

PERBANDINGAN METODE EXTREME LEARNING MACHINE DAN ALGORITMA C4.5 UNTUK MENGLASIFIKASI KUALITAS RED WINE

Okky Putra Barus

Program Studi Teknik Informatika

Fakultas Ilmu Komputer, Universitas Pelita Harapan

E-mail: okky.barus@uph.edu

ABSTRACT

Wine is an alcoholic beverage which is a result of fermentation from grapes, where in the process of making wine is to convert sugar into alcohol. Each wine has a different qualities, the higher the quality of the wine, it will taste better. Therefore, a study is needed to determine in assessing the level of quality of the wine. There are various methods that can be used to determine the level of wine quality, This study will apply artificial neural network method, known as Extreme Learning Machine (ELM) and C4.5 Algorithm. Extreme Learning Machine is a feed-forward neural network with one or more hidden layers or single-layer feed-forward neural, C4.5 Algorithm is an algorithm used to create a decision tree (data tree) from the data. Based on the results of the research that has been done, it can be seen that the C4.5 Algorithm method is able to provide good results compared to extreme learning machine methods. By using a confusion matrix, the results of the study also show the accuracy of the C4.5 algorithm method is 1.5 times better than the extreme machine learning method.

Keywords: *Wine, Red Wine, Artificial Neural Network, Extreme Learning Machine, Algorithm C4.5*

ABSTRAK

Wine merupakan minuman beralkohol yang merupakan hasil fermentasi dari buah khususnya buah anggur. Setiap wine memiliki tingkat kualitas yang berbeda-beda yang mana semakin tinggi tingkat kualitas wine maka semakin baik pula rasa yang dihasilkan oleh wine tersebut. Oleh karena itu, diperlukan suatu penelitian yang dapat menjadi pertimbangan dalam penentuan tingkat kualitas dari wine tersebut. Penelitian ini akan mengimplementasikan metode jaringan saraf tiruan tersebut yaitu Extreme Learning Machine (ELM) dan Algoritma C4.5. ELM merupakan jaringan saraf tiruan feed-forward dengan satu atau lebih hidden layer atau single layer feed-forward neural, Algoritma C4.5 merupakan algoritma yang digunakan untuk membangun sebuah decision tree dari data. Berdasarkan hasil penelitian yang telah dilakukan, terlihat bahwa metode Algoritma C4.5 mampu memberikan hasil akurasi yang baik dibandingkan metode ELM. Dengan menggunakan confusion matrix, hasil penelitian juga menunjukkan bahwa tingkat keakuratan metode algoritma C4.5 sebesar 1.5 kali lebih baik dibandingkan metode ELM.

Kata Kunci: *Wine, Red Wine, Jaringan Saraf Tiruan, Extreme Learning Machine, AlgoritmaC4.5*

PENDAHULUAN

Buah anggur merupakan buah yang cukup populer dan banyak dikonsumsi oleh banyak orang. Buah anggur biasanya dapat dikonsumsi secara langsung atau dapat juga dibuat suatu produk minuman hasil fermentasi buah anggur. Wine adalah minuman beralkohol yang dibuat dari sari buah anggur. Wine dibuat melalui fermentasi gula yang ada di dalam buah anggur. Ada beberapa jenis minuman anggur yaitu, Red Wine, White Wine, Rose Wine, Sparkling Wine, Sweet Wine dan Fortified Wine. Red Wine atau anggur merah adalah anggur yang berasal dari buah anggur yang berwarna merah atau hitam. Warna merah diperoleh dari pencelupan kulit dan biji ke dalam sari buah yang telah diperas untuk difermentasi.

Di zaman teknologi masa kini, kita dapat mengukur tingkat kualitas wine yang ditawarkan dengan menggunakan berbagai macam tools perhitungan data mining seperti Matlab, Rapid Miner, Weka untuk menghitung tingkat kualitas wine.

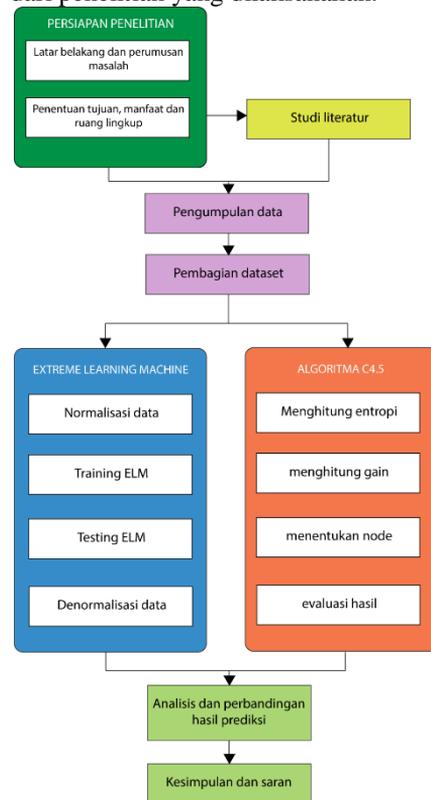
Extreme Learning Machine (ELM) merupakan metode pembelajaran baru dari jaringan syaraf tiruan. ELM merupakan jaringan syaraf tiruan feedforward dengan single hidden layer atau biasa disebut single hidden layer feedforward neural network (SLFN). Metode pembelajaran ELM dibuat untuk mengatasi kelemahan-kelemahan dari jaringan syaraf tiruan feedforward terutama dalam hal learning speed.

Decision Tree adalah metode pembelajaran yang paling banyak digunakan untuk masalah pengklasifikasian. Sebuah decision tree terdiri dari beberapa simpul yaitu tree's root, internal nod dan leafs. Konsep entropy digunakan untuk penentuan pada

atribut mana sebuah pohon akan terbagi (split). Semakin tinggi entropy sebuah sampel, semakin tidak murni sampel tersebut.

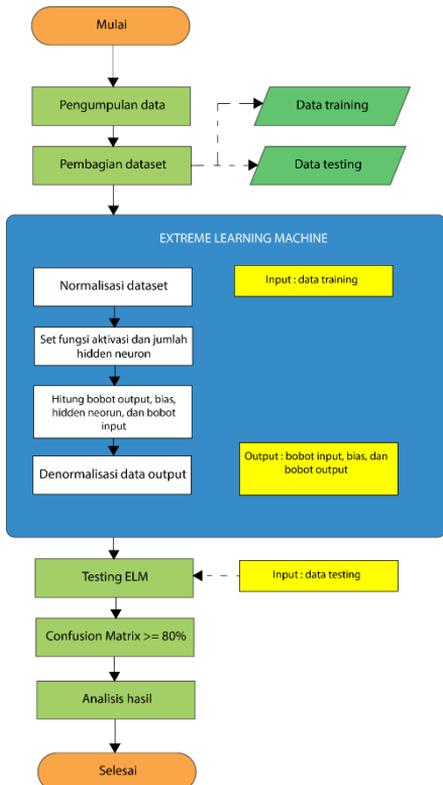
METODE PENELITIAN

Untuk menyelesaikan penelitian ini, peneliti akan menyusun tahapan-tahapan penelitian, serta rencana dan rancangan kegiatan hingga hasil akhir diperoleh dari penelitian yang dilaksanakan.



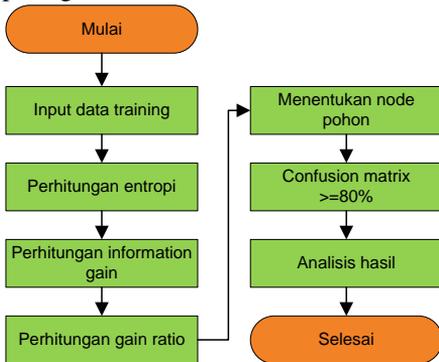
Gambar 1. Tahapan penelitian

Sedangkan proses diagram *Extreme Learning Machine(ELM)* tampak pada gambar dibawah ini :



Gambar 2. Diagram Proses ELM

Sedangkan diagram proses C4.5 tampak pada gambar dibawah ini.



Gambar 3. Diagram Proses C4.5

Metode prediksi kualitas wine yang akan diteliti ini menggunakan tool MATLAB untuk implementasinya dan terbagi menjadi 3 bagian :

1. Proses pembagian Data Training dan Data Testing

Proses training ini akan digunakan untuk mengembangkan dan meningkatkan efektivitas dari penggunaan ELM. Proses testing sendiri akan digunakan untuk mengukur dan mengevaluasi berulang kali kemampuan ELM yang digunakan.

Pembagian Data training dan Data Testing dibagi dengan proporsi sebagai berikut:

- Data Training sebanyak 75% dari total data yang dimiliki.
- Data Testing sebanyak 25% dari total data yang dimiliki.

2. Proses Training Extreme Learning Machine

Proses Training ELM ini dilakukan dengan menggunakan tool MATLAB serta proses training yang terbagi menjadi empat bagian, yaitu:

- Normalisasi Data Training
Rumus yang digunakan dalam normalisasi adalah sebagai berikut:

$$x = 2 \times \frac{(x_p - \min x_p)}{(\max x_p - \min x_p)} - 1$$

Keterangan:

x : nilai hasil normalisasi dengan rentang antara -1 sampai 1

x_p : nilai data asli yang belum dinormalisasi

$\min x_p$: nilai minimum pada data set

$\max x_p$: nilai maksimum pada data set

- Penentuan fungsi aktivasi dan jumlah hidden neuron

Pada tahap proses training, fungsi aktivasi dan jumlah dari hidden neuron harus ditentukan agar dapat menghasilkan fungsi aktivasi sigmoid. Menurut (Sun, Choi, Au, & Yu, 2008), ELM menghasilkan output peramalan yang stabil dengan jumlah hidden neuron 0-30. Tetapi jika output yang dihasilkan ELM

kurang optimal, maka jumlah hidden neuron-nya akan diubah.

- Perhitungan bobot input, bias dari hidden neuron, dan bobot output.

Proses training ELM menghasilkan output sebagai berikut: bobot input, bobot output, dan bias dari hidden neuron dengan tingkat error rendah dan diukur menggunakan Mean Square Error (MSE). Bobot input akan ditentukan secara acak, serta bobot output yang menjadi invers daripada matriks hidden layer dan output dan dapat ditulis sebagai berikut:

$$\beta = H^T T \quad (1)$$

$$H = (W_{i, \dots, W_N}, b_{i, \dots, b_N}, x_{i, \dots, x_N}) \quad (2)$$

$$= \begin{matrix} g(w_i \cdot x_i + b_i) & \cdots & g(w_N \cdot x_i + b_N) \\ \vdots & & \vdots \\ g(w_i \cdot x_N + b_i) & \cdots & g(w_N \cdot x_N + b_N) \end{matrix}$$

$$\beta = \begin{matrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{matrix} \quad T = \begin{matrix} t_1^T \\ \vdots \\ t_N^T \end{matrix} \quad (3)$$

- Denormalisasi hasil output
Proses training yang menghasilkan output harus didenormalisasi kembali dengan rumus denormalisasi sebagai berikut:

$$x = 0.5 \times (x_p + 1) \times (\max x_p - \min x_p) + \min x_p$$

Keterangan:

- x : nilai data setelah denormalisasi
- x_p : nilai data asli yang belum denormalisasi
- $\min x_p$: nilai minimum pada data set sebelum denormalisasi

$\max x_p$: nilai maksimum pada data set sebelum denormalisasi

3. Testing ELM

Pada tahap testing ELM ini, data input akan dinormalisasi dan didenormalisasi berdasarkan bobot input dan output yang didapat dari proses training melalui rumus yang sama dengan rumus data training.

Metode Diagnosis Algoritma C4.5

Pohon (tree) adalah sebuah struktur data yang terdiri dari simpul (node) dan rusuk (edge). Simpul pada sebuah pohon keputusan dibedakan menjadi tiga yaitu akar simpul, simpul percabangan, dan simpul akhir (Kundari, 2015).

Algoritma C4.5 ini merupakan algoritma yang sering digunakan untuk melakukan klasifikasi berdasarkan sifat prediktifnya. Algoritma ini berbentuk seperti pohon atau dapat disebut juga pohon keputusan (Decision Tree). Rumus untuk menghitung entropi pada algoritma C4.5 adalah sebagai berikut:

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

Keterangan:

- S adalah himpunan (dataset) kasus.
- k adalah banyaknya partisi S.
- P_i adalah probabilitas yang didapat dari Sum (Ya) atau Sum (Tidak) dibagi total kasus.

Analisis setiap atribut dan nilainya akan dilakukan serta perhitungan entropinya setelah entropi tersebut telah didapat menggunakan perhitungan sebelumnya. Perhitungan gain dilakukan dengan menggunakan rumus sebagai berikut:

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i) \quad (2)$$

Keterangan:

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke-i
 |S| : Jumlah kasus dalam S

Proses evaluasi untuk memperkuat keakuratan prediksi ELM dapat ditingkatkan dengan classification accuracy, sensitivity, dan specificity. Classification accuracy sendiri merupakan sebuah klasifikasi ketepatan yang diperoleh. Sensitivity merupakan sebuah ukuran parameter ketepatan dari suatu peristiwa yang diharapkan, dan specificity merupakan suatu ukuran parameter untuk menyatakan persentase kejadian yang tidak diinginkan. Classification accuracy, sensitivity, dan specificity dapat ditentukan dari menggunakan nilai yang terdapat dalam confusion matrix. Confusion matrix hanya menggunakan tabel matriks seperti tabel 1 di bawah ini, jika data set terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007).

Tabel 1. Tabel *Confusion Matrix Model*

Klasifikasi benar	Diklasifikasi sebagai	
	+	-
+	True positive	False negative
-	False positive	True negative

Rumus untuk menghitung classification accuracy, sensitivity, dan specificity adalah sebagai berikut:

Classification Accuracy

$$= \frac{TP + TN}{TP + FP + TN + FN}$$

$$= \frac{a + d}{a + b + c + d}$$

$$Specificity (\%) = \frac{TN}{TP + FN}$$

$$Sensitivity (\%) = \frac{TP}{TP + FN}$$

Keterangan:

- TP = jumlah true positives
- TN = jumlah true negatives
- FP = jumlah false positives
- FN = jumlah false negatives

HASIL DAN PEMBAHASAN

1. Akurasi Elm Terhadap Jumlah Hidden Neuron

Seperti yang telah dijelaskan pada sub bab 3.5 bahwa penentuan jumlah hidden neuron akan berpengaruh pada prediksi ELM yang optimal. Prediksi ELM sendiri akan menghasilkan output yang stabil dengan rentang 0 sampai 30 (Sun, Choi, Au, & Yu, 2008). Data tingkat akurasi yang dihasilkan prediksi menggunakan metode ELM berdasarkan nilai jumlah hidden neuron yang dapat dilihat pada Gambar di bawah ini.



Gambar 4. Akurasi prediksi ELM Terhadap Jumlah Hidden neuron

Terlihat pada grafik di atas bahwa jumlah hidden neuron akan berpengaruh pada hasil akurasi prediksi pada data training maupun pada data testing. Dengan rentang 0 sampai 30, didapatkan nilai akurasi terbaik untuk data training adalah 0.6225 dengan nilai jumlah hidden neuron adalah 30. Nilai akurasi

didapatkan dengan proses confusion matrix. Nilai yang mendekati 1 adalah nilai akurasi yang terbaik, yang berarti 100% akurat.

2. Keakuratan Hasil Presiksi Elm Dengan Confusion Matrix

Hasil perhitungan dengan menggunakan confusion matrix dilakukan terhadap proses training dan proses testing data. Tingkat keakuratan ELM dalam memprediksi kualitas red wine pada penelitian ini dapat dilihat pada kategori kualitas red wine pada seluruh data pada proses training dan testing data. Dari hasil proses ELM inilah, didapatkan tabel confusion matrix pada setiap kondisi masing-masing proses training dan testing data yang terlihat pada tabel di bawah ini.

Tabel 2. Confusion Matrix Hasil Proses ELM

Proses	TP	Total
<i>Training</i> ELM	741	1200
<i>Testing</i> ELM	241	399

Berdasarkan hasil confusion matrix di atas, tahap selanjutnya yaitu menghitung classification accuracy untuk proses training dan testing ELM dengan persamaan yang telah dijelaskan pada sub bab 3.7.1

Tabel 3. Kinerja dari Setiap Proses

Proses	<i>Classification Accuracy</i>
<i>Training</i> ELM	0.61750
<i>Testing</i> ELM	0.60401

Pada tabel di atas terlihat hasil dari proses classification accuracy. Dari tabel di atas terlihat nilai akurasi pada proses training ELM adalah 61.75% dan nilai akurasi untuk proses testing ELM adalah 60.40%.

3. Keakuratan Hasil Prediksi Algoritma C4.5 Dengan Confusion Matrix

Hasil perhitungan dengan menggunakan confusion matrix dilakukan terhadap proses training dan proses testing data. Tingkat keakuratan algoritma C4.5 dalam memprediksi kualitas red wine pada penelitian ini dapat dilihat pada kategori kualitas red wine pada seluruh data pada proses training dan testing data. Dari hasil proses algoritma C4.5 inilah, didapatkan tabel confusion matrix pada setiap kondisi masing-masing proses training dan testing data yang terlihat pada tabel di bawah ini.

Tabel 4. Confusion Matrix Hasil Proses ELM

Proses	TP	Total
<i>Training</i> Algoritma C4.5	1105	1200
<i>Testing</i> Algoritma C4.5	363	399

Berdasarkan hasil confusion matrix di atas, tahap selanjutnya yaitu menghitung classification accuracy untuk proses training dan testing ELM dengan persamaan yang telah dijelaskan pada sub bab 3.7.1

Tabel 5. Kinerja dari Setiap Proses

Proses	<i>Classification Accuracy</i>
<i>Training</i> Algoritma C4.5	0.9208
<i>Testing</i> Algoritma C4.5	0.9097

Pada tabel di atas terlihat hasil dari proses classification accuracy. Dari tabel di atas terlihat nilai akurasi pada proses training Algoritma C4.5 adalah 92.08% dan nilai akurasi

untuk proses testing Algoritma C4.5 adalah 90.97%.

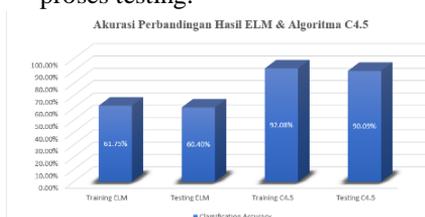
4. Perbandingan Tingkat Keakuratan Elm Dan Algoritma C4.5

Berikut merupakan perbandingan tingkat keakuratan secara keseluruhan baik proses training data dan proses testing data. Tabel berikut ini adalah perbandingan tingkat keakuratan kinerja ELM dan Algoritma C4.5 untuk masing-masing proses perhitungan data training dan data testing.

Tabel 6. Perbandingan training dan testing data ELM dan Algoritma C4.5

Proses	Classification Accuracy
Training ELM	61.75%
Testing ELM	60.40%
Training Algoritma C4.5	92.08%
Testing Algoritma C4.5	90.09%

Tabel di atas menunjukkan perbandingan hasil keakuratan dari metode ELM dan algoritma C4.5 untuk seluruh data baik data training dan data testing. Dari data tersebut diketahui bahwa tingkat keakuratan hasil yang lebih baik terjadi pada penggunaan metode algoritma C4.5 baik dari proses training ataupun proses testing.



Gambar 5. Grafik Perbandingan Hasil Akurasi ELM & Algoritma C4.5

Berdasarkan grafik di atas dapat disimpulkan bahwa hasil yang dihasilkan proses Algoritma C4.5 menunjukkan hasil yang jauh lebih akurat dibandingkan proses ELM.

KESIMPULAN

Dari hasil percobaan dan memprediksi kualitas red wine dengan menggunakan metode extreme learning machine (ELM) dan algoritma C4.5, maka dapat diambil kesimpulan sebagai berikut:

1. Jumlah hidden neuron yang dipakai dalam penggunaan metode ELM akan menentukan hasil akurasi pada penelitian yang akan dilakukan, dimana rentang nilai 0-30 pada penelitian ini jumlah hidden neuron yang dipakai untuk mendapatkan hasil yang optimal adalah 30.
2. Perbandingan tingkat akurasi dalam mengklasifikasikan kualitas red wine menggunakan algoritma C4.5 memiliki tingkat akurasi 1,5 kali lebih baik daripada metode ELM, dimana tingkat akurasi algoritma C4.5 sebanyak 92,08% (data training) dan 90,09% (data testing) dibanding ELM yang akurasinya sebanyak 61,75% (data training) dan 60,40% (data testing).

DAFTAR PUSTAKA

- [1] D. Davis and R. M. Cosenza, Business Research for Decision Making (3rd Ed.), Bellmonte, Calis: Wadsworth, 1993.
- [2] J. Sarwono, Metode Penelitian Kuantitatif dan Kualitatif, Yogyakarta: Graha Ilmu, 2006.
- [3] J. R. Raco, Metode Penelitian Kualitatif: Jenis, Karakteristik dan Keunggulannya, Jakarta: Grasindo, 2010.

- [4] Purwanto, Metodologi Penelitian Kuantitatif untuk Psikologi dan Pendidikan, Yogyakarta: Pustaka Belajar, 2010.
- [5] G. Keraf, Komposisi: Sebuah Pengantar Kemahiran Bahasa, Ende-Flores: Penerbit Nusa Indah, 1997.
- [6] Z. L. Sun, T. M. Choi, K. F. Au and Y. Yu, "Sales Forecasting Using Extreme Learning Machine With Applications in fashion Retailing," pp. 411-419, 2008.
- [7] E. S. Kundari, "Perbandingan Kinerja Metode Naive Bayes dan C4.5 dalam Pengklasifikasikan Penyakit Diabetes Melitus Di Rumah Sakit Kumala Siswi Kudus," 2015.
- [8] M. Ridwan, H. Suyono and M. Sarosa, *Penerapan Data mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayer Classifier*, pp. 59 - 64, 2013.
- [9] C. Vercellis, *Business Intelligence : Data Mining and Optimization for Decision Making*, John Wiley & Sons, Ltd., 2009.
- [10] G.-. B. Huang, Q.-. Y. Zhu and C.-. K. Siew, "Extreme Learning Machine : Theory and Applications," *Extreme Learning Machine : Theory and Applications*, pp. 490-501, 2006.
- [11] I. Rahmayuni, *Perbandingan Performansi Algoritma C4.5 dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang*, pp. 40-46, 2014.
- [12] D. H. Kamagi and S. Hasun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *Ultimatics*, Vols. Vol. VI, No. 1, pp. 15 - 20, 2014.
- [13] O. Kristanto, *Penerapan Algoritma Klasifikasi Data Mining untuk menentukan penjurusan siswa SMAN 6 Semarang*, 2014.
- [14] A. E. Tyasti, D. Ispriyanti and A. Hoyyi, *Algoritma Iterative Dichotomiser 3 (ID3) untuk mengidentifikasi data rekam medis*, pp. 238-246, 2015.
- [15] J. J. Pangaribuan, "Mendiagnosis Penyakit Diabetes Melitus Dengan Menggunakan Metode Extreme Learning Machine," *Journal Information System Development (ISD)*, vol. II, pp. 32-40, 2016.
- [16] F. Pakaja, A. Naba and P. Purwanto, "Peramalan Penjualan Mobil Menggunakan Jaringan Syaraf Tiruan dan Certainty Factor," 2012.
- [17] D. Widiastusti, *ANALISA PERBANDINGAN ALGORITMA SVM, NAIVE BAYES, DAN DECISION TREE DALAM MENGLASIFIKASIKAN SERANGAN (ATTACKS) PADA SISTEM PENDETEKSI INTRUSI*, 2012.
- [18] H. Pawingya, T. W. Widayati, D. Putra and P. Akbar, *Tinjauan Kinetika Pembuatan Rose Wine*,

2010.

- [19] A. D. Saputra, "Perbandingan Kadar Alkohol Dan Cita Rasa Wine Salak Pondoh Dengan Metode Red Wine, White Wine, dan Sake Menggunakan Strain Yeast Polandia," 2018.
- [20] R. A. Ovihapsany, A. Mustofa and N. Suhartatik, *Karakteristik Minuman Beralkohol Dengan variasi Kadar Ekstrak Buah Bit Dan Lama Fermentasi*, vol. V, pp. 55-63.