



## Pendeteksian Aroma Ganja Kering Menggunakan Algoritma Random Forest

Ulul Azmi<sup>#</sup>, Hendrick<sup>#</sup>, Humaira<sup>#</sup>

<sup>#</sup>Jurusan Teknologi Informasi, Politeknik Negeri Padang, Limau Manis, Padang, 25164, Indonesia  
E-mail: [azmipoenya8@gmail.com](mailto:azmipoenya8@gmail.com)

### ABSTRACTS

Cannabis (*Cannabis Sativa*) is the most consumed type of drug with a percentage of 56.7% where cannabis abuse has many negative effects, one of which is a decrease in the work of the brain so that it will experience hallucinations, illusions, thought disturbances, sudden changes in feelings and cause addictive effects. . The high number of cases raises several obstacles such as procuring expensive tools in the search process. So an E-Nose tool was designed to detect marijuana in which this system will mimic the function of the sense of smell (nose) in humans. In this research, we will process cannabis data that has been collected to determine a cannabis detection model by comparing two classification methods, namely the Random Forest and Decision Tree algorithms to determine the best classification in producing an accurate classification model for detecting cannabis. The accuracy of the model obtained from the Random Forest algorithm is 100% where the resulting model can correctly classify dry and non-cannabis samples. The Decision Tree algorithm produces an accuracy of 70% correctly on dry cannabis

Manuscript received 13 Dec. 2022; revised 16 Feb. 2023; accepted 22 Mar. 2023 Date of publication 31 Mar. 2023.  
International Journal, JITSI : Jurnal Ilmiah Teknologi Sistem Informasi licensed under a Creative Commons Attribution-Share Alike 4.0 International License



### ABSTRAK

Ganja (*Cannabis Sativa*) sebagai jenis narkoba terbanyak dikonsumsi dengan persentase sebesar 56,7% dimana penyalahgunaan ganja menimbulkan banyak dampak negatif salah satunya penurunan kerja otak sehingga akan mengalami reaksi halusinasi, ilusi, gangguan berpikir, perubahan perasaan secara tiba – tiba dan menimbulkan efek kecanduan. Tingginya kasus menimbulkan beberapa kendala seperti pengadaan alat yang mahal dalam proses pencarian. Maka dirancnglah sebuah alat E-Nose untuk mendeteksi ganja yang mana sistem ini akan meniru fungsi dari indera penciuman (hidung) pada manusia. Pada penelitian ini akan melakukan pengolahan data ganja yang telah terkumpul untuk menentukan model pendeteksian ganja dengan membandingkan dua buah metode klasifikasi yaitu algoritma *Random Forest* dan *Decision Tree* untuk menentukan klasifikasi yang terbaik dalam menghasilkan akurasi pengklasifikasian model untuk mendeteksi ganja. Hasil akurasi model yang diperoleh dari algoritma Random Forest adalah sebesar 100% dimana model yang dihasilkan dapat mengklasifikasikan secara benar sampel ganja kering dan bukan ganja. Pada algoritma *Decision Tree* menghasilkan akurasi sebanyak 70% secara benar pada ganja kering

**Keywords / Kata Kunci** — *Data Mining; Decision Tree; Ganja; Klasifikasi; Random Forest*

### 1. PENDAHULUAN

Mayoritas jenis narkoba (narkotika dan obat – obatan) yang pertama kali dikonsumsi adalah ganja dengan persentase sebanyak 56,7%. Hal tersebut menjadikan ganja sebagai jenis narkoba yang paling banyak dikonsumsi dari lima jenis narkoba teratas[1].

Ganja atau yang memiliki nama ilmiah *Cannabis Sativa* merupakan tanaman berserat atau tanaman obat psikotropika, karena memiliki kandungan zat narkotika tetrahidrokanabinol (THC). Psikotropika sendiri merupakan zat kimia atau obat yang menimbulkan penurunan fungsi kerja otak sehingga menyebabkan pengguna mengalami reaksi berupa halusinasi, ilusi, gangguan cara berpikir, perubahan perasaan yang tiba-tiba dan menimbulkan efek kecanduan jika disalahgunakan[2].

Tingginya kasus dari penggunaan narkoba khususnya ganja juga menimbulkan beberapa kendala seperti proses pencarian peredaran narkoba yang masih terbatas seperti menggunakan anjing pelacak, serta penggunaan alat yang masih minim dikarenakan mahalnya pengadaan alat.

Di perkembangan teknologi 4.0 masa sekarang ini, sudah ditemukan sebuah sistem untuk pendeteksian yang memanfaatkan aroma gas sebagai subjeknya yaitu Electronic Nose (E-Nose). Sistem E-Nose sudah meliputi berbagai aspek kehidupan seperti sebuah paper yang berjudul *Non-Invasive Method for Tuberculosis Exhaled Breath Classification Using Electronic Nose*, pada tulisan tersebut berhasil menerapkan sistem E-Nose untuk mengklasifikasikan napas melalui pernapasan ekspirasi dalam pendeteksian penyakit Tuberculosis dengan tingkat keberhasilan sebesar 92% menggunakan metode Support Vector Machine (SVM), 94.87% menggunakan metode Artificial Neural Network (ANN), 88.24% menggunakan metode Random Forest dan 88.24% menggunakan XGBoost[3]. Pada paper lainnya yang berjudul *A portable electronic nose system for the identification of cannabis-based drugs*, sang penulis berhasil mendeteksi klasifikasi narkoba jenis cannabis dengan tingkat keberhasilan 91.96% menggunakan metode PCA, 98.5% menggunakan metode Support Vector Machine (SVM)[4].

Berdasarkan bahasan diatas, dirancanglah sebuah E-Nose untuk mendeteksi ganja kering yang mana sistem ini akan meniru fungsi dari bagian tubuh manusia, yaitu indera penciuman (hidung) guna membantu BNN dalam upaya pencegahan / peredaran narkoba khususnya pada ganja. Pada penelitian ini akan melakukan pengolahan data ganja yang telah terkumpul untuk menentukan model pendeteksian ganja kering dengan membandingkan dua buah metode klasifikasi yaitu algoritma Random Forest dan Decision Tree untuk menentukan klasifikasi yang terbaik dalam menghasilkan akurasi pengklasifikasian model untuk mendeteksi ganja.

## 2. METODOLOGI PENELITIAN

### 2.1 Ganja (*Cannabis Sativa*)

Ganja merupakan tanaman berserat dengan bentuk daun menjari. Ganja biasanya akan dikeringkan termasuk bagian tangkai, daun, biji serta bunganya saat diolah. Ganja mengandung lebih 60 senyawa kimia cannabinoid. Delta-9-tetra hydro cannabinol (THC) sebagai kandungan yang paling tinggi yang terdapat pada tanaman ganja yang bersifat psikoaktif yang akan mempengaruhi fungsi tubuh seseorang berdasarkan dosis yang dikonsumsi. Penyalahgunaan ganja tentu akan menimbulkan berbagai masalah kesehatan, mempengaruhi struktur dan fungsi otak, sistem kardiovaskular, sistem pernapasan dan serta sistem reproduksi [5].



GAMBAR 1. Database Ganja

### 2.2 Metodologi

#### 2.2.1 Random Forest

RF (Random Forest) merupakan salah satu jenis algoritma klasifikasi yang terdiri dari lebih satu pohon keputusan yang setiap pohon keputusan dibentuk bergantung pada nilai-nilai vector acak sampel secara independen dan identik didistribusikan yang sama untuk semua pohon. Metode ini merupakan salah satu metode klasifikasi yang sangat akurat digunakan dalam melakukan prediksi, bisa menangani inputan variabel yang sangat besar jumlahnya tanpa overfitting, dan membantu menghilangkan korelasi antara pohon keputusan seperti karakteristik ensemble methods[3], [6].

Pohon keputusan dimulai dengan cara menghitung nilai entropy sebagai penentu tingkat ketidakmurnian atribut dan nilai information gain. Untuk menghitung nilai entropy digunakan rumus seperti pada persamaan 1, sedangkan nilai information gain menggunakan persamaan 2[7].

$$Entropy(Y) = - \sum_{i=1}^n p_i \log_2 p_i \tag{1}$$

Dimana Y adalah himpunan kasus dan pi merupakan proporsi Yi terhadap Y serta n merupakan jumlah partisi.

$$Information\ Gain(Y, a) = Entropy(Y) - \sum_{i=1}^n \frac{|Y_i|}{|Y|} Entropy(Y_i) \quad (2)$$

Dimana a merupakan atribut dan n adalah jumlah partisi atribut a. |Y<sub>i</sub>| adalah jumlah kasus pada partisi ke-i dan |Y| jumlah kasus dalam Y.

2.2.2 Decision Tree

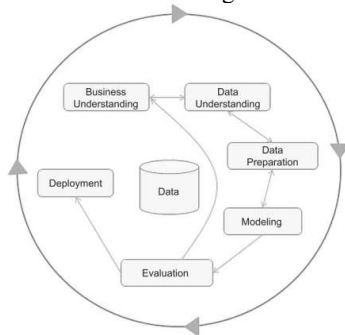
Decision Tree ini merupakan algoritma pengambilan keputusan yang melakukan partisi rekursif atas ruang instance, sebuah pohon keputusan tipikal terdiri dari simpul internal, tepi dan simpul daun. Setiap simpul internal disebut simpul keputusan yang mewakili tes pada atribut atau subset atribut, dan masing-masing edge diberi label dengan nilai spesifik atau rentang nilai atribut input. Pengklasifikasi Decision Tree memperoleh akurasi yang serupa dan terkadang lebih baik jika dibandingkan dengan metode klasifikasi lainnya[6].

Cara membangun sebuah Decision Tree dapat menggunakan perhitungan Gini Impurity untuk leaf dari Decision Tree dengan rumus sebagai berikut:

$$Gini\ Impurity = 1 - (Probabilitas(Ya))^2 - (Probabilitas(Tidak))^2 \quad (3)$$

2.2.3. CRISP – DM (Cross Industry Standard Process for Data Mining)

CRISP – DM merupakan salah satu metodologi dalam penyelesaian data mining berdasarkan permasalahan dalam penelitian ini. Metodologi ini terdiri atas 6 tahapan seperti berikut[8] :



**GAMBAR 2.** Diagram Metodologi CRISP – DM

- 1) *Business Understanding* : atau pemahaman terhadap masalah mengetahui permasalahan apa saja yang ada yang kemudian memahami permasalahan dengan penyelesaian data mining.
- 2) *Data Understanding* : pemahaman pada data dengan melakukan penentuan pada kebutuhan data, pengumpulan serta eksplorasi data. Dalam penelitian ini dibutuhkan data gas dari tanaman ganja kering yang dikumpulkan dari hasil sensor e-nose. Data yang terkumpul merupakan dataset berupa data numerik dari hasil sensor gas yang terpasang pada e-nose.

- 3) *Data preprocessing* : persiapan data adalah tahapan yang memiliki output untuk membuat dataset akhir dari data mentah. Berdasarkan kebutuhan model, pengolahan data dapat diulang hingga tercipta dataset terbaik. Tahap pengolahan data dipisahkan menjadi pemilihan tabel, perekaman data, pemilihan atribut, pembersihan data, dan transformasi data berdasarkan input.
- 4) *Modeling* : Tahap pemodelan diawali dengan pelatihan dataset. Pelatihan dilakukan dengan menerapkan metode machine learning. Selama proses pelatihan, juga memberikan informasi tentang kinerja model. Pelatihan dataset membandingkan dua metode klasifikasi yaitu Random Forest dan Decision Tree untuk menentukan klasifikasi yang terbaik dalam menghasilkan akurasi pengklasifikasian model untuk mendeteksi ganja kering.
- 5) *Evaluation* : Tahap ini dilakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh algoritma yang digunakan. Parameter yang digunakan untuk evaluasi kinerja algoritma adalah confusion matrix. Penghitungan confusion matrix bertujuan untuk memperoleh akurasi yang di dapat, seperti pada rumus berikut :

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

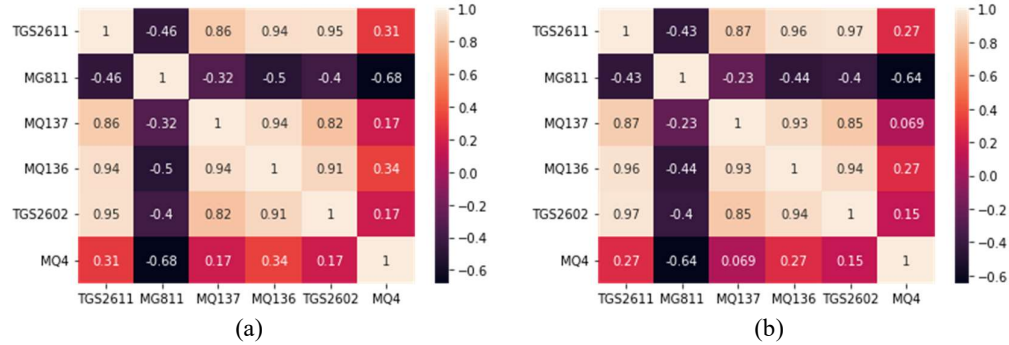
Ket :  
 TN = True Negative  
 TP = True Positive  
 FP = False Positive  
 FN = False Negative

- 6) *Deployment* : deployment merupakan tahapan akhir dari pembuatan model. Pada penelitian ini model di-deploy ke dalam website dengan menggunakan flask. Flask dipilih karena cocok untuk pemrograman python. Sehingga hasil prediksi akan ditampilkan dalam bentuk website dari.

### 3. HASIL DAN PEMBAHASAN

Dataset yang digunakan merupakan data yang telah direkam menggunakan alat E-Nose pendeteksi aroma ganja kering. Data tersebut langsung direkam di BNNP SUMBAR (Badan Narkotika Nasional Provinsi Sumatera Barat).

#### 3.1 Analisa Korelasi Sensor

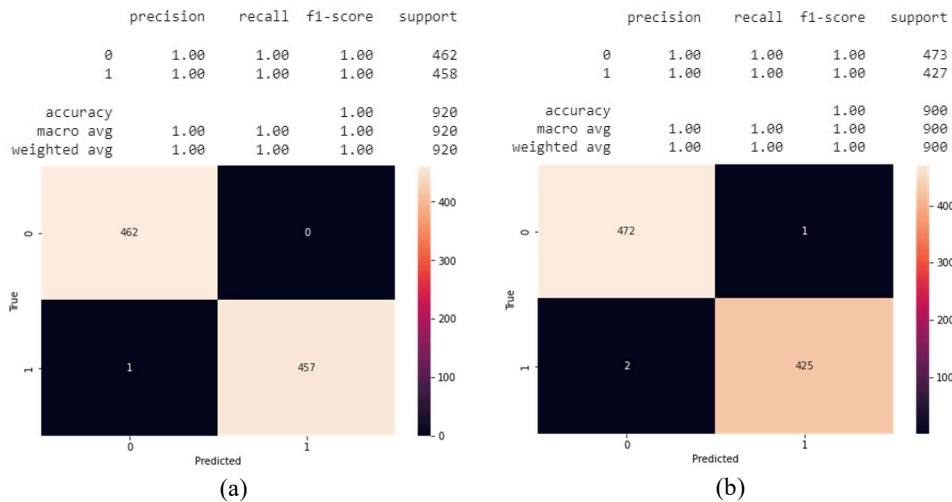


GAMBAR 3. Korelasi data (a)sebelum dibersihkan outlier dan (b)setelah dibersihkan outlier

Dapat dilihat pada gambar 3 dalam perbandingan korelasi data setiap sensor sebelum dibersihkan dan setelah dibersihkan outlier memiliki cukup perbedaan. Secara umum, terjadi kenaikan pada setiap korelasi saat outlier dibersihkan. Semakin mendekati nilai 1 maka korelasi akan semakin bagus atau disebut juga berkorelasi linear secara positif. Jika bernilai 0 berarti tidak berkorelasi sama sekali. Dan jika semakin -1 berarti berkorelasi secara terbalik atau berkorelasi linear secara negatif.

Pada sensor MG811 yang berkorelasi linear secara negatif atau berketerbalikan, meski begitu tingkatan korelasi rendah karena mendekati angka 0 sehingga ada kemungkinan adanya kerusakan pada alat. Pada sensor MQ4 yang berkorelasi linear secara positif. Meski begitu nilai korelasi rata – rata pada nilai 0,1 hingga 0,3 yang mengartikan rendahnya tingkat korelasi karena mendekati 0.

#### 3.2 Random Forest



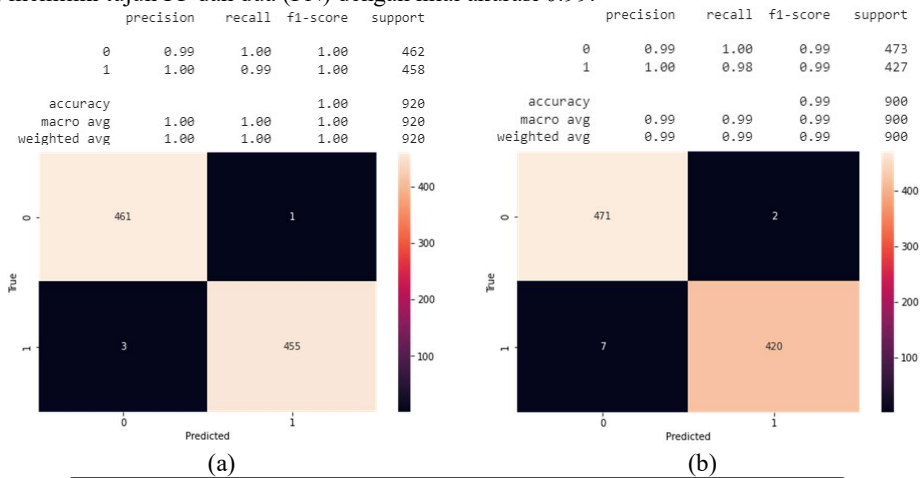
GAMBAR 4. Perbandingan Classification Report dan Confusion Matrix model Random Forest

Gambar 4 merupakan perbandingan antara classification report dan confusion matrix model random forest (a) sebelum outlier dibersihkan dan (b) setelah outlier dibersihkan. Terlihat gambar (a) dan (b) sama – sama menghasilkan akurasi 1 serta pada confusion matrix nya juga hanya memiliki satu False Positive (FP) pada gambar (a) dan confusion matrix pada gambar (b) memiliki 2 FP dan 1 False Negative (FN).

#### 3.3 Decision Tree

Gambar 5 merupakan perbandingan antara classification report dan confusion matrix model decision tree (a) sebelum outlier dibersihkan dan (b) setelah outlier dibersihkan. Terlihat gambar (a) menghasilkan nilai akurasi

satu (1) serta pada confusion matrix nya juga hanya memiliki tiga FP dan satu FN dan confusion matrix pada gambar (b) memiliki tujuh FP dan dua (FN) dengan nilai akurasi 0.99.



**GAMBAR 5.** Perbandingan Classification Report dan Confusion Matrix model Decision Tree

Gambar 5 merupakan perbandingan antara classification report dan confusion matrix model decision tree (a) sebelum outlier dibersihkan dan (b) setelah outlier dibersihkan. Terlihat gambar (a) menghasilkan nilai akurasi satu (1) serta pada confusion matrix nya juga hanya memiliki tiga FP dan satu FN dan confusion matrix pada gambar (b) memiliki tujuh FP dan dua (FN) dengan nilai akurasi 0.99.

Dari model yang telah ada, kemudian dilakukan pengujian prediksi pada setiap model sehingga memperoleh nilai akurasi 100% pada kedua sampel baik pada ganja kering dan bukan ganja untuk model random forest (tabel 1 dan tabel 2) dan 70% pada sampel ganja kering (tabel 3) serta 100% bukan ganja untuk model decision tree (tabel 4).

**TABLE 1.** HASIL Pengujian Sampel Bukan Ganja model Random Forest

| Sample      | Uji ke | Waktu Pengujian     | Hasil Deteksi |
|-------------|--------|---------------------|---------------|
| Bukan Ganja | 1      | 2022-11-21 09:10:15 | Bukan Ganja   |
|             | 2      | 2022-11-21 09:11:37 | Bukan Ganja   |
|             | 3      | 2022-11-21 09:12:11 | Bukan Ganja   |
|             | 4      | 2022-11-21 09:12:55 | Bukan Ganja   |
|             | 5      | 2022-11-21 09:13:29 | Bukan Ganja   |
|             | 6      | 2022-11-21 09:14:07 | Bukan Ganja   |
|             | 7      | 2022-11-21 09:14:55 | Bukan Ganja   |
|             | 8      | 2022-11-21 09:15:28 | Bukan Ganja   |
|             | 9      | 2022-11-21 09:15:54 | Bukan Ganja   |
|             | 10     | 2022-11-21 09:16:15 | Bukan Ganja   |

**TABLE 3.** Hasil Pengujian Sampel Ganja Kering model Decision Tree

| Sample       | Uji ke | Waktu Pengujian     | Hasil Deteksi |
|--------------|--------|---------------------|---------------|
| Ganja Kering | 1      | 2022-11-21 08:55:06 | Ganja Kering  |
|              | 2      | 2022-11-21 08:55:47 | Ganja Kering  |
|              | 3      | 2022-11-21 08:56:22 | Ganja Kering  |
|              | 4      | 2022-11-21 08:56:53 | Ganja Kering  |
|              | 5      | 2022-11-21 08:57:23 | Ganja Kering  |
|              | 6      | 2022-11-21 08:58:43 | Ganja Kering  |
|              | 7      | 2022-11-21 08:59:23 | Ganja Kering  |
|              | 8      | 2022-11-21 08:59:47 | Ganja Kering  |
|              | 9      | 2022-11-21 09:00:13 | Ganja Kering  |
|              | 10     | 2022-11-21 09:00:37 | Ganja Kering  |

**TABLE 2.** Hasil Pengujian Sampel Ganja Kering model Random Forest

| Sample       | Uji ke | Waktu Pengujian     | Hasil Deteksi |
|--------------|--------|---------------------|---------------|
| Ganja Kering | 1      | 2022-11-21 09:36:58 | Ganja Kering  |
|              | 2      | 2022-11-21 09:38:02 | Ganja Kering  |
|              | 3      | 2022-11-21 09:38:52 | Ganja Kering  |
|              | 4      | 2022-11-21 09:39:43 | Ganja Kering  |
|              | 5      | 2022-11-21 09:40:31 | Ganja Kering  |
|              | 6      | 2022-11-21 09:41:08 | Ganja Kering  |
|              | 7      | 2022-11-21 09:41:41 | Ganja Kering  |
|              | 8      | 2022-11-21 09:42:20 | Ganja Kering  |
|              | 9      | 2022-11-21 09:43:30 | Ganja Kering  |
|              | 10     | 2022-11-21 09:43:52 | Ganja Kering  |

**TABLE 4.** Hasil Pengujian Sampel Bukan Ganja model Decision Tree

| Sample      | Uji ke | Waktu Pengujian     | Hasil Deteksi |
|-------------|--------|---------------------|---------------|
| Bukan Ganja | 1      | 2022-11-21 08:42:33 | Bukan Ganja   |
|             | 2      | 2022-11-21 08:43:06 | Bukan Ganja   |
|             | 3      | 2022-11-21 08:43:43 | Bukan Ganja   |
|             | 4      | 2022-11-21 08:44:21 | Bukan Ganja   |
|             | 5      | 2022-11-21 08:44:46 | Bukan Ganja   |
|             | 6      | 2022-11-21 08:45:21 | Bukan Ganja   |
|             | 7      | 2022-11-21 08:45:52 | Bukan Ganja   |
|             | 8      | 2022-11-21 08:46:28 | Bukan Ganja   |
|             | 9      | 2022-11-21 08:46:58 | Bukan Ganja   |
|             | 10     | 2022-11-21 08:47:33 | Bukan Ganja   |

#### 4. KESIMPULAN

Korelasi yang dihasilkan data sebelum outlier dibersihkan dan setelah dibersihkan mengalami kenaikan nilai. Kenaikan nilai yang tidak begitu besar dikarenakan data outlier yang dibersihkan tidak begitu banyak, dari data awal sebanyak 3.065 data menjadi 2.999 data saja setelah dibersihkan. Akurasi yang dihasilkan model Random Forest menghasilkan nilai 1 baik sebelum outlier dibersihkan dan setelah dibersihkan. Pada model Decision Tree yang belum dibersihkan outlier menghasilkan nilai akurasi 1 dan setelah dibersihkan menghasilkan nilai akurasi 0.99. Akurasi prediksi model Random Forest yang dihasilkan pada pengujian sampel ganja kering didapatkan nilai 100% dari 10 kali percobaan begitu juga dengan sampel bukan ganja. Akurasi prediksi model Decision Tree yang dihasilkan pada pengujian sampel ganja kering didapatkan nilai 70% dari 10 kali percobaan. Sehingga dari pengujian terbukti bahwa pengklasifikasian menggunakan Algoritma Random Forest jauh lebih baik dibandingkan dengan menggunakan Decision Tree. Semakin banyak jumlah data training nilai akurasi yang dihasilkan pada model akan semakin bagus.

#### REFERENSI

- [1] Ppid.bnn.go.id, "Indonesia Drug Report," 2022. <https://ppid.bnn.go.id/informasipublik/informasi-berkala/>
- [2] E. Isnaini, "Penggunaan Ganja Dalam Ilmu Pengobatan Menurut Undang-Undang Nomor 35 Tahun 2009 Tentang Narkotika," *J. Indep.*, vol. 5, no. 2, p. 46, 2017, doi: 10.30736/ji.v5i2.71.
- [3] H. Hendrick, R. Hidayat, G. J. Horng, and Z. H. Wang, "Non-Invasive Method for Tuberculosis Exhaled Breath Classification Using Electronic Nose," *IEEE Sens. J.*, vol. 21, no. 9, pp. 11184–11191, 2021, doi: 10.1109/JSEN.2021.3061616.
- [4] Z. Haddi, A. Amari, H. Alami, N. El Bari, E. Llobet, and B. Bouchikhi, "A portable electronic nose system for the identification of cannabis-based drugs," *Sensors Actuators, B Chem.*, vol. 155, no. 2, pp. 456–463, 2011, doi: 10.1016/j.snb.2010.12.047.
- [5] F. Harun, Dasrul, Sugito, Zuhrawaty, Nazaruddin, and E. Rahmi, "Pengaruh Paparan Asap Ganja (*Cannabis sativa*) terhadap Patologi Anatomi Testis Tikus Putih (*Rattus novergicus*) Strain Wistar," *Jimvet*, vol. 01, no. 2, pp. 226–234, 2017.
- [6] S. Rahayu and J. J. Purnama, "Klasifikasi Konsumsi Energi Industri Baja Menggunakan Teknik Data Mining," *J. Teknoinfo*, vol. 16, no. 2, p. 395, 2022, doi: 10.33365/jti.v16i2.1984.
- [7] S. Saadah and H. Salsabila, "Prediksi Harga Ponsel Menggunakan Metode Random Forest," *J. Komput. Terap.*, vol. 7, no. 1, pp. 24–32, 2021.
- [8] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.