

## Selection of Home Wifi Internet: Machine Learning Implementation With Decision Tree C4.5 Algorithm Method

Dewi Khairani<sup>1</sup>, Muhammad Ammaridho Romdhan Siregar<sup>2</sup>, Siti Ummi Masruroh<sup>3</sup>, Miftakhul Nuuril Azizah<sup>4</sup>

<sup>1,2,3</sup>Informatics, Science and Technology

<sup>1,2,3</sup>Islamic State University Syarif Hidayatullah Jakarta

<sup>1,2,3</sup>Ir. H. Juanda No. 95. South Tangerang, Banten, Indonesia

<sup>4</sup>Kyushu Univeristy, Japan

E-mail: <sup>1</sup>dewi.khairani@uinjkt.ac.id, <sup>2</sup>ridho.siregar18@mhs.uinjkt.ac.id, <sup>3</sup>ummi.masruroh@uinjkt.ac.id, <sup>4</sup>azizah.miftakhul.341@s.kyushu-u.ac.jp

### ABSTRACT

#### Article:

Accepted: October 13, 2022

Revised: September 02, 2022

Issued: November 15, 2022

© 2022 The Author(s).



This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

\*Correspondence Address:

dewi.khairani@uinjkt.ac.id

The multiple bandwidths that internet service providers offer make it difficult for people to choose, especially for regular people unfamiliar with the internet; therefore, most people choose because the price is reasonable. Numerous users also lament the difficulty and slow internet usage. The issue is then concentrated on internet service providers, who are thought to be poor at offering services. The quantity of bandwidth consumed, which does not correspond to the user's needs, is one factor contributing to slow internet. As a result, the appropriate bandwidth must be chosen based on the requirements of each user. Compared to other algorithms, the C4.5 decision tree method can deliver the best and correct decision, according to the current literature. As a result, this project will develop a web application based on the C4.5 decision tree algorithm that can assist in determining bandwidth and internet following community needs. Using this C4.5 Decision Tree, decisions are based on patterns identified in previously collected data. Predictions about various forms of internet use in the neighborhood may subsequently be produced from these patterns. Based on the calculation, the accuracy obtained is 0.54, or a percentage of 54%. The black box testing indicated that the bandwidth determination application was functioning correctly.

**Keywords:** *Machine Learning, Decision Tree C4.5, Home-Wifi Selection*

## I. INTRODUCTION

During the Covid-19 pandemic, we are forced to carry out various activities from home, be it shopping, work, socialization, or school, all

from home. With this, these activities require the internet as their connection access. Moreover, more people realize that the internet is necessary and has become a basic need. According to a survey conducted by the

Association of Indonesian Internet Service Providers (AAPJII), Indonesia achieved a very high internet penetration, reaching 73.7%. %, Surabaya 83%, Banten >100%” [1].

The characteristics of the cellular and WiFi networks are distinct; the cellular network can be characterized as being less able to support shared networks due to limited access, unstable network conditions, and low power consumption due to employing batteries. Long-term cell phone use can be pretty uncomfortable [2]. Compared to WiFi, which is directly connected to electricity and has a more reliable network, more devices can access it, and do not need to recharge it constantly. On the other hand, cellular use has advantages when utilized while traveling; however, with the current situation that we are compelled to work from home, such as due to the COVID-19 virus, the features of WiFi are ideal for families who stay at home [3].

According to a simple survey the author conducted, many people are unaware of the differences between different bandwidth amounts, and many even have no idea what bandwidth is. Of the 104 respondents, 81.7% were unaware of internet bandwidth, 86.5% were unaware of its size, and 97.1% were unsure how to calculate the amount of bandwidth that would best meet their needs. We can implement a decision-making system to make decisions suitable for the community, mainly residential users. DSS is a technique for analytically examining decision-making and computer programming [4][5]. With this approach, a program calculation is developed, the user will inputs the data, and the computer automatically bases the decision on the algorithm we programmed. The decision-making process usually uses the KNN[6], Decision Tree[7], Naïve Bayes [8], SVM [9], Neural Network[10], or Fuzzy methods[11].

In a prior study by Puspita [12], the three approaches of KNN, Decision Tree, and Nave Bayes were examined to see the difference in each degree of accuracy. The KNN technique has a class precision for pred. negative of 52.17%, pred. positive of 0.00%, and pred. neutral of 97.27%. The accuracy percentage then rises to 96.13%. In the Decision Tree approach, the class precision for pred. negative is 55.00%, pred. positive is 0.00%, and pred. neutral is 97.28%. Sentiment analysis using Twitter data for BPJS services achieved an accuracy level of 96.01. And the latter achieves

an accuracy of 89.14%. Where is the class precision for pred. negative is 16.67%, pred. positive is 1.64%, and pred. neutral is 98.40% with the Naïve Bayes method. In this study, it can be seen that the Decision Tree method has a higher accuracy level than the other two methods, with an accuracy rate of 96.13%.

Research works by Wang [13] aim to analyze scholarships to ensure that they are effective and equitable for students. Fuzzy, Set Pair Analysis (SPA), and Self-organizing Maps (SOM) were among the retrieval techniques previously tried, but the outcomes were unsatisfactory. The decision tree was ultimately selected because it is significantly more efficient than the prior approach (C4.5 algorithm). In this investigation, the accuracy differences between the methods C4.5, ID3, Fuzzy, and Set Pair analysis were 91.59%, 87.32%, 90.45%, and 83.68%, respectively. The speed levels between the four ways are 1.7, 1.9, 2.5, and 2.2 seconds, with a noticeable variation between them.

The goal of subsequent research by Luigi, et al. [14] is to categorize the several situations that can impair voltage stability, possibly due to increased usage or interference. The decision tree has a black box character and can be used for unknown or complex data, with much data accessible. Other machine learning techniques include the clustering algorithm, neural network, and static method. From the explanation regarding the implementation of Machine Learning to assist decision-making that has been described, the Decision Tree has the best decision-making accuracy. So the author will select home internet bandwidth with the Decision Tree decision-making method as shown in Figure 1.

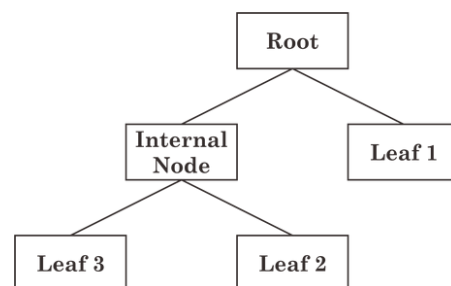


Figure 1. Decision tree concept [15]

However, the Decision Tree has a significant challenge, especially if it is to be realized in implementation, namely how to find which attributes to choose at each level; this is also known as Attribute Selection [16]. Various

measures of attribute retrieval to select attributes can perform well. There are various algorithms in this Decision Tree method, including ID3 and C4.5.

Among the studies about the C4.5[17] and ID3 Algorithms[18] is a decision tree algorithm used by Yudhi Pratama [19]. The research was carried out by creating a website for using this method to make decisions. The pattern or decision tree model produced by the two algorithms is the study's outcome. The accuracy of ID3 is 62%, according to the self-testing findings, whereas the accuracy of the C4.5 algorithm is 88%.

In this research, we chose the decision tree approach for research on selecting home internet bandwidth with the C4.5 algorithm so that individuals may obtain the necessary WiFi bandwidth. After comparing the two algorithms, it was discovered that C4.5 has a greater level of accuracy.

## II. METHODOLOGY

This study is classified as experimental design because one or more independent variables are manipulated and applied to one or more dependent variables to measure their effect on the latter [20]. To help researchers come to a logical conclusion about the link between these two variable types, the effect of the independent variables on the dependent variables is typically monitored and recorded over some time.

At first, we perform some data collection, in looking for material for comparison with the created web application, the authors look for similar applications or in line with methods to obtain information. Some literature gather to be analyzed and reviewed its impact on the study [13], [14], [19], [21], [22]. The researcher then limited the use of the Decision Tree approach for the purposes of this research based on the determination of the machine learning mechanism.

The next phase is by distributing questionnaires on the WiFi internet user survey website and generating datasets based on the application, and data was collected in two different methods. Purposive sampling will be employed to gather data through surveys to make the information more thorough and targeted to the test cases, particularly Home WiFi users. A closed-open survey is utilized to collect data, and there is 50 samples total, in

contrast to Kerlinger and Lee's assertion that there should always be at least 30 samples used in quantitative research [23].

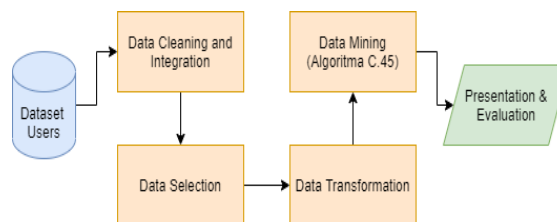


Figure 2. Research flow on decision tree method

The collected data, then analyzed, and the stages carried out in this C4.5 decision tree process research are shown in Figure 2, described as follows.

1. Data Integration and Cleaning: Sorting data to determine whether it is necessary or not, discarding duplicate or inconsistent data, for example. Then add more useful information or enhance the already existing data.
2. Data Selection: The training dataset and the testing dataset are the two components of the dataset. The mining process will use the training dataset to find patterns, and the testing dataset will be used to validate any patterns that are discovered.
3. Data Transformation: This process is the same as grouping data where data is bandwidth medium
4. Data Mining: The process of finding patterns from existing datasets, the research that the author did uses the Decision tree C4.5 algorithm.
5. Presentation & Evaluation: All forms of results are explained in full, clearly and in detail; this stage is also a correction for whether there are errors in the process.

The next step is to implement the model into development, the construction flow describes in Figure 3.

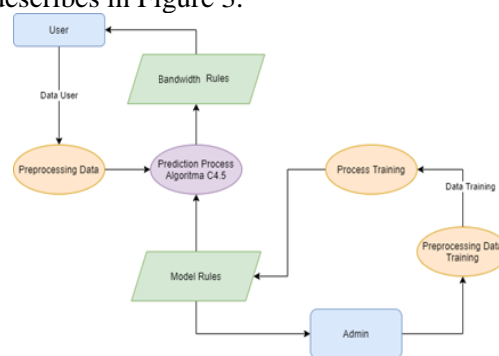


Figure 2. Construction method

The steps taken in this implementation are as follows:

- A. Create Prediction Pattern
  1. Input training data.
  2. The data preprocessing process is carried out so that the data can be managed during the training process
  3. Conducting the Data Training Process with the Decision Tree Algorithm, the process of looking for data patterns.
  4. Get the Model in the form of Rules.
- B. Making Predictions based on Prediction Patterns
  1. User Input Data.
  2. They preprocess data so that the prediction process can be managed.
  3. Taking the rules that the training process results have formed by the admin.
  4. Prediction process with patterns generated by the results of the training process by the admin.
  5. Bandwidth Decision Results are presented.

Accuracy testing with the following accuracy formula:

$$Accuracy = \frac{T}{T+F} \quad (1)$$

With T is the total conclusions that are under the prediction results (The data states Enough and the prediction results state Enough (All True)); while F is the Total conclusions that do not match the prediction results (Data states Enough, but the prediction results state Less (All False)). We also carry out testing based on black box testing[24], to ensure all application functions can run properly.

### III. RESULTS AND DISCUSSION

#### 3.1 Data Collection and Analysis

By distributing questionnaires about internet usage in the home, information was gathered. The data can be gathered in bulk or by personally completing the online questionnaire on the website for determining internet bandwidth. Purposive sampling will be used to choose the data to be collected to make the data more specific and focused on the case being evaluated; however, due to the enormous volume of data needed, the procedure is based

on a dataset generator that uses a random method.

The data needed to make prediction patterns include the Number of Occupants in the house, Number of Devices, Device average usage range, Large bandwidth currently used, Inferior internet Less / Enough / More, The data required for the Best Price Provider; for Recommendation Process it includes: Provider Name, Bandwidth, and Monthly Fee. The following are the Tables consisting of the classifications.

*Table 1. Bandwidth classifications*

Bandwidth Classifications	Not enough	Enough	More
Low (<= 20)	16	5	3
Middle (21 - 40)	2	6	5
High (> 40)	4	3	6

*Table 2. Number of occupants classifications*

Number of Occupants Classifications	Not enough	Enough	More
Low (<= 3)	6	3	9
Middle (4 - 5)	6	8	5
High (> 5)	10	3	0

*Table 3. Bandwidth classifications*

Num of Devices Classifications	Not enough	Enough	More
Low (<= 5)	4	3	6
Middle (6 - 7)	2	2	4
High (> 7)	16	9	4

*Table 4. Bandwidth classifications*

Usage range Classifications	Not enough	Enough	More
Low (< 10)	3	3	5
Middle (10 - 15)	3	2	5
High (> 15)	16	9	4

The issue that needs to be resolved is how to make it simpler for people to select internet bandwidth that meets their needs. This is because many people select service providers simply because they are affordable, without considering the bandwidth offered, even though this is one of the most crucial factors to consider. From this point on, the neighborhood will likely experience many complaints, particularly about slow internet due to inappropriate bandwidth selection.

In this application, a prediction tree based on various historical data on the community's

internet usage experience will be generated from the observations, and a decision tree will be made from the predictions. The created Decision Tree will be used as a predictive pattern for those who want to know whether or not the bandwidth used is appropriate. In addition, it can be used to determine how much bandwidth is appropriate for it, along with recommendations of the provider who has the highest efficiency compared to the highest price based on bandwidth requirement. Next is creating a pattern shelter for the predicted C4.5 decision tree patterns, which can later be selected for use. This application, will provide a concise and easy-to-use interface so that both novice and familiar users can use it immediately.

### 3.2 Process Analysis

In the process, the author will process historical data on home WiFi internet usage, and then the data is processed by a mining process, namely Decision Tree C4.5. In a Decision Tree, there are attributes in it; these attributes are divided into two parts, namely the first input attribute, which then produces the target attribute as the second part; these attributes function as determinants of the calculation of gain and ratio which will later be compared.

In the process, two types of data are needed, namely the training dataset and the testing data set, the training dataset is needed to carry out the decision tree process, which then results from the process to produce a rule, then the rule is checked using dataset testing. Each of the two datasets must have a target attribute; in this case, the target attribute is less, enough, and more. The input attribute that has the largest gain ratio is the attribute that will be the root.

### 3.3 Measuring the Model

In the design for this bandwidth selection process, the author uses the decision tree algorithm C4.5. in the process of changing the table into a tree structure, then from that structure, we get a rule and then conclude it. First, the selection of the attribute as the root is based on the highest gain value for each attribute. In determining the gain, we must first determine the entropy value; here is the equation for the entropy:

$$Entropy(S) = -\sum_{i=1}^n p_i \log_2 p_i \quad (2)$$

$S$ : case set  
 $n$ : number of partitions  $S$   
 $p_i$ : the proportion of  $S_i$  to  $S$

Furthermore, after getting the Entropy, we continue to calculate the gain using the Formula (3)

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (3)$$

$S$ : case set  
 $A$ : attribute  
 $n$ : number of attribute partition  $A$   
 $|S_i|$ : number of cases on partition  $i$   
 $|S|$ : number of cases in  $S$

After Gain ( $S, A$ ) is calculated, create a branch for each value (Attribute of root), divide cases into branches, and repeat the process for each branch until all cases on the branch have the same class. The calculation is as follows for each Gain ( $S, A$ ):

$$\begin{aligned} &Gain(\text{Total}, \text{Bandwidth}) \\ &= 1.549587 - \left( \left( \frac{24}{50} * 1,236440 \right) + \left( \frac{13}{50} * 1,460484 \right) + \left( \frac{13}{50} * 1,526234 \right) \right) = 0.179549 \end{aligned}$$

$$\begin{aligned} &Gain(\text{Total}, \text{Num of Occupants}) \\ &= 1.549587 - \left( \left( \frac{18}{50} * 1.459147 \right) + \left( \frac{19}{50} * 1.557431 \right) + \left( \frac{13}{50} * 0 \right) \right) = 0.432470 \end{aligned}$$

$$\begin{aligned} &Gain(\text{Total}, \text{Num of Devices}) \\ &= 1.549587 - \left( \left( \frac{13}{50} * 1.526234 \right) + \left( \frac{8}{50} * 1.5 \right) + \left( \frac{29}{50} * 1.391452 \right) \right) = 0.105724 \end{aligned}$$

$$\begin{aligned} &Gain(\text{Total}, \text{Usage Range}) \\ &= 1.549587 - \left( \left( \frac{11}{50} * 1.539484 \right) + \left( \frac{10}{50} * 1.485475 \right) + \left( \frac{29}{50} * 1.391452 \right) \right) = 0.106763 \end{aligned}$$

Based on the calculation table above on node 1, it is known that the attribute with the highest gain is the number of occupants, which is 0.432470. It can be seen from these attributes that we have three branches, Low, Middle, and High. The Low and Middle branches must be calculated again, while the High branches can be classified because the entropy is 0 (zero). High branches have a ratio of Less and Middle respectively 10 versus 3, therefore, Low is higher, and then it is concluded to be "Low" with an accuracy percentage of 76.92% based on the comparison of 10 and 3.

The results of our calculations also demonstrate that Bandwidth, which has a size of 0.581976, is the property with the highest gain. This attribute has three branches: Low, Medium, and High. The Low branch requires recalculation, however, the Medium and High branches can be identified because both have entropy 0 (zero). The ratio in the Medium branch is Enough and More, each 1 vs 2, so the highest result is "More" with a percentage of accuracy of 66.67%, and in the High branch, it can be determined right away to have a "More" result with a percentage of 100% because all of the final data concluded More.

The Usage range property has three branches for calculating purposes: Light, Medium, and Heavy. Recalculation is required for the Light branch, but the Medium and Heavy branches can be completed because each has an entropy value of 0 (zero). Given that both have their final results in the Less area, where accuracy is 100%, it can be said that the medium and heavy branches are less accurate.

As for the calculation of the Number of Devices, because many gadgets only have one attribute content, "Small Number of Devices" while the others (Normal and Many Branches), are empty or zero, more branches cannot be made. A comparison of Less, Enough, and More, respectively is 3, 2, and 3, from here, the highest and worst values are taken, namely three and Less.

### 3.4 Comparing the Manual Calculation and Predictions using the Model

The results obtained by the system and detailed calculations in the process design can be correct because they have the same tree results. We can see with the same data set the tree results are obtained as below:

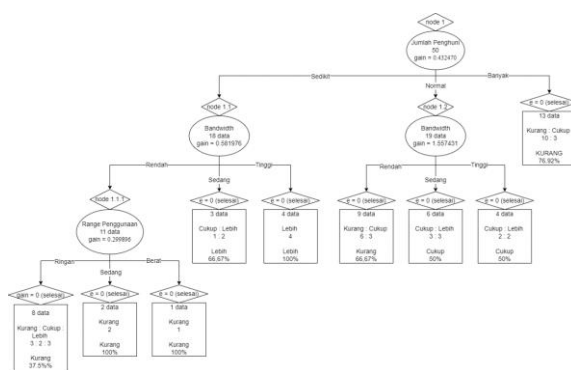


Figure 3. Manual calculation

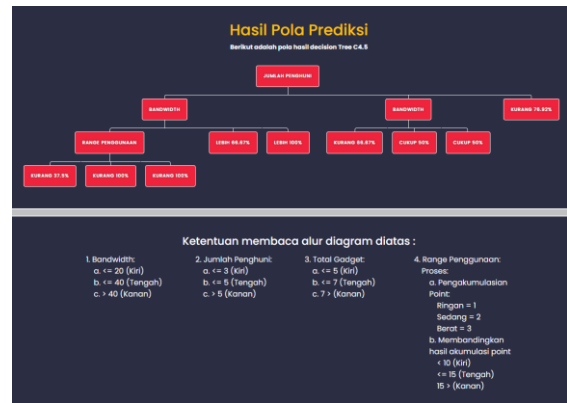


Figure 4. Calculation using predictions

From the results of the resulting tree, it can be ascertained that the C4.5 decision tree calculation system has run as expected because it produces a tree that is precisely the same.

The best top provider in each distribution of the bandwidth range using manual calculations and forecasts have similarities based on the tests done with the C4.5 decision tree analysis data set; therefore, it can be concluded that the output of the best website provider is correct as expected.

By comparing the usage range with manual calculations, we can see the test results based on the usage range. We can also test the system using the same dataset, namely C4.5 decision tree analysis, whereas the output results based on manual calculations are bandwidth with a medium or low range. in the 20 to 40 Mbps range. It can be said that the system's output, which ranges from 20 to 40 Mbps and matches both the results of the manual calculation process and the provider's recommendations based on manual calculations with a moderate range, is consistent with the output obtained through manual calculation.

### 3.5 Accuracy Measurement

Testing accuracy with the following accuracy formula (1), using the following data:

Table 5. True false conclusion

Conclusion	Predictions	Total	Value	Total
More	More	9	True	27 (True)
Enough	Enough	8	True	
Less	Less	10	True	
More	Enough	5	False	23 (False)
Enough	More	3	False	
Less	More	6	False	
Enough	Less	3	False	6
Less	Enough	6	False	

Then it can be calculated as follows:

$$Accuracy = \frac{27}{27+23} = 0.54$$

Based on the calculation, the accuracy obtained is 0.54, or a percentage of 54%.

### 3.6 Model Implementation on the Web

In order to use the model as the finished product, the research's next step is to build an application interface, which is an implemented product. At this step, interface design describes how each module's website interface works.

Users can interact with the system more easily thanks to the interface's design. This design includes the initial appearance of the system, entering login information, entering predicting data, and adding a list of prediction results to the main display that explains the prediction results. The interface is made so that it can make it easier for users in general. The display in Figure 6 is the initial view when you first open the internet bandwidth selection website. This page has two kinds of buttons, each of which functions to view the website and admin page. To make predictions the user does not have to log in and can make predictions directly based on the pattern that the super admin has determined.

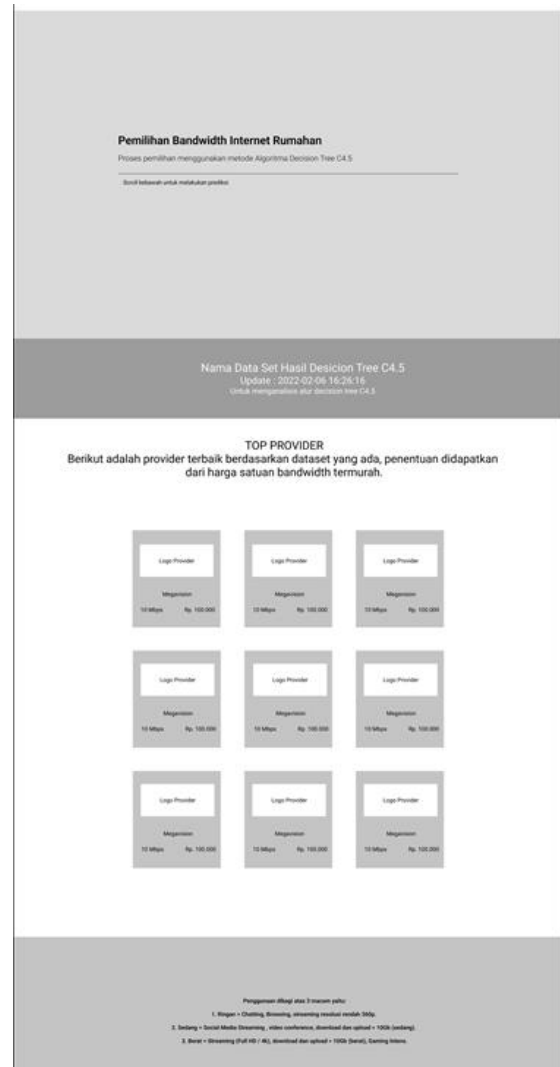


Figure 5. Interface Design of the Application

While the display in Figure 7 is the top main view after selecting the predicted pattern, the title is presented along with the detailed data and the resulting tree pattern.



Figure 6. Predictions Result.

After the implementation process is complete, the following process is testing the system. Previously the construction or coding phase has been completed, this phase is the next to check whether the application that has been made can work properly as planned or not. This stage also prevents problems when the user uses the application. The testing using the black box method.

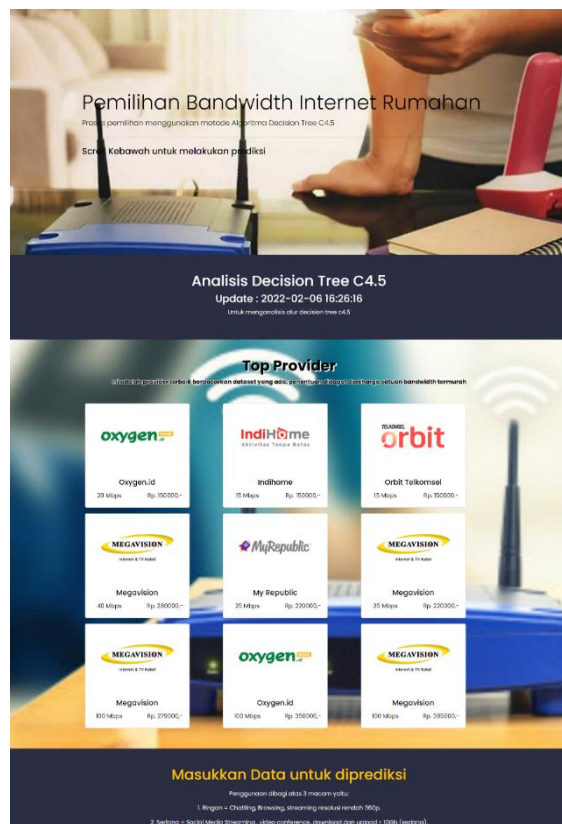


Figure 7. Working Application for Predictions

Figure 8 is the initial view of the website using the internet network. This view includes the name of the data used, the list of the best providers based on the dataset used, and the prediction form. General users can make predictions without having to log in first.

#### IV. CONCLUSION

The bandwidth selection application has been built based on the theoretical basis, framework, and model that has been described in the previous chapter, this website has successfully run according to its design by implementing the Decision Tree C4.5 algorithm. This system can determine how much bandwidth is needed by projecting the various uses of people who have used wifi at

home, from the various kinds of data, then processed using the C4.5 algorithm and a decision-making pattern is obtained to make the prediction process for determining the right bandwidth. Here, the system also provides provider recommendations based on the efficiency of provider prices from various community usage data.

Based on the application that has been made, the user can determine the amount of bandwidth that suits the needs of a particular home. The application that has been made can provide recommendations for the best provider based on needs, sourced from imported datasets. The decision tree C4.5 algorithm can help solve the problem of choosing bandwidth recommendations well.

By using the system development recommendations, this application can be improved. You can automate data development with system integration when retrieving datasets from a sub-district, for instance, as this dataset will grow in tandem with the sub-internet district's usage. It can be improved in terms of determining variables so that they will determine the decision tree process in forming patterns even better, leading to a higher level of accuracy.

In order to prevent further confusion about slow internet, which mistakes in bandwidth selection may actually cause, we hope that the community will make use of this bandwidth selection website application and use the available bandwidth following their own needs. Hopefully, this application can be improved upon in the future and made more complex to benefit the community more.

#### BIBLIOGRAPHY

- [1] "Asosiasi Penyelenggara Jasa Internet Indonesia." <https://apjii.or.id/content/read/39/559/Laporan-Survei-Profil-Internet-Indonesia-2022> (accessed Aug. 19, 2022).
- [2] S. R. Pokhrel and M. Mandjes, "Improving Multipath TCP Performance over WiFi and Cellular Networks: An Analytical Approach," *IEEE Trans. Mob. Comput.*, 2019, doi: 10.1109/TMC.2018.2876366.
- [3] A. J. A. M. van Deursen, "Digital inequality during a pandemic: Quantitative study of differences in



- COVID-19-related internet uses and outcomes among the general population,” *J. Med. Internet Res.*, 2020, doi: 10.2196/20073.
- [4] S. H. Chan, Q. Song, S. Sarker, and R. D. Plumlee, “Decision support system (DSS) use and decision performance: DSS motivation and its antecedents,” *Inf. Manag.*, 2017, doi: 10.1016/j.im.2017.01.006.
- [5] J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, and C. Carlsson, “Past, present, and future of decision support technology,” *Decis. Support Syst.*, 2002, doi: 10.1016/S0167-9236(01)00139-7.
- [6] M. K. Anam, B. N. Pikir, and M. B. Firdaus, “Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, 2021, doi: 10.30812/matrik.v21i1.1092.
- [7] Y. Y. Song and Y. Lu, “Decision tree methods: applications for classification and prediction,” *Shanghai Arch. Psychiatry*, 2015, doi: 10.11919/j.issn.1002-0829.215044.
- [8] Z. F. Firmansyah and N. F. Puspitasari, “ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 BERDASARKAN OPINI PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES,” *J. Tek. Inform.*, vol. 14, no. 2, pp. 171–178, Oct. 2021, doi: 10.15408/JTI.V14I2.24024.
- [9] L. Sun, S. Fu, and F. Wang, “Decision tree SVM model with Fisher feature selection for speech emotion recognition,” *Eurasip J. Audio, Speech, Music Process.*, 2019, doi: 10.1186/s13636-018-0145-5.
- [10] P. J. Lisboa and A. F. G. Taktak, “The use of artificial neural networks in decision support in cancer: A systematic review,” *Neural Networks*, 2006, doi: 10.1016/j.neunet.2005.10.007.
- [11] F. Blanco-Mesa, J. M. Merigó, and A. M. Gil-Lafuente, “Fuzzy decision making: A bibliometric-based review,” *Journal of Intelligent and Fuzzy Systems*. 2017, doi: 10.3233/JIFS-161640.
- [12] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, 2021, doi: 10.32493/informatika.v5i4.7622.
- [13] X. Wang, C. Zhou, and X. Xu, “Application of C4.5 decision tree for scholarship evaluations,” 2019, doi: 10.1016/j.procs.2019.04.027.
- [14] L. Vanfretti and V. S. N. Arava, “Decision tree-based classification of multiple operating conditions for power system voltage stability assessment,” *Int. J. Electr. Power Energy Syst.*, 2020, doi: 10.1016/j.ijepes.2020.106251.
- [15] A. R. Panhalkar and D. D. Doye, “Optimization of decision trees using modified African buffalo algorithm,” *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2021, doi: 10.1016/j.jksuci.2021.01.011.
- [16] B. Joa, A. Paulus, R. Mikoleit, and G. Winkel, “Decision making in tree selection – contemplating conflicting goals via marteloscope exercises,” *Rural Landscapes*, 2020, doi: 10.16993/rl.60.
- [17] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Edik Inform.*, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [18] L. R. Haidar, E. Sedyono, and A. Iriani, “ANALISA PREDIKSI MAHASISWA DROP OUT MENGGUNAKAN METODE DECISION TREE DENGAN ALGORITMA ID3 dan C4.5,” *J. Transform.*, 2020, doi: 10.26623/transformatika.v17i2.1609.
- [19] Y. P. Tanjung, S. R. Sentinuwo, and A. Jacobus, “Penentuan Daya Listrik Rumah Tangga Menggunakan Metode Decision Tree,” *J. Tek. Inform.*, 2016, doi: 10.35793/jti.9.1.2016.14141.
- [20] O. Mitchell, “Experimental Research Design,” in *The Encyclopedia of Crime and Punishment*, 2015.
- [21] A. Nugroho *et al.*, “Implementasi metode decision tree c4.5 dalam pemberian subsidi listrik kepada masyarakat,” *Proceeding SENDIU 2020*, pp. 978–979, 2020.
- [22] A.-A. B. Bugaje, J. L. Cremer, M. Sun,

- and G. Strbac, "Selecting decision trees for power system security assessment," 2021, doi: 10.1016/j.egyai.2021.100110.
- [23] P. W. Wardhani, "Hubungan antara nilai," *Hub. antara nilai*, pp. 42–52, 2009.
- [24] M. Kumar, A. Professor, S. Kumar Singh, R. K. Dwivedi, and A. Professor, "A Comparative Study of Black Box Testing and White Box Testing Techniques," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, 2015.