



IDENTIFIKASI PEMALSUAN KUMIS KUCING DARI BABADOTAN DAN TEKELAN MENGGUNAKAN *MACHINE LEARNING*

ARY PRABOWO



ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2022



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA¹

Dengan ini saya menyatakan bahwa tesis dengan judul “Identifikasi Pemalsuan Kumis Kucing dari Babadotan dan Tekelan Menggunakan *Machine Learning*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2022

Ary Prabowo
G651180464

¹ *Pelimpahan hak cipta atau karya tulis dari penelitian kerjasama dengan pihak luar IPB harus didasarkan pada perjanjian kerjasama yang terkait



RINGKASAN

ARY PRABOWO. Identifikasi Pemalsuan Absorban Kumis Kucing dari dan Tekelan Menggunakan *Machine Learning*. Dibimbing oleh WISNU ANANTA KUSUMA, ANNISA, dan MOHAMAD RAFI.

Identifikasi pemalsuan terhadap tanaman obat merupakan hal penting dalam kendali mutu. Pemalsuan tanaman obat pada kumis kucing (*Orthosiphon Aristatus*) dengan tanaman lain yang mirip seperti babadotan dan tekelan sulit untuk diidentifikasi, hal ini disebabkan karena tingkat kemiripan yang tinggi. Metode yang diterapkan pada penelitian sebelumnya menggunakan PCA dan *discriminant analysis* untuk praproses dan klasifikasi termasuk dalam metode reduksi dimensi. Metode reduksi dimensi memiliki keunggulan yaitu kompresi data, efisiensi waktu komputasi, dan juga menghapus fitur-fitur yang *redundant* namun memiliki kelemahan seperti akan mengubah representasi aslinya, sulit memahami hubungan antar fitur sehingga tidak informatif. Selain itu, metode reduksi dimensi belum bisa memisahkan dengan jelas, tingkat kemiripan yang tinggi sehingga sulit untuk mengidentifikasi pemalsuan kumis kucing dengan tekelan dan babadotan. Penelitian ini menggunakan pendekatan *machine learning* untuk mengklasifikasikan absorban campuran kumis kucing dengan babadotan dan tekelan serta hasil validasi silang tergolong rendah yaitu 74% (Fitria *et al.* 2018).

Penelitian ini menggunakan dataset absorban dari penelitian Fitria (2018) yang terdiri atas 29 sampel yaitu 10 sampel kumis kucing, 10 sampel babadotan, dan 9 sampel tekelan. Tahapan selanjutnya dilakukan praproses dan didapatkan variabel dependen yaitu panjang gelombang sebagai fitur sebanyak 1201, sedangkan variabel independen yaitu kelas. Kelas terdiri atas enam kelas yaitu kelas kumis kucing dengan 5%, 25%, 50% babadotan dan kumis kucing dengan 5%, 25%, 50% tekelan. Data yang telah dilakukan praproses selanjutnya diidentifikasi dengan menggunakan metode klasifikasi antara lain *Support Vector Machine*, *k-NN*, *Naive Bayes*, *Decision Tree*, *Random Forest*, serta seleksi fitur yaitu PCA dan *Recursive Fitur Elimination* (RFE).

Pembentukan model klasifikasi dengan tiga skenario percobaan yang pertama tanpa menggunakan PCA, dengan PCA, dan *Recursive Feature Elimination* untuk dilakukan seleksi fitur yang bertujuan mereduksi dimensi data dan optimalisasi kinerja model. Berdasarkan tiga percobaan tersebut didapat dua model yang terbaik yang pertama metode SVM dan PCA tingkat akurasi 79,55% dengan waktu training dataset 4.300 detik, sedangkan hasil tertinggi didapat dengan metode *Random Forest* dan (RFE) mencapai akurasi 81,82% dengan waktu training 259.000 detik dan jika dibandingkan dengan penelitian sebelumnya mengalami peningkatan sebesar 8%.

Kata kunci: Klasifikasi, *Machine Learning*, Pemalsuan, *Random Forest*, *Recursive Fitur Elimination*.

SUMMARY

ARY PRABOWO. Identification of *Orthosiphon Aristatus Adulteration* by *Babadotan* and *Tekelan* Using *Machine Learning*. Supervised by WISNU ANANTA KUSUMA, ANNISA, and MOHAMAD RAFI.

Identification of adulteration medicinal plants is important in quality control. The adulteration of medicinal plants in Java Tea (*Orthosiphon Aristatus*) with other similar plants such as *babadotan* and *tekelan* is difficult to identify, cause high of similiarity. The method applied in the previous research using PCA and discriminant analysis for preprocessing and classification, it is part of dimension reduction method. The reduction method has the advantages of data compression, computational time efficiency, and removes redundant features but has disadvantages such as changing the original representation, it is difficult to understand the relationship between features, so it is not informative. In addition, the reduction in dimensions has not been able to separate clearly, high of similiarity, so it is difficult to identify adulteration of *Java Tea* with *babadotan* and *tekelan*. This study used a machine learning approach to classify the adulteration of *Java Tea* with *babadotan* and *tekelan* and the cross-validation results are low at 74% (Fitria *et al.* 2018).

This study used the absorbance dataset from Fitria *et al.* (2018) which consists of 29 samples, namely 10 samples of *Java Tea*, 10 samples of *babadotan*, and 9 samples of *tekelan*. In the next stage, preprocessing and the dependent variable is the wavelength as a feature of 1201 features, whereas the independent variable is class. The class consists of 5 classes, *Java Tea* class with 5%, 25%, 50% *babadotan* and *Java Tea* with 5%, 25%, 50% *tekelan*. The preprocessed data is then identified using classification methods as follows Support Vector Machine, k-NN, Naive Bayes, Decision Tree, Random Forest, and feature selection is PCA and Recursive Feature Elimination (RFE).

The formation of a classification model with three experimental scenarios, the first without using PCA, with PCA, and Recursive Feature Elimination for feature selection aimed at reducing dimensions and optimizing model performance. Based on these three experiments, the two best models were obtained, the first SVM and PCA methods, the accuracy rate was 79.55% with a dataset training time of 4,300 seconds, while the highest results were obtained using the Random Forest method and (RFE) achieving an accuracy of 81.82% with training time. 259,000 seconds and if it's compared with previous studies, there was an increase of 8%.

Keywords: Adulterants, Classification, Machine Learning, Random Forest, Recursive Fitur Elimination.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2022
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



IDENTIFIKASI PEMALSUAN KUMIS KUCING DARI BABADOTAN DAN TEKELAN MENGGUNAKAN *MACHINE LEARNING*

ARY PRABOWO

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister Ilmu Komputer pada
Program Studi Ilmu Komputer

**ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2022**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



**Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Toto Haryanto, S.Kom., M.Si.**

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Tesis : Identifikasi Pemalsuan Kumis Kucing dari Babadotan dan Tekelan Menggunakan *Machine Learning*

Nama : Ary Prabowo

NIM : G651180464

Disetujui oleh

Pembimbing 1:

Dr. Eng Wisnu Ananta Kusuma, S.T., M.T.



Pembimbing 2:

Dr. Eng Annisa, S.Kom., M.Kom.



Pembimbing 3:

Dr. Mohamad Rafi, S.Si., M.Si.



Diketahui oleh

Ketua Program Studi:

Irman Hermadi, S. Kom., M.S., Ph.D
NIP 197503112006041009



Dekan Fakultas Matematika dan Ilmu
Pengetahuan Alam:

Dr. Berry Juliandi, S.Si., M.Si.
NIP 19780723 200701 1 001



Tanggal Ujian: 8 Agustus 2022

Tanggal Lulus: 11 AUG 2022

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PRAKATA

Alhamdulillah rabbi'l'alamin, puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Juni 2020 sampai bulan Juli 2022 ini ialah *data mining* dan *bioinformatics* dengan judul “Identifikasi Pemalsuan Kumis Kucing dari Babadotan dan Tekelan Menggunakan *Machine Learning*”.

Terima kasih penulis ucapkan kepada para pembimbing Dr. Eng Wisnu Ananta Kusuma, S.T., M.T, Dr. Eng Annisa, S.Kom., M.Kom., dan Dr. Mohamad Rafi, S.Si., M.Si yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada moderator seminar, dan penguji luar komisi pembimbing. Di samping itu, penghargaan penulis sampaikan kepada Papa Sudyanto (alm), Mama Kartini, Bapak Drs. Sentot Supriyono, M.Pd, Ibu Sri Kinetika M, M.M, Istri Priska Nur Kartika, S.TP, Anak Abdurrahman Aysar dan Muhammad Aqlan Zuhdi, Kak Lia, Kak Yenni, Adik Resty, dan Adik Dwiki Darma S, SP keluarga yang telah membantu, memberikan dukungan doa dan kasih sayang selama penelitian. Ungkapan terima kasih juga disampaikan kepada sahabat dan teman-teman terdekat yang telah membantu serta memberikan dukungan, dan doa untuk penulis dalam melakukan penelitian.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2022

Ary Prabowo

DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	2
II TINJAUAN PUSTAKA	3
2.1 Proses Identifikasi Pemalsuan	3
2.2 Spektrofotometri UV-Vis	3
2.3 Cara Kerja Spektrofotometri UV-Vis	4
2.4 Instrumentasi Spektrofotometri UV-Vis	4
2.5 Absorbansi	5
2.6 Analisis sidik jari	5
2.7 Identifikasi	6
2.8 Lingkup Kerja	6
2.9 Proses <i>Sampling</i> dan preparasi	6
2.10 Pengukuran Sampel Campuran	6
2.11 <i>Feature Selection</i>	7
2.12 <i>Random Forest (RF)</i>	7
2.13 <i>Multi Class Classification</i>	8
2.14 <i>K-Fold Cross Validation</i>	9
2.15 <i>Multiclass Confusion Matrix</i>	9
III METODE	11
3.1 Waktu dan Lokasi Penelitian	11
3.2 Alat dan Bahan	11
3.3 Tahapan Penelitian	11
3.4 Dataset	11
3.5 Praproses data	13
3.6 Pembagian data	13
3.7 Seleksi fitur (<i>Feature Selection</i>)	13
3.8 Klasifikasi	13
3.9 Evaluasi	14
IV HASIL DAN PEMBAHASAN	15
4.1 Praproses Data	15
4.2 Proses Seleksi Fitur	15
4.3 Pembuatan Model Prediksi dan Tuning Parameter	16
4.4 Evaluasi Model <i>Classifier</i>	16
4.5 Analisis Perbandingan Hasil	19
V SIMPULAN DAN SARAN	20
5.1 Simpulan	20
5.2 Saran	20

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



DAFTAR PUSTAKA
LAMPIRAN
RIWAYAT HIDUP

21
24
26

@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Bentuk Umum <i>Confusion Matrix</i> untuk <i>Multiclass</i>	9
2	Spesifikasi perangkat keras	11
3	Kode dan asal sampel	12
4	Representasi dataset Absorban Campuran Kumis Kucing	12
5	Evaluasi model classifier tanpa PCA	16
6	Evaluasi model <i>classifier</i> dengan PCA	16

DAFTAR GAMBAR

1	Kumis kucing beserta dua tanaman pemalsu (Fitria, 2018)	3
2	Diagram skematis spektrofotometer UV-Vis (Underwood 1999)	4
3	Ilustrasi <i>Random Forest</i> (Goel dan Abhilasha 2017)	8
4	Tahapan penelitian	11
5	Prosedur menghitung vote pada metode RFE (Markoulidakis <i>et al.</i> 2021)	13
6	Hasil akurasi menggunakan metode RFE	15
7	hasil akurasi seluruh model	17
8	Hasil perbandingan PCA dan RFE berdasarkan waktu	17
9	ROC pada Metode Random Forest	18
10	Perbandingan hasil akurasi seluruh metode	18

DAFTAR LAMPIRAN

1	Dataset Absorban Campuran Kumis Kucing	25
---	--	----

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.