Machine Learning

Testing using the KNN and SVM algorithms is a system development methodology after data pre-processing in analyzing processing data quantitatively. Techniques for processing quantitative data using mathematical equations and statistics are then assisted by the Rapid Miner application to calculate the accuracy, precision, and recall values for the results of data classification in the algorithm testing process.

3.1.1. K-nearest Neighbor

The number of closest neighbors, denoted by the variable parameter k in the K-nearest Neighbor (KNN) algorithm The KNN algorithm locates a data point or nearest neighbor given a query from the training data set. According to the shortest distance from the query point, the nearest data point is located. It uses a majority vote rule to determine which class appears the most after locating the k nearest data points. The final classification of the query will be based on the class that appears the most [25]. In the KNN algorithm, a formula is needed to calculate the shortest distance, the distance value in the KNN method can be calculated using the Euclidean distance formula [26]. The following is the Euclidean distance formula in Equation 2

$$d(x, y) = \sqrt{\sum_{i=1}^{m}(x_i + y_i)^2} \quad (2)$$

Information:
x = data 1
y = data 2
d = Euclidean distance
i = iteration

Where x = x1, x2 ,…xn , and y = y1 , y2…,yn . n represents the attribute values of the two records.

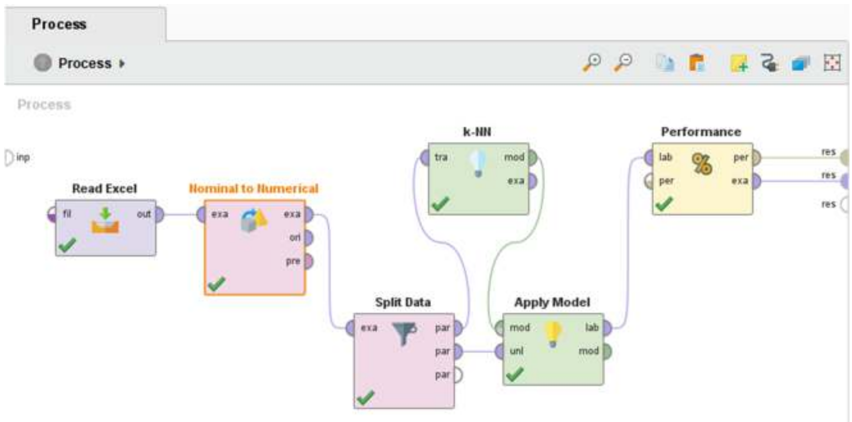

Figure 7. Data Test Design using K-Nearest neighbor

Figure 7 shows the testing process on the KNN algorithm. The dataset is obtained as input in excel data format. Then the nominal data on the category attribute is converted into a numerical form. Data is separated by a ratio of 0.7 as train data and 0.3 as test data. Then the stage is continued by selecting the KNN model to classify the data. After that the process continues with the apply model to apply the classification model. After obtaining the model from test data and train data, performance is obtained which shows the accuracy, precision, and recall values of the model being trained and tested.

3.1.2. Support Vector Machine

Support Vector Machine (SVM) is a machine learning technique based on statistical learning theory that can identify predictive systems and handle non-linear regression in a spherical space. This algorithm is also flexible enough to be used to the field of data modeling, where data classification and analysis follow a regressive pattern. SVM is an algorithm for producing predictions in the context of regression or classification [27]. In this test using SVM with the Radial Basis Function (RBF) kernel function, also known as the Gaussian kernel function. Where γ is a parameter to set the distance. The RBF kernel function can be formulated by Equation 3 [28].

$$K(x_i , x_j) = \exp\left(-\gamma \left|X_i^T X_j\right|^2\right), \gamma > 0 \quad (3)$$

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Information:

K = kernel function
γ = distance
Xi = i^{th} data
Xj = j^{th} data

There are several steps for classifying with the SVM algorithm, including determining the hyperplane or dividing line between two support vectors, determining the margin or distance line between support vectors and hyperplanes, and mapping support vectors into a class in the same dimension class.

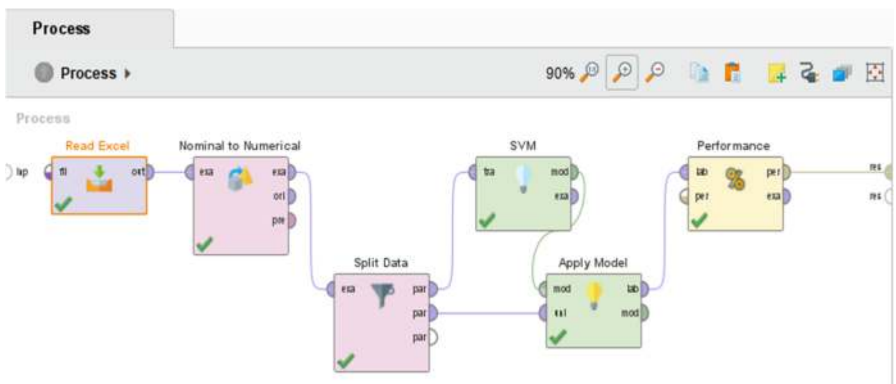

Figure 8. Data Testing Design using Support Vector Machine

Figure 8 shows the testing process on the SVM algorithm. The dataset is obtained as input in excel data format. The table data obtained has category attributes with nominal types which are converted to numerical attributes to make it easier to classify using SVM. Then the data is separated by a ratio of 0.7 as train data and 0.3 as test data. The next step is to choose the SVM model to classify the data. After that the process continues with the apply model to apply the classification model. After obtaining the model from test data and train data, performance is obtained which shows the accuracy, precision and recall values of the model being trained and tested.

3.2. Accuracy Measurement with Confusion Matrix

The confusion matrix is a classification method evaluation method based on the accuracy of the classification results. Classification accuracy will affect classification performance. The confusion matrix is a prediction matrix that will be compared with the original categories in the input data [29]. In order to compare the classification results produced by the system with the actual classification results used the confusion matrix. The importance of the confusion matrix will provide information on how well the model that has been made previously through existing accuracy measurements to find out how accurate the model that has been made is. The classification model's performance on a set of test data with known true values is described by the confusion matrix. In the case of multi-class classification, the metrics specified for binary classification, do not fully apply [30]. Table 1 shown the confusion matrix for multi-class problems with k class numbers [31].

Table 1. Confusion Matrix in Multi-Class Classification

		Predicted Class			
		Class 1	Class 2	…	Class k
Actual Class	Class 1	f_{11}	f_{12}	…	f_{1k}
	Class 2	f_{21}	f_{22}	…	f_{2k}
	...	…	…	…	…
	Class k	f_{k1}	f_{k2}	…	f_{kk}

In calculating accuracy, precision, and recall can be calculated using Equation 4, Equation 5, and Equation 6 [32].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\% \quad (4)$$

$$\text{Class Precision} = \frac{TP}{FP+TP} \text{ x } 100\% \quad (5)$$

$$\text{Class Recall} = \frac{TP}{TP+FN} \text{ x } 100\% \quad (6)$$

Information:
TP : The amount of accurate positive data.
TN : The amount of accurate negative data.
FN : The amount of false negative data
FP : The amount of false positive data.

In this study there were five different classes namely low, moderate, high, very high, and extreme. The results will discuss the accuracy, precision, and recall values obtained using the confusion matrix in the multi-class classification of the K-NN and SVM algorithm calculations.

3.3. Accuracy results of the K-nearest neighbor algorithm

Testing of the machine learning algorithm produces class recall, class precision, and accuracy values. The method for determining the accuracy value is a multi-class confusion matrix. Table 2 shows the results of the classification test with the UV index confusion matrix using the K-NN algorithm.

Table 2. The results of the UV index data classification in the confusion matrix using K-NN

	true Low	true Moderate	true High	true Very High	true Extreme	Class precision
pred. Low	203	0	0	0	0	100 %
pred. Moderate	0	40	2	0	0	95.24 %
pred. High	0	0	9	0	0	100 %
pred. Very High	0	0	0	29	0	100 %
pred. Extreme	0	0	0	0	363	100 %
Class Recall	100 %	100 %	81.82 %	100 %	100 %	

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Based on the calculation of Equation 4 for the confusion matrix shows the accuracy for the whole class is:

$$\text{Accuracy} = \frac{644}{644+2} x\ 100\% = 99.69\ \%$$

Then the precision value can be calculated using equation 5 accumulated to the total precision value in the calculation below.

$$\text{Precision} = \frac{1+0.95+1+1+1}{5} = 99.00\ \%$$

Furthermore, the recall value can be calculated using Equation 6 which is accumulated to the total recall value in the calculation below.\

$$\text{Recall} = \frac{1+0.81+1+1+1}{5} = 96.20\ \%$$

The results of the classification consisted of 644 data that were correctly classified and there were 2 high data which were classified as moderate data. From the calculation of Equation 4, Equation 5, and Equation 6, an accuracy value of 99.69%, a precision value of 99.00%, and a recall value of 96.20% are obtained. Table 3 shown the results of the classification test with the UV index confusion matrix using the SVM algorithm.

Table 3. The results of the UV index data classification in the confusion matrix using SVM

	true Low	true Moderate	true High	true Very High	true Extreme	Class precision
pred. Low	187	4	0	0	0	97.91%
pred. Moderate	0	56	6	0	0	90.32%
pred. High	0	0	15	1	0	93.75%
pred. Very High	0	0	6	49	0	89.09%
pred. Extreme	0	0	0	8	315	97.52%
Class Recall	100%	93.33%	55.56%	84.48%	100%	

Based on the calculation of Equation 4 for the confusion matrix shows the accuracy for the whole class using SVM is.

$$\text{Accuracy} = \frac{622}{622+25} x\ 100\% = 96.13\ \%$$

Then the precision value can be calculated using Equation 5 which is accumulated to the total precision value in the calculation below.

$$\text{Precision} = \frac{0.97+0.90+0.93+0.89+0.97}{5} = 93.20\ \%$$

Furthermore, the recall value can be calculated using Equation 6 which is accumulated to the total recall value in the calculation below.

$$\text{Recall} = \frac{1+0.93+0.55+0.84+1}{5} = 86.40\ \%$$

The results of the classification consisted of 622 data that were correctly classified and 25 data that had misclassifications. At 4 moderate data are classified as low data. Then there are 6 high data classified as moderate data. There is 1 very high data classified as high data. Then there are 8 very high data classified as extreme data. From the calculation of Equation 4, Equation 5, and Equation 6, an accuracy value of 96.13%, a precision value of

93.20%, and a recall value of 86.40% are obtained. Table 4 shown the results of testing the KNN and SVM algorithms based on the results of accuracy, precision and recall values.

Table 4 Comparison results of KNN and SVM testing

Machine Learning Algorithm	Accuracy Value (%)	Precision Value (%)	Recall Value (%)
K-NN	99.69 %	99.00 %	96.20 %
SVM	96.13 %	93.20 %	86.49 %

Applying the K-NN and SVM algorithms, a comparative analysis of accuracy testing on UV index data classification was successfully completed. The test results reveal that SVM has an accuracy value of 96.13%, a precision value of 93.20%, and a recall value of 86.40% while K-NN has an accuracy value of 99.69%, a precision value of 99.00%, and a recall value of 96.20%.

4. CONCLUSION

The results of this study indicate a comparative analysis of the level of accuracy of the supervised machine learning research method in classifying UV index data in the recommendation system used, namely K-nearest neighbors and Support Vector Machine. Evaluation of the classification method based on the accuracy of the classification results is done by calculating the confusion matrix. The results of the confusion matrix calculation produce class recall values and class precision values which show various levels of accuracy. The classification method that produces the highest accuracy value is K-NN which is equal to 99.69%. In research using the SVM classification method, it shows the lowest accuracy value, which is equal to 97.91%. In this study, the K-nearest neighbors and Support Vector Machine methods were used and used as much data as data, therefore suggestions for further research can use more diverse methods and add the amount of data to improve higher accuracy.

ACKNOWLEDGEMENTS

In this study the author thanks to Dr. Rianto, S. Kom., M. Eng. as a lecturer in the Intelligent Recommender System Course for the support and guidance given so that research can be completed, the authors also thank parents and families for the support in this learning activity.

REFERENCES

[1] A. Adzhani, F. Darusman e R. Aryani, “Kajian Efek Radiasi Ultraviolet Terhadap Kulit,” *Bandung Conference Series: Pharmacy,* vol. 2 no. 2, pp. 106-112, 2022. https://proceedings.unisba.ac.id/index.php/BCSP/article/download/3551/2048/.

[2] A. Addas, M. Ragab, A. Maghrabi, S. M. Abo-Dahab e E. F. El-Nob, “UV Index for Public Health Awareness Based on OMI/NASA Satellite Data at King Abdulaziz University, Saudi Arabia,” *Hindawi Advances in Mathematical Physics,* pp. 1-11, 2021. https://doi.org/10.1155/2021/2835393.

[3] S. Ristanto, C. Huda e A. F. Kurniawan, “Pengukuran Indeks Ultraviolet Matahari dan Atenuasinya oleh Beberapa Bahan untuk Rekomendasi Waktu Aman Berjemur,” *Indonesian Journal of Applied Physics,* vol. 11 no. 2, pp. 248-255, 2021. https://jurnal.uns.ac.id/ijap/article/view/52753.

[4] H. OTT, “Kipp Zonen UV A Radiometer,” 7 February 2017. [Online]. Available: https://www.kippzonen.com/Product/426/SUV-A-UVA-Radiometer#.Y8cpAnZBzIX. [Acesso em 18 January 2023].

[5] H. OTT, “Kipp Zonen UV B Radiometer,” 7 February 2017. [Online]. Available: https://www.kippzonen.com/Product/427/SUV-B-UVB-Radiometer#.Y8drT3ZBzIW. [Acesso em 18 January 2023].

[6] S. Subbidang Produksi Informasi Iklim dan Kualitas Udara, “Indeks Sinar Ultraviolet,” BMKG, [Online]. Available: https://www.bmkg.go.id/kualitas-udara/indeks-uv.bmkg. [Acesso em 18 January 2023].

[7] B. Thomas e A. K. John, “Machine Learning Techniques for Recommender Systems – A Comparative Case Analysis,” *IOP Conference Series: Materials Science and Engineering,* vol. 1085, pp. 1-14, 2020. doi 10.1088/1757-899X/1085/1/012011.

[8] S. Sauda e M. R. Ramadhan, “Implementasi Memory-Based dan Model-Based Collaborative Filtering pada Sistem Rekomendasi Sepeda Gunung,” *Syntax Literate: Jurnal Ilmiah Indonesia,* vol. 7 no. 6, pp. 7836-7848, 2022. doi http://dx.doi.org/10.36418/syntax-literate.v7i6.7559.

[9] D. Theodorus, S. Defit e G. W. Nurchayo, “Machine Learning Rekomendasi Produk dalam Penjualan Menggunakan Metode Item-Based Collaborative Filtering,” *Jurnal Informasi dan Teknologi,* vol. 3 no. 4, pp. 202-208, 2021. doi https://doi.org/10.37034/jidt.v3i4.151.

[10] S. Sharda e G. S. Josan, “Machine Learning Based Recommendation System: A Review,” *International Journal of Next-Generation Computing,* vol. 12, nº 2, pp. 134-144, 2021. doi.org/10.47164/ijngc.v12i2.200.

[11] A. Roihan, P. A. Sunarya e A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang:Review Paper,” *Indonesian Journal on Computer and Information Technology,* vol. 5 Nomor 1, pp. 75-82, 2020. doi https://doi.org/10.31294/ijcit.v5i1.7951.

[12] S. A. Rajagukguk, “Tinjauan Pustaka Sistematis: Prediksi Prestasi Belajar Peserta Didik dengan Algoritma Pembelajaran Mesin,” *Jurnal SNATi,* vol. 1 No 1, pp. 22-30, 2021.

[13] D. Rohmayani, “Analysis of Student Tuition Fee Pay Delay Prediction using Naive Bayes Algorithm with Particle Swarm Optimation Optimazation (Case Study: Politeknik TEDC Bandung),” *Jurnal Teknologi Informasi dan Pendidikan,* Vols. %1 de %213, No. 2 , pp. 1-8, 2020. doi https://doi.org/10.24036/tip.v13i2.

[14] A. Tangkelayuk e E. Mailoa, “Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree,” *Jurnal Teknik Informatika dan Sistem Informasi,* vol. 9 No. 2, pp. 1109-1119, 2022.

[15] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi e J. Akinjobi, “Supervised Machine Learning Algorithms: Classification and Comparison,” *International Journal of Computer Trends and Technology (IJCTT),* vol. 48 No 3, pp. 128-138, 2017.

[16] A. P. Permana, K. Ainiyah e K. F. H. Holle, “Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up,” *JISKa,* vol. 6 No. 3, pp. 178-188, 2021.

[17] R. Puspita e A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *Jurnal Informatika Universitas Pamulang,* vol. 5 No. 4, pp. 646-654, 2020.

[18] I. Ernawati, “Naïve Bayes Classifier dan Support Vector Machine sebagai Alternatif Solusi untuk Text Mining,” *JTIP Jurnal Teknologi Informasi dan Pendidikan,* vol. 12 no. 2, pp. 33-39, 2019. doi: https://doi.org/10.24036/tip.v12i2.219.

[19] J. Cervantesa, F. Garcia-Lamonta, L. Rodríguez-Mazahuab e A. Lopezc, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing,* vol. 408, pp. 189-215, 2020.

[20] N. L. Anggreini, “Teknik Clustering dengan Algoritma K-Medoids untuk Menangani Strategi Promosi di Politeknik TEDC Bandung,” *JTIP Jurnal Teknologi Informasi dan Pendidikan,* vol. 12 no. 2, pp. 1-7, 2019. doi: https://doi.org/10.24036/tip.v12i2.215.

[21] T. E. Arijaje, T. V. Omotosho, S. A. Akinwumi, O. O. Ometan e O. O. Fashade, “Analysis of Solar Ultraviolet radiation Index over Nigeria, West Africa,” *5th International Conference on Science and Sustainable Development ,* vol. 993, pp. 1-11, 2022.

[22] A. F. McKinlay e B. L. Diffey, “McKinlay, A.F. and Diffey, B.L. (1987) A Reference Action Spectrum for Ultraviolet Induced Erythema in Human Skin,” *CIE Journal,* vol. 6, pp. 17-22, 1987.

[23] I. Rasila e U. Ristian, “Implementasi Metode Naive Bayes Classifier pada Sistem Pengklasifikasi Berita Otomatis berbasis Website (Studi Kasus: Berita Lokal Dari Mediamassa Online Kalimantan Barat),” *Coding : Jurnal Komputer dan Aplikasi,* vol. 7 no. 2 , pp. l 49-60, 2019.

[24] M. Fatchan e H. Sugeng, “Analisa Terpilihnya Tri Rismaharini sebagi Menteri Sosial dengan Pendekatan Algorithma Naïve Bayes,” *Journal of Practical Computer Science,* vol. 1 no. 2, pp. 50-57, 2021.

[25] S. Uddin, I. Haque, H. Lu, M. A. Moni e E. Gide, “Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction,” *Scientific Reports,* vol. 12, nº 1, pp. 1-11, 2022.

[26] V. V. Pramansah, D. I. Mulyana e T. Silfia, “Analisis Perbandingan Algoritma SVM Dan KNN Untuk Klasifikasi Anime Bergenre Drama,” *Jurnal J-COM (Jurnal Informatika dan Teknologi Komputer) ,* vol. 03 no. 01 , pp. 49-55, 2022.

[27] Saikin, S. Fadli e M. Ashari, “Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results,” *JISA (Jurnal Informatika dan Sains),* vol. 04 no. 01, pp. 22-27, 2021.

[28] N. Arifin, U. Enri e N. Sulistiyowati, “Penerapan Algoritma Support Vector Machine dengan TF-IFD N-Gram untuk Text Classification,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi),* vol. 6 no. 2 , pp. 129-136, 2021 .

[29] Ainurrohmah, “Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka,” *PRISMA Prosiding Seminar Nasional Matematika ,* vol. 4, pp. 493-499, 2021.

[30] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis e N. Doulamis, “Multiclass Confusion Matrix Reduction Method and ItsApplication on Net Promoter Score Classification Problem,” *technologies,* vol. 9, pp. 1-22, 2021.

[31] S. A. A. Kharis, I. Hadi e K. A. Hasanah, “Multiclass Classification of Brain Cancer with Multiple Multiclass Artificial Bee Colony Feature Selection and Support Vector Machine,” *Journal of Physics: Conference Series ,* vol. 1417 , pp. 1-8, 2019.

[32] K. L. Kohsasih e Z. Situmorang, “Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular,” *JURNAL INFORMATIKA,* vol. 9 no. 1, pp. 13-17, 2022.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id