Available at https://journalenrichment.com/index.php/jr/

Enrichment: Journal of Multidisciplinary Research and Development

Classification of Sentiment Towards BPJS Services Using the C50 Algorithm

Amellia Fahezha Cahyaningrum[1*], Yulison Herry Chrisnanto[2], Ade Kania Ningsih[3]
Universitas Jenderal Achmad Yani, Indonesia[1,2,3]
[*]Email: amel@gmail.com[1*]

ABSTRACT

Social media provides a timely source of public feedback on services delivered by the Social Security Administering Body for Health (BPJS Health), a State-Owned Enterprise responsible for Indonesia's public health insurance program. This study aimed to evaluate the ability of the C5.0 algorithm to classify positive and negative sentiment toward BPJS services in Twitter data. Applied quantitative research with an experimental text-classification design was conducted using a secondary dataset obtained from Kaggle. The implemented database displayed 3,060 documents. Data were processed through cleaning, case folding, tokenization, filtering, stemming, and TF-IDF weighting, followed by C5.0 classification and confusion-matrix evaluation using an 80:20 split. The reported test matrix comprised 621 cases: 6 true positives, 579 true negatives, 4 false positives, and 32 false negatives. These values produced 94.2% accuracy, 60.0% precision, and 15.8% recall. Although the aggregate accuracy was high, the low recall shows that the model detected only a small proportion of the positive class and was strongly influenced by the majority class. Therefore, the current model demonstrates the technical feasibility of applying C5.0 to BPJS-related tweets but cannot yet be considered balanced or fully reliable for service evaluation. Future optimization should address class imbalance, verify dataset labeling, and report complementary metrics before the results are used to support BPJS service-improvement decisions.

Keywords: text mining, sentiment classification; health BPJS; c50 algorithm

INTRODUCTION

Digital platforms have become important channels through which citizens communicate evaluations of public services. Social-media posts provide timely and unsolicited feedback that can complement conventional surveys, while their volume and linguistic variability make manual interpretation impractical. Sentiment analysis can support transparency, citizen engagement, and service improvement, but its validity depends on the dataset, preprocessing, classifier, and evaluation metrics (Verma, 2022; Xu et al., 2022). In Indonesia, Twitter has also been used to assess government-service quality because it captures diverse public opinions at relatively low collection cost, although complaint dominance and limited representativeness require cautious interpretation (Wilantika & Wibisono, 2021). This broader context makes BPJS Health a relevant case for automated analysis of public feedback.

BPJS (Social Security Administering Body) Health is a national health insurance program in Indonesia that aims to provide affordable health services for the entire community. As a large public service, BPJS Health receives various reviews and opinions from the public regarding the quality of its services. Health services include efforts made individually or collectively in certain relationships to develop government assistance, prevent disease, and improve the healing process (Mega et al., 2020). This research aims to determine public sentiment by combining positive and negative responses to BPJS services (Similarity, 2022). In this case, sentiment classification is a way to find out a person's or group's assessment of a certain subject, topic, opinion, or opinion in data mining. Classification is one of the most important steps in grouping new data or objects into categories or labels based on certain attributes (Dan et al., 2021).

This research uses the C5.0 Algorithm method. According to Revathy & Lawrence (2019), the C5.0 Algorithm can handle large-scale data well and provides higher accuracy compared to the C4.5 algorithm. In addition, the performance of the C5.0 algorithm is faster and uses memory more efficiently, resulting in a smaller decision tree than the C4.5 algorithm (Limbong et al., 2022; Amalda et al., 2022). The rules in the C5.0 algorithm also have a lower error rate for invisible cases and a high accuracy value compared to other C4.5 algorithms. In addition, this algorithm can automatically remove useless attributes in the analysis of the data as it is a feature of today's modern techniques to help improve the performance of predictive models significantly (Technology & Science, 2022). The C5.0 algorithm has several weaknesses, including being sensitive to unbalanced or proportional data which can cause the model to tend to predict the majority class and ignore the minority class. In addition, this algorithm is not effective for handling high-dimensional data and cannot handle unstructured data such as text or image data.

The C5.0 algorithm is also susceptible to overfitting if the model is too complex and fits well on the training data so it is difficult to generalize to new testing data. Therefore, it is necessary to set the right parameters so that the results are optimal in predictive modeling using the C5.0 algorithm on certain specific datasets (Matematika, 2022). One of the classification algorithms commonly used in data mining strategies is C50. In previous research by Ridawati Manik, Pristiwanto, and Kennedi Tampubolon entitled 'Credit Collection Forecasting Using the C5.0 Algorithm', it was stated that the C5.0 algorithm can be applied well to identify problems and see the relationship between the factors that contribute to the problem (Zamasi, 2021).

Prior studies confirm the relevance of classifying BPJS-related tweets but use different analytical approaches. Mega et al. (2020) applied a support vector machine to sentiment associated with BPJS tariff increases, whereas Yunus et al. (2021) used Naive Bayes on 780 BPJS-related tweets and reported 86.25% accuracy, 84.92% precision, and 87.78% recall. Wilantika and Wibisono (2021) combined sentiment and topic classification for public-service evaluation and cautioned that complaint-heavy social-media data can distort interpretation. Systematic syntheses likewise show that no classifier is universally optimal and that imbalanced labels, preprocessing choices, and metric selection materially affect performance (Ligthart et al., 2021; Xu et al., 2022). Against this literature, the present study examines whether TF-IDF and C5.0 can classify BPJS service sentiment in an Indonesian Twitter dataset. Its novelty lies in evaluating a decision-tree-based C5.0 pipeline in this specific public-service context and interpreting the confusion matrix rather than assuming that high overall accuracy alone denotes balanced classification.

The classification stages consist of model creation, model application, and evaluation. In developing the model, training data is used which already has attributes, classes, and labels. Classification is a type of learning that aims to find relationships between input and target attributes (Westley et al., 2022). The data from this classification provides an overview of the current sentiment of the BPJS user community. This research aims to evaluate the effectiveness of the C50 algorithm, measuring the extent to which the C50 classification algorithm is effective in classifying customer sentiment towards services provided by the Social Security Administering Body (BPJS). Then identify the ability of the C50 algorithm to recognize and classify positive or negative sentiments expressed by customers regarding BPJS services (Law et al., 2019). Then identify influential service aspects. To identify the attributes or aspects of BPJS services that most influence positive and negative customer sentiment. Specific recommendations will be prepared based on sentiment analysis of customer feedback, which is expected to help BPJS respond with more appropriate solutions (Hari et al., n.d.). By providing in-depth insight through sentiment classification results,

this research is expected to help BPJS understand customer perceptions and design improvement strategies based on the feedback received.

This evaluation is urgent because aggregate accuracy can conceal failure to identify a minority sentiment class, which may cause an organization to underestimate critical public feedback. Practically, the findings show whether the current C5.0 model is sufficiently balanced for interpreting BPJS-related opinions and where resampling, class weighting, or parameter optimization may be required. Theoretically, the study contributes a context-specific assessment of a decision-tree classifier for Indonesian public-service sentiment and reinforces the need to interpret accuracy together with precision and recall. The intended implication is therefore both methodological and managerial: automated sentiment outputs should inform BPJS decisions only after minority-class performance has been validated, while future studies can use the reported pipeline and limitations for comparative evaluation.

METHOD

Research Design and Data Source

This study employed applied quantitative research with an experimental text-classification design. The unit of analysis was an individual tweet concerning BPJS services. Secondary data were obtained from Kaggle and imported into MongoDB; the database implementation displayed in this manuscript contains 3,060 documents. Because the research used an existing public dataset, it did not involve a physical research site or direct recruitment of human participants.

Research Instrument and Data Collection

The computational instrument was the web-based sentiment-classification system implemented with JavaScript and MongoDB. The application supported data upload and storage, preprocessing, TF-IDF weighting, C5.0 classification, presentation of results, and confusion-matrix evaluation. The five application functions were separately checked through black-box testing. The dataset was collected indirectly by downloading the secondary Twitter dataset from Kaggle and uploading it to the system database.

Data Analysis. Each record passed through cleaning, case folding, tokenization, filtering or stop-word removal, and stemming

TF-IDF transformed the cleaned tokens into weighted features, after which C5.0 generated a decision-tree classifier. The experiment used an 80:20 training-and-test split; the available confusion matrix reports 621 test cases. Model performance was evaluated using accuracy, precision, and recall derived from true-positive, true-negative, false-positive, and false-negative counts. Black-box testing evaluated only application functionality and was not treated as evidence of classification performance.

Current System Analysis

With the help of ongoing system analysis, the problem of disassembling a complete and real system into computer parts or components is solved to identify and evaluate the problems caused by the system, thereby leading to solutions for improvement and development in a better direction that meets technological development needs. The system that will be developed in this research does not yet have a working system because it is a new system whose data is collected via the Kaggle website and processed through the C5.0 classification algorithm and TF-IDF weighting, where the results can be used to identify sentiment.

Analysis of the BPJS Service Sentiment Classification System

System analysis is a process for evaluating and analyzing classification systems, this analysis process consists of a data collection process (selection), then a preprocessing process is carried out, which is a stage to eliminate noisy and missing values by using features such as cleaning, case folding, tokenizing, filtering, stemming. Then carry out the data transformation process using TF-IDF as a weighting for the terms. The data mining process is the application of the c5.0 algorithm such as determining root and node as initial determinants for decision trees. The last one is the validation test process or measuring the performance of the model that has been implemented.

RESULTS AND DISCUSSION

System Implementation

Software implementation of the BPJS service sentiment classification system that has been designed by UML. This software was built using the Javascript programming language and MongoDB as the database (Syuhada et al., 2021).

Interface Implementation

The interface implementation displays dashboard pages, data management, preprocessing, sentiment classification, and results.

Dashboard Implementation

The dashboard page interface is the initial display for entering the system. There are several buttons in the system. The implementation of the Dashboard page can be seen in Figure 1.

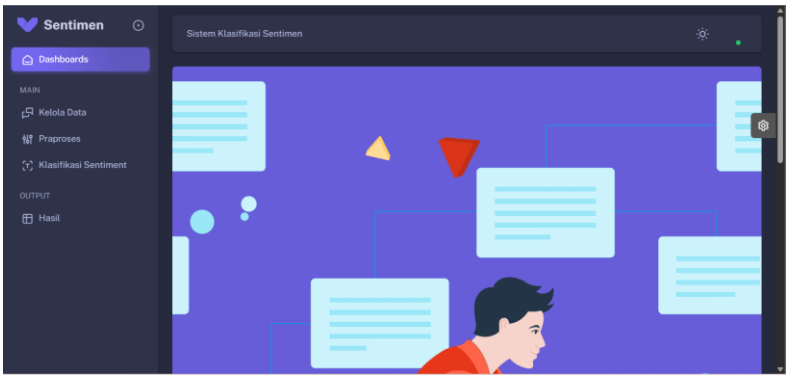

Figure 1. Dashboard Implementation

Implementation of Data Management

The data management page interface is a display that shows the data. The implementation of the Dashboard page can be seen in Figure 2.

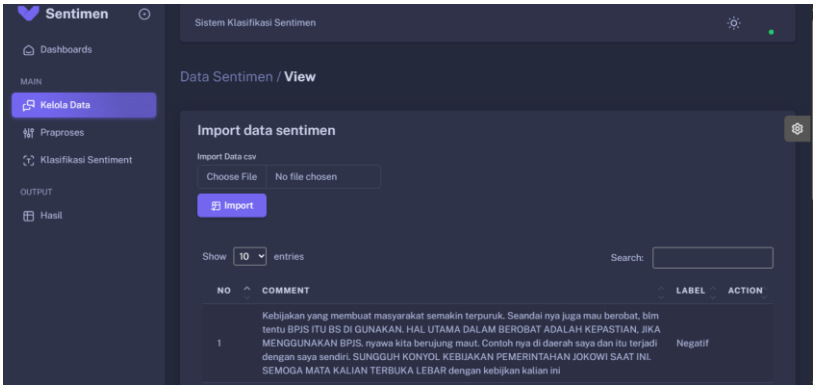

Figure 2. Implementation of the data management page

Implementation of Data Preprocessing

The preprocess data page interface is a display on the system used to retrieve sentiment data from the database and then display it (Zamasi, 2021). The implementation of the data preprocessing page can be seen in Figure 3.

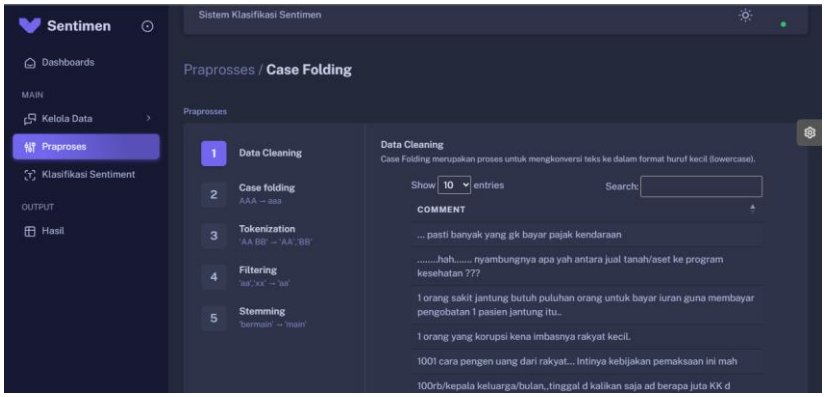

Figure 3. Implementation of data preprocessing

Implementation of sentiment classification

The sentiment classification interface is a display of the system used to retrieve data and then classify it. The implementation of the sentiment classification page can be seen in Figure 4.

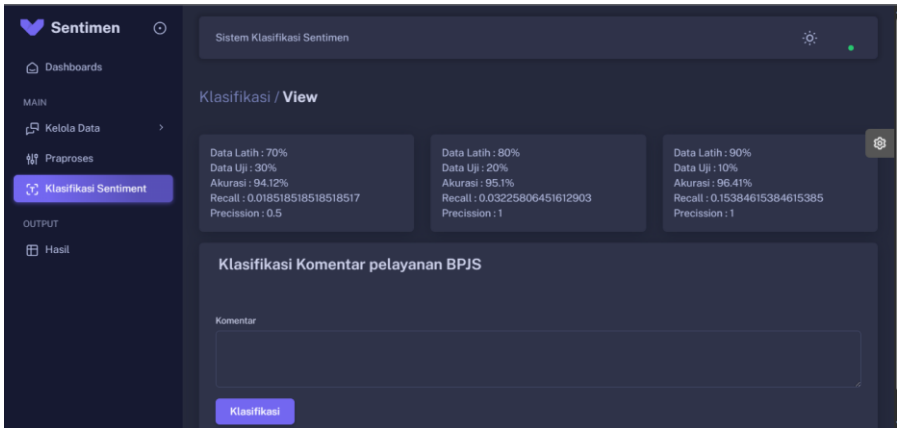

Figure 4. Implementation of sentiment classification

Implementation of Results

Implementation of results is built based on the final results that have been implemented in the program (Purba et al., 2022). Implementation results can be seen in Figure 5.

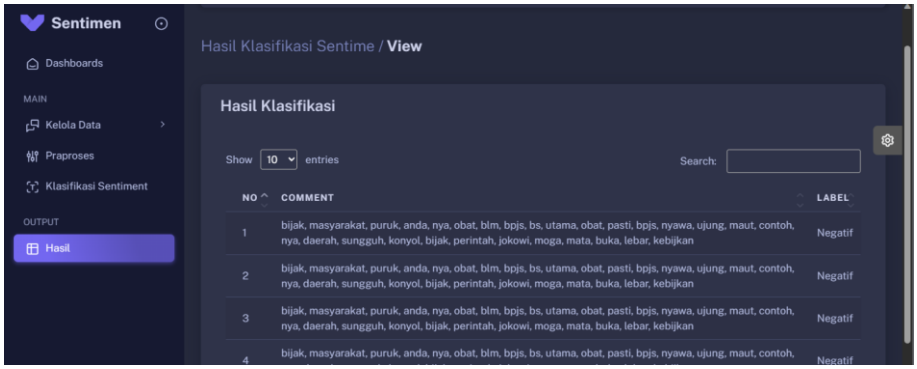

Figure 5. implementation results

Database Implementation

The database implementation is built based on the database design that was created previously. A database is needed for implementation using MongoDB software, the following is a table in the database which can be seen in Figure 6.

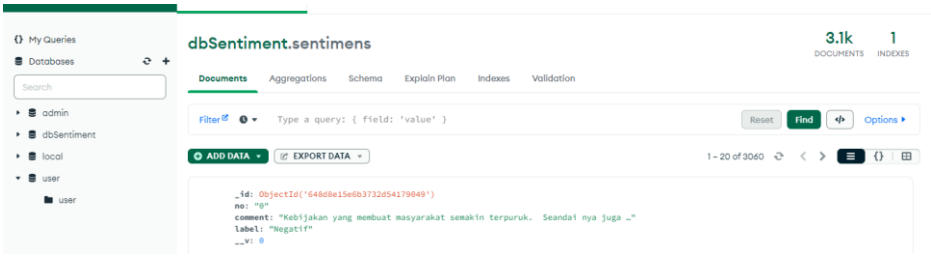

Figure 6. Database implementation

Confusion Matrix Calculation

In this research, the model evaluation process was carried out using Confusion Matrix Technology to achieve the accuracy obtained using the model in data mining. Accuracy is used in the accuracy identification process. If the indicators to be calculated about achieving accuracy are:

1. True Positive (TP) is the total positive cases included in the correct classification.
2. True Negative (TN) total negative cases included in the true classification.
3. False Positive (FP) is the total number of negative cases incorrectly classified as positive.
4. False Negative (FN) is the total number of negative cases included in the wrong classification.

Calculating the Confusion Matrix

The following is a way to calculate the confusion matrix, available in the following image:

Calculating Confusion Matrix training/data test (80/20)

Figure 7. Confusion matrix 80/20

Note. Figure 7 is retained as the original system output. The accuracy and the first precision values visible in the screenshot are inconsistent with the displayed TP, TN, FP, and FN counts. The verified calculations below (94.2% accuracy, 60.0% precision, and 15.8% recall) should be treated as authoritative, and the output code should be corrected.

a. Calculating Accuracy = (TP+TN) / (TP+TN+FP+FN)
   Accuracy = (6+579) / (6+579+4+32)
   = 0.942 x 100%
   = 94.2%
b. Calculating Precision = TP / (TP + FP)
   Precision = 6 / (6 + 4)
   = 0.6 x 100%
   = 60.0%
c. Calculating Recall = TP / (TP + FN)
   Recall = 6 / (6 + 32)

= 0.157 x 100%
= 15.8%

The 80:20 evaluation produced 6 true positives, 579 true negatives, 4 false positives, and 32 false negatives. Accordingly, accuracy was 94.2%, precision was 60.0%, and recall was 15.8%. The high aggregate accuracy must be interpreted cautiously because the model detected only 6 of 38 actual positive cases and missed the remaining 32. The combination of high accuracy and low recall indicates majority-class dominance and weak sensitivity to the minority class rather than uniformly strong classification performance.

These results partially address the objective of evaluating C5.0 effectiveness: the model classified the dominant class accurately, but its low recall shows that it was not balanced across sentiment classes. Accuracy alone is therefore insufficient for claiming effective BPJS sentiment classification. The finding is consistent with the documented sensitivity of C5.0 to imbalanced class distributions and with broader reviews that emphasize dataset and metric selection in social-media sentiment analysis (Ligthart et al., 2021; Xu et al., 2022). Furthermore, the present results contain no feature-importance, keyword, topic, or aspect-level analysis. Consequently, they do not support claims about officer professionalism, response time, facility accessibility, claims procedures, or other specific service drivers. Future experiments should report class distribution and complementary measures such as F1-score or balanced accuracy and examine resampling, class weighting, or parameter tuning before the model is used for substantive service evaluation.

Software Testing

Black-box testing was used only to verify whether the five application functions operated as designed. This functional software check complements, but does not replace, classifier evaluation based on the confusion matrix.

Test Method

Testing using black box testing techniques focuses on the functional requirements of the software and allows system analysis to obtain a set of input conditions that are carried out by the functional program (Sungkar & Qurohman, 2021). The black box testing method aims to find the cause if it occurs in:

1. Determine quality testing objectives.
2. Determine the category of quality test results
3. Designing quality tests based on use case groupings
4. Implementation of quality testing
5. Conclusions and quality testing results

Test objectives

The purpose of quality testing of the system being built can be seen in Table 1.

Table 1. Testing Objectives

No	Use case	Purpose
1	Manage data	Test the software's ability to upload review sentiment data and then store it in a database and test the ability to view review sentiment data
2	Dashboard	Test the software's ability to display the program's start menu

No	Use case	Purpose
3	Preprocessing	Testing the software's ability to process sentiment data with case folding, tokenizing, stopword removal, and stemming
4	Sentiment classification	Testing the software's ability to calculate and classify sentiment data using TF-IDF, C50 algorithm, and Cosine similarity
5	Results	Test the software's ability to produce results and perform accurate calculations using a confusion matrix

Quality Test Result Categories

In determining the category in quality testing, it is divided into 2 categories, namely:

1. Appropriate
   If the quality and function of the software being tested are to its planning and usability objectives, then it is included in the successful category.
2. Not suitable
   If the quality and function of the software being tested do not match its planning and usability objectives, then it is included in the failed category.

Quality Testing Scenarios

Software testing scenarios are designed to look for errors in the system, thereby creating quality testing scenarios with incorrect and correct steps or stages of system use. The quality testing scenario can be seen in Table 2.

Table 2. Quality Testing Scenarios

Use case	Function name	Test code	Test case
Dashboard	Dashboard	KU.101	Tests software capabilities when the user displays the home page
Manage data	Manage data	KU.102	Test software capabilities when users upload data and display data
Preprocessing	Preprocessing	KU.103	Testing software capabilities when users process sentiment data with case folding, tokenizing, stopword removal, and stemming
Sentiment classification	Sentiment classification	KU.104	Testing software capabilities when the user calculates and classifies sentiment data with TF-IDF, C50 algorithm, and Cosine similarity
Results	Results	KU.105	Test the software's capabilities when the user displays the results and performs accuracy calculations using a confusion matrix
Dashboard	Dashboard	KU.101	Tests software capabilities when the user displays the home page

Software Testing

Testing is carried out to compare the suitability of the system being built with the system being designed to produce test results that are suitable and non-compliant. Software testing can be seen in Table 3.

Table 3. Software Testing

No	Test case	Expected output	Results obtained
1	Dashboard	Users can open the home page	Displays the home page
2	Manage data	Users can upload data and view data in the system	Save uploaded data and display stored data
3	Preprocessing	The system processes data by case folding, tokenizing, stopword removal, and stemming.	Storing results in the database
4	Sentiment classification	Carry out the process of displaying the results	Display the calculation results and then save them to the database
5	Results	Displays results and accuracy tests with the c50 algorithm and confusion matrix	Displays accuracy tests for training and test data with c50 using the confusion matrix

Furthermore, the conclusion of the black box testing that has been carried out is in a table with a total of 5 functions, namely, dashboard, data management, preprocessing, classification as well as results and validation. then it can be calculated the number of functional suitability presentations in the following system (Sudarma, 2018):

Number of test codes = 5 codes
Test code with appropriate results = 5 codes
Test code with inappropriate results = 0 codes
Percentage

$$percentage\ \frac{(number\ of\ test\ codes - non - matching\ test\ codes)}{number\ of\ test\ codes}\ x\ 100\%$$

$$= \frac{(5-0)}{5}\ x\ 100\%$$

$$= 100\%$$

CONCLUSION

This study implemented a TF-IDF and C5.0 pipeline to classify sentiment toward BPJS services in Twitter data. Using the reported 80:20 evaluation confusion matrix, the model correctly classified 585 of 621 test cases, corresponding to 94.2% accuracy. However, positive-class performance was uneven: precision was 60.0% and recall was 15.8%, with only 6 true positives and 32 false negatives. Thus, the high aggregate accuracy should not be interpreted as satisfactory balanced classification because the model was strongly influenced by the majority class. The study demonstrates the technical feasibility of integrating preprocessing, TF-IDF, C5.0, and a web application, but the present evidence does not identify specific BPJS service aspects or justify aspect-based operational recommendations. BPJS sentiment outputs should therefore be used cautiously until minority-class

performance improves. Future research should verify dataset provenance and labeling, reconcile sample counts, report class distribution and complementary metrics, and compare resampling, class weighting, parameter tuning, or alternative classifiers before the model is used for service-quality decisions.

REFERENCES

Amalda, R. N., Millah, N., & Fitria, I. (2022). *Implementasi algoritma c5.0 dalam menganalisa kelayakan penerima keringanan ukt mahasiswa itk*. *7*(1), 101–116.

Bidang, P., Sains, K., Mardi, Y., Gajah, J., No, M., & Barat, S. (n.d.). *Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika*.

Dan, P., Tni, K., & Rsal, D. I. (2021). *Landosar Parsaulian. Analisis Pelayanan BPJS Bagi… | 596*. 596–604.

Hari, E., Prastyo, A., Studi, P., Teknik, S., Informasi, F. T., Hasyim, U., Studi, P., Teknik, S., Informasi, F. T., Hasyim, U., Informasi, P. S. S., Informasi, F. T., & Hasyim, U. (n.d.). *Implementasi Teknik Web Scraping Pada Situs Berita Menggunakan Metode Supervised learning Implementasi Web Scraping Pada Situs Berita Menggunakan Metode Supervised learning IGL Putra Eka Prismana Radityo Wiratsongko Implementasi Teknik Web Scraping Pada Situs Berita Menggunakan Metode Supervised learning*.

Law, A., Hukum, F., Diponegoro, U., Services, P., Publik, P., & Pendahuluan, A. (2019). *Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Sebagai Pelayanan Publik*. *2*(4), 686–696.

Limbong, J. J. A., Sembiring, I., Hartomo, K. D.,(2022). *Analysis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Analysis Of Review Sentiment Classification On E-Commerce Shopee Word Cloud Based With Naïve Bayes And K-Nearest Neighbor Methods*. *9*(2), 347–356. https://doi.org/10.25126/jtiik.202294960

Matematika, J. I. (2022). *Digital Digital Repository Repository Universitas Universitas Jember Jember Digital Digital Repository Repository Universitas Universitas Jember Jember*.

Mega, P., Dharmapatni, N., Luh, N., & Merawati, P. (2020). *Jurnal Bumigora Information Technology (BITe) Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan Jurnal Bumigora Information Technology (BITe) Jurnal Bumigora Information Technology (BITe) Jurnal Bumigora Information Technology (BITe)*. *2*(2), 105–112. https://doi.org/10.30812/bite.v2i2.904

Purba, W., Caprio, C. Di, Ryan, M., & Program, S. (2022). *Machine Translated by Google*. *10*(2), 1050–1054.

Similarity, C. (2022). *Temu Kembali Informasi Berita Kegiatan Program Studi Menggunakan Algoritma Pembobotan TF-IDF Dan Cosine Similarity*. *September*, 10–11.

Sudarma, I. M. (2018). *Implementasi Algoritma C5 . 0 pada Penilaian*. *17*(3), 1–6.

Sungkar, M. S., & Qurohman, M. T. (2021). *Penerapan Algoritma C5 . 0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer*. *5*, 1166–1172. https://doi.org/10.30865/mib.v5i3.3116

Syuhada, A. S., Simanullang, A. M., Lewa, D. S., & Marthin, S. J. (2021). *Fakultas teknik universitas maritim raja ali haji 2021*.

Westley, V., Thomas, D., & Rumaisa, F. (2022). *Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF*. *6*, 1767–1774. https://doi.org/10.30865/mib.v6i3.4218

Zamasi, N. (2021). *Implementasi Algoritma C 5 . 0 Pada Analisa Data Potensi Pertanian dan Perternakan*. *2*(4), 184–190.