

## Identifikasi Pemalsuan Kumis Kucing dari Babadotan dan Tekelan menggunakan *Machine Learning*

### *Identification of Java Tea Adulteration by Babadotan and Tekelan using Machine Learning*

Penulis Ary Prabowo<sup>1\*</sup>, Wisnu Ananta Kusuma<sup>1,3</sup>, Annisa<sup>1</sup>, Mohamad Rafi<sup>2,3</sup>

Afiliasi <sup>1</sup>Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Kampus IPB Dramaga, Jl. Kamper, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680  
<sup>2</sup>Departemen Kimia, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Kampus IPB Dramaga, Jl. Kamper, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680  
<sup>3</sup>Pusat Studi Biofarmaka Tropika, Lembaga Penelitian dan Pengabdian kepada Masyarakat., Institut Pertanian Bogor, Kampus IPB Taman Kencana, Jl. Taman Kencana No.3, RT.03/RW.03, Babakan, Kecamatan Bogor Tengah, Kota Bogor, Jawa Barat 16128

#### Kata Kunci

- pemalsuan kumis kucing
- random forest
- Recursive Feature Elimination
- data berdimensi tinggi
- multi class

#### Keywords

- *Orthosiphon aristatus* adulteration
- Random Forest
- Recursive Feature Elimination
- High dimensional
- multi class

Diterima 2 Agustus 2022  
Direvisi 8 Agustus 2022  
Disetujui 29 Desember 2022

#### \*Penulis Koresponding

Ary Prabowo

email:

ary\_p2306@apps.ipb.ac.id,  
ary.prabowo@ui.ac.id

#### ABSTRAK

Kumis kucing yang lebih dikenal sebagai tanaman obat herbal berfungsi sebagai pengobatan kesehatan dan mengobati berbagai penyakit. Tingginya permintaan kumis kucing menyebabkan tingginya harga dan penurunan jumlah bahan baku tanaman obat, sehingga menimbulkan berbagai masalah pengendalian mutu seperti kandungan berbagai komponen bioaktif, serta pemalsuan dengan tanaman lain. Selama ini pendeteksian pemalsuan dilakukan dengan berbagai analisis termasuk analisis kimia dan metode statistik untuk mengolah data. Data yang digunakan berdimensi tinggi dengan tingkat kerapatan yang sangat tinggi, sehingga menyebabkan kesulitan dalam klasifikasi. Data campuran kumis kucing terdiri dari 1201 fitur dengan jumlah sampel 216. Penelitian ini menggunakan metode random forest dengan jumlah tree sebanyak 100 dan metode *Random Forest* (RF) digabungkan dengan metode *Recursive Feature Elimination* (RFE). RF dan RFE menghasilkan nilai optimum pada jumlah fitur 244. Evaluasi eksperimen mengungkapkan bahwa metode yang diusulkan dapat mencapai akurasi tinggi sebesar 81,82% untuk mengidentifikasi pemalsuan kumis kucing.

#### ABSTRACT

*Java Tea (Orthosiphon aristatus) is a common herbal medicinal plant that functions as a health treatment and treats various diseases. The high demand for Java Tea causes high prices and a decrease in the amount of medicinal plant raw materials, causing various quality control problems such as the content of various bioactive components and adulteration from babadotan and tekelan. So far, the detection of adulteration has been carried out by various analyzes, including chemical analysis and statistical methods to process data. The data used is of high dimension with a very high-density level, thus causing difficulties in classification. The mixed data of Orthosiphon aristatus consists of 1201 features with a total sample of 216. This study uses a Random Forest (RF) method with a total of 100 trees, and the RF method is combined with the Recursive Feature Elimination (RFE) method. In the RF and RFE that can be produced, the optimum value for the number of features is 244. The experimental evaluation results revealed that the proposed method could achieve a high accuracy of 81.82% in identifying Orthosiphon aristatus.*



## PENDAHULUAN

Tanaman herbal telah banyak digunakan sebagai obat sejak jaman dahulu. Berbagai jenis tanaman yang dapat dijadikan obat, seperti Jahe (*Zingiber officinale*) (Murugesu *et al.* 2020), Kunyit (*Curcuma domestica*) (Masuda *et al.* 1993), Kencur (*Kaempferia galanga*) (Ichwan *et al.* 2019), dan Kumis Kucing (*Orthosiphon aristatus*) juga digunakan sebagai bahan obat tradisional oleh masyarakat lokal. Menurut WHO 80% penduduk dunia menggunakan produk herbal (Srirama *et al.* 2017). WHO juga merekomendasikan penggunaan obat herbal dalam pemeliharaan kesehatan masyarakat, terutama untuk mencegah dan mengobati penyakit kronis, degeneratif, dan kanker.

Kumis Kucing (*Orthosiphon aristatus*) merupakan famili *Lamiaceae*. Umumnya dikenal sebagai kumis kucing dan obat herbal di Asia Tenggara. *Orthosiphon aristatus* adalah tanaman penting dalam pengobatan tradisional. Kumis kucing (*Orthosiphon aristatus*) secara tradisional digunakan untuk mengobati berbagai penyakit seperti Diabetes Mellitus, hipertensi, rematik, radang amandel, gangguan menstruasi, dan penyakit yang mempengaruhi saluran kemih, dan masalah terkait angiogenesis seperti kanker (Umar *et al.* 2014; Ucuzian *et al.* 2010; Geng *et al.* 2013, Vijayan *et al.* 2013).

Penelitian yang dilakukan oleh Hernadi *et al.* (2018) menyebutkan bahwa kumis kucing sebagai bahan baku obat herbal memiliki permasalahan dengan kendali mutu seperti kandungan komponen bioaktif yang beragam serta pemalsuan dengan tanaman lain. Pemalsuan juga sering terjadi terhadap suplemen/obat penurun berat badan (Dastjerdi *et al.* 2018). Pemalsuan ini terjadi karena keterbatasan ketersediaan kumis kucing terhadap permintaan pasar yang begitu tinggi. Berbagai metode telah dilakukan untuk menganalisis tingkat pencemaran (Hernadi *et al.* 2018) menggunakan kromatografi cair kinerja tinggi (KCKT) dan kemometrik pada *fingerprint* kumis kucing. Dastjerdi *et al.* (2018) telah menganalisis menggunakan *principal component analysis* (PCA). Penelitian pemalsuan terhadap suplemen diet menggunakan metode *multivariate partial least squares regression* dan dikombinasi dengan *discriminant analysis* (PLS-DA) (Walkowiak *et al.* 2018). Fitria *et al.* (2018) juga menerapkan kombinasi spektrum UV-Vis, PCA, dan *discriminant analysis* untuk mendeteksi pemalsuan kumis kucing.

Penelitian yang dilakukan Fitria *et al.* (2018) hanya mengidentifikasi pemalsuan kumis kucing dengan takelan dan babadotan saja. Selain itu, dataset kumis kucing yang dihasilkan memiliki jumlah fitur sebanyak 1201 fitur. Jumlah fitur sebanyak itu akan menimbulkan

data berimensi tinggi sehingga terjadi tumpang tindih. Metode yang digunakan oleh fitria (2018) dengan menggunakan PCA sebagai metode reduksi dimensi, namun metode ini hanya mendapatkan hasil akurasi persilangan sebesar 74%.

Berdasarkan permasalahan di atas, penelitian ini menggunakan pendekatan fitur seleksi *Recursive Feature Elimination* (RFE) dan metode klasifikasi yaitu *Random Forest* (RF). Chen *et al.* (2018), metode RFE sering digunakan dalam memilih fitur yang berguna, sehingga dapat mengatasi masalah *high* dimensi dan peningkatan akurasi. *Random Forest* (RF) merupakan salah satu metode *machine learning* berbasis pohon yang mutakhir (Khammas, 2020) dan banyak penerapannya di berbagai bidang seperti pertanian (Hitziger *et al.* 2014), transportasi (Zhang *et al.* 2015), energi (Papadopoulos *et al.* 2018), dan kesehatan (Ellahham *et al.* 2019). Berdasarkan penelitian sebelumnya dengan berbagai keunggulan tersebut, maka penelitian ini menggunakan metode *random forest* yang digunakan untuk mengidentifikasi pemalsuan kumis kucing. Selain itu, penelitian ini mengembangkan identifikasi pemalsuan berdasarkan komposisi atau persentase campuran tumbuhan babadotan dan tekelan menjadi 6 kelas dengan hasil persilangan kumis kucing x babadotan 5%, kumis kucing x babadotan 25%, kumis kucing x babadotan 50%, kumis kucing x takelan 5%, kumis kucing x takelan 25% dan kumis kucing x takelan 50%.

## METODE

### Data Penelitian

Penelitian ini menggunakan data absorbansi dari tanaman kumis kucing, takelan, dan babadotan yang dikumpulkan oleh Fitria (2018) sebanyak 29 sampel yang terdiri atas 10 sampel kumis kucing, 10 sampel babadotan, dan 9 sampel takelan yang berasal dari 3 provinsi antara lain, Banten, Jawa Barat, dan Jakarta, Indonesia.

### Alat dan Bahan

Penelitian ini dilakukan menggunakan bantuan perangkat keras (*lihat Tabel 1*) dan Bahasa pemrograman yang digunakan untuk memproses data yaitu python dengan IDE *Jupyter Notebook*.

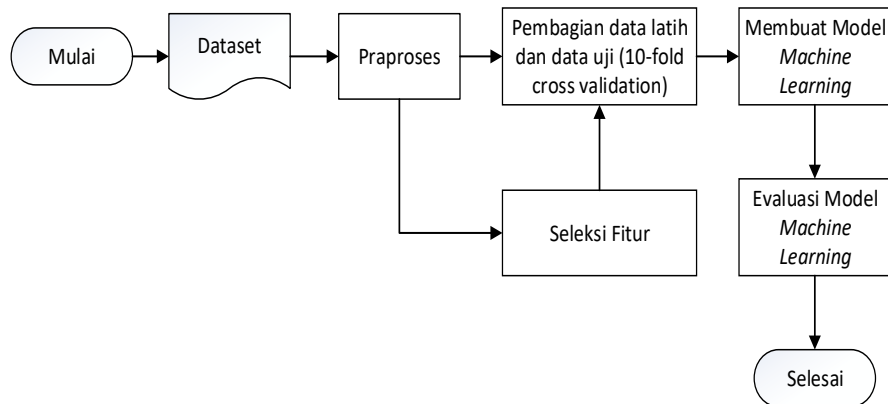
### Tahapan Penelitian

Penelitian ini dilakukan terdiri atas beberapa tahapan seperti pada **Gambar 1**.



**Tabel 1.** Perangkat Keras Penelitian

Spesifikasi	Client	Server
Processor	11th Gen Intel(R) Core(TM) i7-1195G7 @ .90GHz	Intel(R) Xeon(R) Silver 4110 CPU@ .10GHz
Memory	24GB (23700MB usable)	65580MB (46346MB Used)
Machine type	Laptop	Desktop
OS	Windows 11 Home 64-bit	Ubuntu 18.04.3 LTS

**Gambar 1.** Tahapan Penelitian**Dataset**

Data penelitian ini berupa panjang gelombang dan absorbansi pada tanaman kumis kucing, tekelan, dan babadotan. Rentang Panjang gelombang yang digunakan yaitu berkisar 200-800 nm. Ilustrasi tanaman kumis kucing, tekelan, dan babadotan dapat dilihat pada **Gambar 2**.

**Data Preprocessing**

Praproses data merupakan tahapan awal dari pengolahan data yang bertujuan mengoptimalkan data. Praproses yang dilakukan antara lain transpose yang bertujuan mengubah posisi target dan atribut data.

Setelah itu melakukan pengecekan data meliputi *redundancy*, dan *missing value*.

**Recursive Feature Elimination (RFE)**

*Recursive feature elimination* (RFE) merupakan bagian dari metode seleksi fitur seleksi bertujuan untuk mengatasi masalah dimensi dan peningkatan akurasi. Metode RFE bekerja dengan memilih seluruh fitur terlebih dahulu, kemudian menghilangkan fitur yang paling tidak penting, memeriksa hasilnya, dan mengulangi proses untuk membuat model baru hingga dapat meningkatkan akurasi.



Kumis kucing  
(*Orthosiphon Aristatus*)



Tekelan  
(*Eupatorium Riparium*)



Babadotan  
(*Acalypha Siamensis*)

**Gambar 2.** Kumis Kucing dan 2 Tanaman Pemalsu

### Model Klasifikasi dan Tuning Parameter

Setelah dilakukan praproses selanjutnya adalah pemodelan prediksi. Penelitian ini menggunakan lima metode klasifikasi machine learning yaitu *Random Forest (RF)*, *Support Vector Machine (SVM)*, *Decision Tree (DT)*, *Naïve Bayes*, *k-NN*. Kelima model tersebut dilakukan tuning parameter untuk mengoptimalkan kinerja dari metode tersebut. Penentuan parameter pada metode machine learning menggunakan metode *exhaustive grid search*. Metode ini adalah memungkinkan untuk mengkombinasi seluruh parameter yang digunakan.

Penelitian ini menggunakan beberapa parameter yang berpengaruh terhadap kinerja model RF, yaitu: *n\_estimator*, penetapan jumlah *n\_estimator* bergantung pada ketersediaan *hardware*, *max\_feature* = "auto", dan parameter lain yang digunakan yang terdapat pada *package* RF. Parameter yang berpengaruh terhadap kinerja SVM, yaitu: *gamma*, *kernel*, dan C. Metode DT juga memiliki parameter yang sama dengan RF, namun terdapat perbedaan pada parameter *criterion*. Selanjutnya metode k-NN dipengaruhi oleh parameter *n\_neighbors*. Sedangkan, metode *Gaussian Naïve Bayes* menggunakan parameter *var\_smoothing*.

### Evaluasi Model Klasifikasi

Pengukuran kemampuan suatu model klasifikasi dilakuka berdasarkan banyaknya data yang terklasifikasin dengan benar atau salah dan dapat disajikan secara lebih terperinci dengan menggunakan *confusion matrix*. Berikut adalah bentuk umum dari *confusion matrix* untuk klasifikasi (lihat **Tabel 2**).

**Tabel 2.** Bentuk Umum *Confusion Matrix* untuk Dua Kelas

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	True positive (TP)	False negative (FN)
<i>Actual Negative</i>	False positive (FP)	True negative (TN)

Berdasarkan tabel *confusion matrix* di atas maka *precision*, *recall*, *accuracy*, dan *F-measure* dapat dihitung dengan persamaan (1).

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad Accuracy = \frac{TP}{TP + TN + FN + FP}$$

$$F-measure = \frac{2 \times recall \times precision}{recall + precision} \quad (1)$$

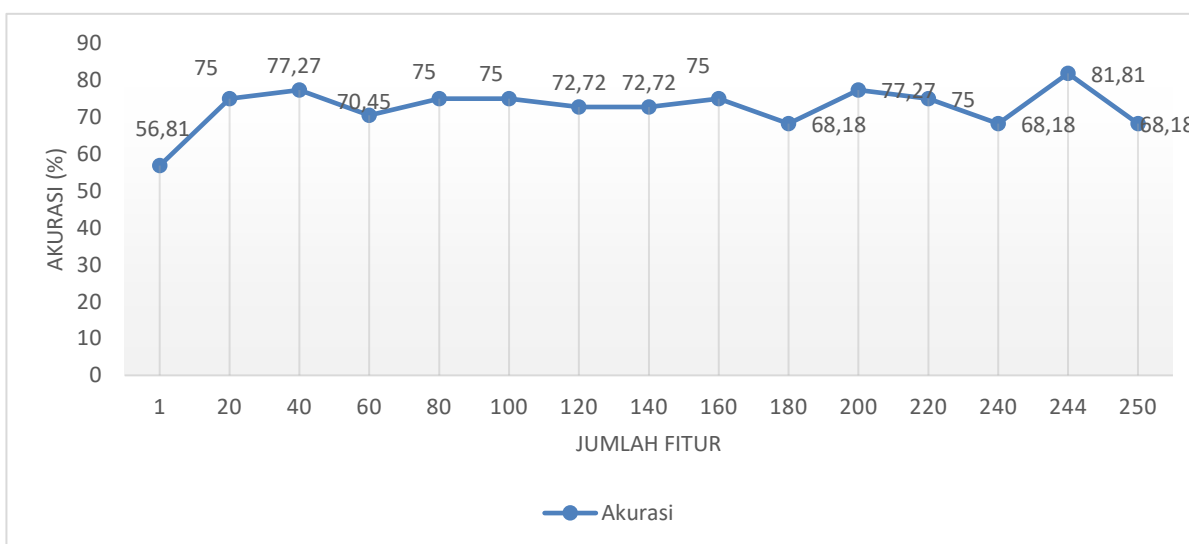
## HASIL DAN PEMBAHASAN

### Praproses Data

Tahapan praproses data pada dataset absorban campuran kumis kucing menghasilkan data dengan target kelas diubah menjadi 6 kelas yang terdiri atas (kumis kucing x babadotan 5%), (kumis kucing x babadotan 25%), (kumis kucing x babadotan 50%), (kumis kucing x takelan 5%), (kumis kucing x takelan 25%) dan (kumis kucing x takelan 50%).

### Seleksi Fitur

Setelah dilakukan praproses tahapan selanjutnya melakukan seleksi fitur. Metode seleksi fitur yang digunakan yaitu *Recursive Fitur Elimination (RFE)* dan PCA. Berdasarkan percobaan yang telah dilakukan



**Gambar 3.** Hasil Akurasi Menggunakan Metode RFE



seleksi fitur RFE dengan kombinasi metode Random Forest (RF) dapat dilihat pada **Gambar 3**.

**Gambar 3** menunjukkan bahwa dengan jumlah fitur sebanyak 1201 fitur dengan menerapkan metode RF dan RFE didapatkan jumlah fitur yang optimal yaitu 244 fitur, dengan nilai akurasi 81.82%. Selain metode RFE juga digunakan metode reduksi dimensi lainnya yaitu *Principal Component Analysis* (PCA). Penggunaan PCA dengan merepresentasikan 95 persen *variance* dari dataset sehingga diperoleh dimensi data PCA sebesar [126, 60].

### Perbandingan Hasil Evaluasi Model

Penelitian ini menggunakan beberapa metode classifier antara lain Random Forest, Support Vector Machine, Decision Tree, Naïve Bayes, dan K-NN. Jenis validasi yang digunakan adalah dengan membagi dataset menjadi 80% data latih dan 20% data uji. Perbandingan hasil evaluasi berupa akurasi, precision, recall, dan F-measure. **Gambar 3** menunjukkan hasil akurasi, precision, recall dan F-measure untuk setiap classifier tanpa PCA, dengan PCA, dan RFE.

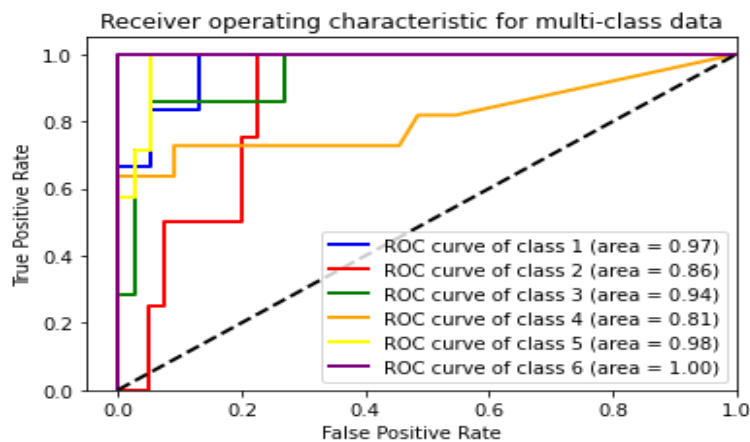
Hasil dari metode support vector machine memberikan hasil yang baik dengan rata-rata akurasi 79,55%, nilai tersebut naik daripada tanpa menggunakan PCA. Sehingga dapat diduga, metode SVM dengan baik dapat digunakan pada data berdimensi tinggi. Namun, metode *Random forest* (RF) dipadukan dengan RFE mendapatkan hasil akurasi tertinggi yaitu 81,82%. Kinerja lagoritma RF ditampilkan dalam bentuk grafik Receiver Operating Characteristic (ROC) dapat dilihat pada **Gambar 4**.

Berdasarkan grafik ROC pada Gambar 4 yang menunjukkan ROC pada masing-masing kelas. Nilai ROC terbaik terdapat pada kelas 6 yaitu kelas campuran kumis kucing dengan tekelan 50%.

### Perbandingan waktu eksekusi

Dalam melatih model pada masing-masing *classifier* maka dibutuhkan waktu eksekusi (*lihat Tabel 4*).

Berdasarkan **Tabel 4** menunjukkan bahwa metode klasifikasi dengan menggunakan fitur seleksi RFE menghasilkan waktu yang lebih lama dalam melatih *dataset* daripada metode yang lainnya.



**Gambar 4.** ROC pada Metode *Random Forest*

**Tabel 4.** Perbandingan Waktu *Training Model*

Classifier	Waktu (dalam second)	
	PCA	RFE
SVM	4300	6559
k-NN	9,53	4400
NB	5,88	1800
DT	4500	38619
RF	10800	259000



**Perbandingan seluruh akurasi *classifier***

Bagian ini menunjukkan perbandingan hasil akurasi pada masing-masing model *classifier* dapat dilihat pada **Gambar 5**. Metode *Random Forest* (RF) yang digabungkan dengan *Recursive Feature Elimination* (RFE). Metode ini juga menjadi yang paling baik jika dibandingkan dengan metode sebelumnya yang menggunakan *PCA* dan *discriminant analysis* (DA).

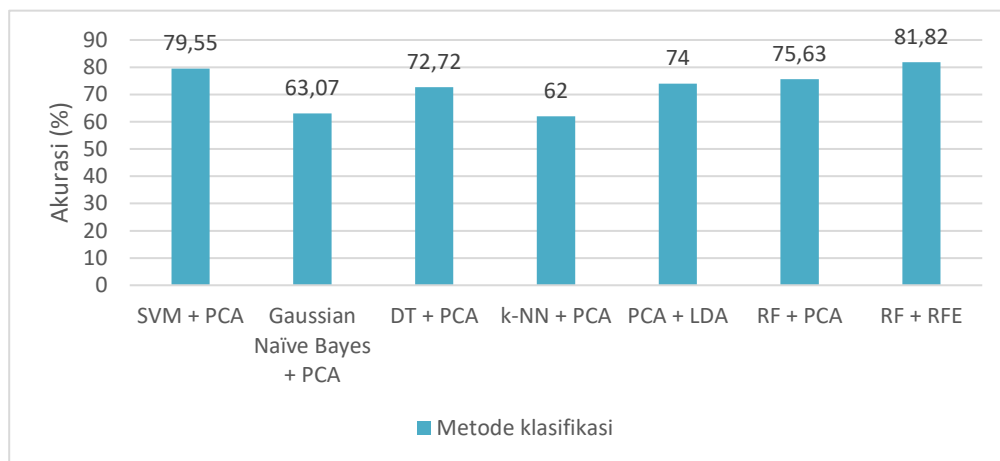
**Analisis Perbandingan Hasil**

Dari sisi hardware metode RFE membutuhkan hardware yang memadai daripada metode PCA. Metode RFE membutuhkan waktu yang lama (**Gambar 4**) dibandingkan dengan metode PCA. Meskipun begitu, RFE juga memiliki hasil evaluasi dengan kinerja terbaik terutama pada metode *Random Forest* (RF) dapat dilihat pada **Tabel 3**.

**SIMPULAN**

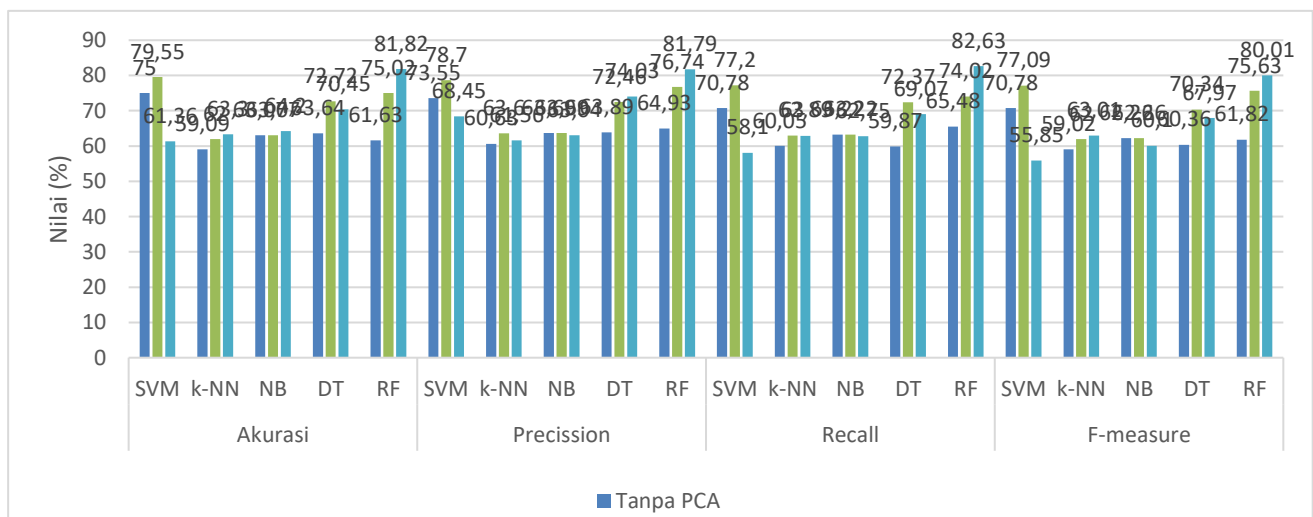
Penelitian ini telah berhasil mengidentifikasi pemalsuan kumis kucing dengan pencampuran 6 jenis komposisi dan jumlah fitur sebanyak 1201 fitur menggunakan metode *machine learning*. Hasil penelitian menunjukkan bahwa metode PCA dan RFE sangat berpengaruh terhadap akurasi model klasifikasi. Metode RF dengan RFE mendapatkan hasil terbaik yaitu 81,82%, walaupun SVM dengan PCA juga mendapatkan skor tinggi yaitu 79,55%.

Berdasarkan hasil perbandingan dapat disimpulkan bahwa penggunaan metode RFE membutuhkan sumber daya yang besar, baik itu waktu maupun hardware yang memadai. Tetapi akurasi yang didapat tidak signifikan dalam meningkatkan kinerja model klasifikasi.



**Gambar 5.** Perbandingan akurasi model *classifier*

**Tabel 3.** Perbandingan Hasil Evaluasi Model Klasifikasi



**DAFTAR PUSTAKA**

- Chen Q, Meng Z, Liu X, Jin Q, Su R. 2018. Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE. *Genes*, 9, 301: doi: 10.3390/genes9060301.
- Dastjerdi AG, Akhgari M, Kamali A, Mousavi Z. 2018. Principal component analysis of synthetic adulterants in herbal supplements advertised as weight loss drugs. 236-241: doi:10.1016/j.ctcp.2018.03.007.
- Ellahham S, Ellahham N, Simsekler MCE. Application of artificial intelligence in the health care safety context: opportunities and challenges. *Am J Med Qual* 2019. <https://doi.org/10.1177/1062860619878515>. 1062860619878515.
- Fitria AR. 2018. Deteksi pemalsuan kumis kucing dari babadotan Dan tekelan menggunakan kombinasi spektrum Uv-vis dan analisis diskriminan. Bogor. Indonesia. IPB University.
- Geng L, Chaudhuri A, Talmon G, Wisecarver JL, Wang J. 2013. TGF-Beta Suppresses VEGFA-Mediated Angiogenesis in Colon Cancer Metastasis. *PLoS ONE* 8(3): e59918. doi:10.1371/journal.pone.0059918.
- Hernadi E, Rohaeti E, Rafi M, Wahyuni WT, Putri, SP, Fukusaki E 2018. HPLC fingerprinting coupled with linear discriminant analysis for the detection of adulteration in *Orthosiphon aristatus*. *Journal of Liquid Chromatography & Related Technologies* 42 (15-16): 513-520. doi:10.1080/10826076.2019.1629956.
- Hitziger M, Ließ M. Comparison of three supervised learning methods for digital soil mapping: application to a complex terrain in the ecuadorian andes. *Soil Sci.* 2014. <https://doi.org/10.1155/2014/809495>.
- Ichwan SJA, Husin A, Suriyah WH, Lestari W, Omar MN, Kasmuri AR. 2019. Anti-neoplastic potential of ethyl-p-methoxycinnamate of *Kaempferia galanga* on oral cancer cell lines. *Material Today: Proceedings* 16(4): 2115-2121.
- Khammas BM. 2020. Ransomware Detection using Random Forest Technique. vol 6. pp.325-331. *ICT Express*. <https://doi.org/10.1016/j.ict.2020.11.001>.
- Masuda T, Jitoe A, Isobe J, Nakatani N, Yonemori S. 1992. Anti-oxidate and anti-inflammatory curcumin-related phenolics from rhizomes of *curcuma domestica*. *Phytochemistry*. 32(6): 1557-1560. [https://doi.org/10.1016/0031-9422\(93\)85179-U](https://doi.org/10.1016/0031-9422(93)85179-U).
- Papadopoulos S, Azar E, Woon W-L, Kontokosta CE. Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. *J Build Perform Simul* 2018;11:322–32. <https://doi.org/10.1080/19401493.2017.1354919>.
- Ucuzian AA, Gassman AA, East AT, Greisler HP. Molecular mediators of angiogenesis. *J Burn Care Res.* 2010 Jan-Feb;31(1):158-75. doi: 10.1097/BCR.0b013e3181c7ed82. PMID: 20061852; PMCID: PMC2818794.
- Umar MI, Asmawi MZ, Sadikun A, Majid AM, Al-Suede FS, Hassan LE, et al. 2014. Ethyl-p-methoxycinnamate isolated from *kaempferia galanga* inhibits inflammation by suppressing interleukin-1, tumor necrosis factor- $\alpha$ , and angiogenesis by blocking endothelial functions. *Clinics*. 69(2):134-144.
- Vijayan C, Adersh M, Reji SR, Nair GM. 2013. Screening biological activities of *orthosiphon aristatus*. *International Journal of Pharmacy and Pharmaceutical Sciences* 5(4): 594-600.
- Walkowiak A, Ledzinski L, Zapadka M, Kupcewicz. Detection of adulterants in dietary supplements with *Ginkgo biloba* extract by attenuated total reflectance Fourier transform infrared spectroscopy and multivariate methods PLS-DA and PCA. 222-228: doi:10.1016/j.saa.2018.10.008.
- Zhang Y, Sun S, Dai J, Wenyu W, Huijuan C, Jianbing W, Xiaojun G. 2011. Quality Control Method for Herbal Medicine—Chemical Fingerprint Analysis. 10.5772/23962.
- [WHO] World Health Organization. 2013. WHO Traditional Medicine Strategy 2014-2023. Geneva. Switzerland.

