

PENDEKATAN MODEL GENERATIF PADA *MACHINE LEARNING* UNTUK INDIKASI GEOGRAFIS (IG) MADU LEBAH *STINGLESS* BERBASIS FLUORESENSI

HATA MAULANA



**PROGRAM STUDI DOKTOR ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@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 DISERTASI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa disertasi dengan judul “Pendekatan Model Generatif Pada *Machine Learning* Untuk Indikasi Geografis (IG) Madu Lebah *Stingless* Berbasis Fluoresensi” 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 disertasi ini.

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

Bogor, Juli 2026

Hata Maulana
G6601211009

RINGKASAN

HATA MAULANA. Pendekatan Model Generatif Pada *Machine Learning* untuk Indikasi Geografis (IG) Madu Lebah *Stingless* Berbasis Fluoresensi, Dibimbing oleh YOHANES ARIS PURWANTO, SONY HARTONO WIJAYA, HERU SUKOCO dan DIDING SUHANDY.

Madu merupakan bahan pangan yang dibutuhkan manusia untuk mempertahankan kebutuhan nutrisi dan vitamin dalam tubuh. Manusia dapat memanfaatkan madu baik dikonsumsi secara langsung maupun menjadi campuran bahan makanan lain. Petani madu atau pembudidaya lebah memproduksi madu asli atau langsung dari sarang lebah, kemudian dipasarkan ke perdagangan komersial. Masyarakat dihadapkan pada potensi manipulasi produk madu antara lain terjadinya pemalsuan yang diproduksi tanpa memiliki mekanisme yang jelas dan hanya berfokus pada keuntungan.

Peluang kehadiran sebuah inovasi di sisi hilir dapat membantu secara signifikan dalam pemenuhan kebutuhan produk madu berjenis *stingless* yang terjaga kemurniannya. Standarisasi produk madu, terutama jenis madu lebah tanpa sengat belum dilakukan oleh pihak berwenang di Indonesia. Oleh karena itu, Indikasi Geografis (IG) yang merupakan identitas dari sebuah produk dengan karakteristik spesifik, dalam membedakan satu jenis madu lebah tanpa sengat dari karakteristik produk yang lebih spesifik dan belum dilakukan standarisasi. Hal ini menjadi penting sebagai identitas suatu produk yang setiap wilayah dapat juga memproduksi madu dari jenis lebah, vegetasi dan asal geografis yang berbeda. Akan tetapi, pasar di Indonesia masih sangat memerlukan edukasi terkait hal tersebut. Selain itu IG menjadi salah satu tahap awal sebuah produk madu dengan kualitas kemurnian yang memiliki standar nilai yang lebih spesifik.

Berdasarkan permasalahan tersebut, objek penelitian ini menggunakan madu dari lebah *stingless*, dengan alasan memiliki nilai jual yang lebih tinggi dari madu pada umumnya. Berdasarkan hasil penelusuran literatur, penelitian ini memiliki tujuan untuk membangun dataset dan model klasifikasi IG madu dari lebah *stingless* untuk memudahkan pengujian kualitas, mendukung pemasaran yang diterima masyarakat sehingga membantu peningkatan konsumsi madu lebah *stingless* di Indonesia. Spektroskopi berjenis *ultraviolet* (UV) berbasis fluoresensi dengan rentang panjang gelombang 357-750 nm dikolaborasikan dengan data hasil uji klinis berdasarkan 6 (enam) parameter standar SNI 8664:2018 yaitu nilai enzim diastase, *hydroxymethylfurfural* (HMF), persentase gula reduksi, sukrosa, keasaman, dan kadar air untuk menghasilkan dataset terakuisisi.

Kebaruan dalam penelitian ini adalah akuisisi dataset dengan mengimplementasikan teknik imputasi data kosong (*missing data*) atau data tidak seimbang (*imbalanced data*). selanjutnya data spektrum terakuisisi dilakukan seleksi fitur dengan metode analisis korelasi Pearson antara fitur data spektrum dan data hasil uji klinis. Dataset yang telah diseleksi, pada tahap selanjutnya dilakukan reduksi dimensi dengan algoritma *Probability Principal Component Analysis* (PPCA) dibahas sebagai salah satu jenis algoritma dengan pendekatan model generatif, dan sebagai penentuan nilai pendugaan (*likelihood*) terdekat digunakan algoritma *Expectation Maximization* (EM) untuk mendapatkan keterwakilan fitur utama dari seluruh fitur data spektrum yang terseleksi. Pendekatan model generatif



dengan algoritma PPCA dan EM diusulkan menjadi alternatif solusi dari permasalahan yang dimiliki model diskriminan. Sebagai acuan klasterisasi dari hasil pra-proses tersebut dihasilkan Skor *Silhouette* (SS) adalah 0,64.

Model klasifikasi IG dibangun dengan menggunakan algoritma *machine learning*, kemudian dilakukan perbandingan akurasi dan performa kinerja model yang diterapkan dengan beberapa algoritma klasifikasi populer seperti *Support Vector Machine* (SVM), *Random Forest* (RF), *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* (NB). Akurasi terbaik 98,75% dihasilkan oleh algoritma klasifikasi NB dengan nilai presisi 0,9917 dan *recall* 0,9762 dalam waktu komputasi (*runtime*) 1,702ms. Pemodelan klasifikasi IG produk madu dari lebah *stingless* yang diperoleh dengan nilai akurasi dan performa terbaik, dilengkapi dengan hasil penyusunan karakteristik informasi kualitas madu, hasil uji klinis dominan dan nilai spektrum yang mendeteksi kandungan klinis dominan.

Solusi yang berfokus pada area hilir ini hingga menerapkan model klasifikasi IG menjadi prototipe aplikasi bergerak dalam *platform* android. Hal tersebut diharapkan dapat memberikan kontribusi nyata dan tepat guna untuk meningkatkan kepercayaan masyarakat pada kualitas sebuah madu. Secara tidak langsung tantangan yang dihadapi para pembudidaya berkurang seiring meningkatnya konsumsi madu jenis ini. Sehingga pemerintah dapat merespon dengan pengelolaan standarisasi kualitas kemurnian madu secara nasional.

Kata kunci: fluoresensi, indikasi geografis, *machine learning*, madu lebah *stingless* dan seleksi fitur.



@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.

SUMMARY

HATA MAULANA. Generative Model Approach in Machine Learning for Geographical Indication (GI) of Stingless Honey Bees Based on Fluorescence, Supervised by YOHANES ARIS PURWANTO, SONY HARTONO WIJAYA, HERU SUKOCO and DIDING SUHANDY.

Honey is a food ingredient needed by humans to maintain the body's nutritional and vitamin needs. Humans can benefit from honey either consumed directly or mixed with other foods. Honey farmers or beekeepers produce pure honey directly from the hive, which is then marketed to the commercial trade. The public faces the potential for manipulation of honey products, including counterfeiting, which is processed without clear mechanisms and focused solely on profit.

The potential for downstream innovation could significantly contribute to meeting the demand for stingless honey products with guaranteed purity. Standardization of honey products, particularly stingless honey, has not yet been implemented by the authorities in Indonesia. Therefore, Geographical Indications (GI), which identify a product with specific characteristics, differentiate one type of stingless honey from other, more specific, unstandardized products. This is crucial for product identification, as each region can produce honey from different bee species, vegetation, and geographic origins. However, the Indonesian market still needs education on this matter. Furthermore, GI is one of the initial stages of producing honey with purity quality that has more specific value standards.

Based on these issues, this research used honey from stingless bees as the object of this study, as it has a higher selling value than regular honey. Based on the literature review, this study aims to develop a dataset and GI classification model for stingless bee honey to facilitate quality testing, support market acceptance, and help increase stingless bee honey consumption in Indonesia. Fluorescence-based *ultraviolet* (UV) spectroscopy with a wavelength range of 357-750 nm was combined with clinical test data based on six SNI 8664:2018 standard parameters: diastase enzyme values, *hydroxymethylfurfural* (HMF), reducing sugar percentage, sucrose, acidity, and water content to generate the acquired dataset.

The novelty in this research is the acquisition of datasets by implementing the technique of imputable empty data or unbalanced data. Next, the acquired spectrum data is selected for feature selection using the Pearson correlation analysis method between the spectrum data features and clinical test result data. The selected dataset, in the next stage, is reduced in dimension using the *Probability Principal Component Analysis* (PPCA) algorithm which is discussed as one type of algorithm with a generative model approach, and to determine the closest estimated value (likelihood) the *Expectation Maximization* (EM) algorithm is used to obtain the representation of the main features of all selected data spectrum features. The generative model approach with the PPCA and EM algorithms is proposed as an alternative solution to the problems faced by the discriminant model. As a reference for clustering from the pre-processing results, the *Silhouette* Score (SS) is 0.64.

The GI classification model was built using a machine learning algorithm. The accuracy and performance of the applied model were then compared with several popular classification algorithms, such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and Naive Bayes (NB). The NB



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

classification algorithm achieved the best accuracy of 98,75%, with a *precision* of 0,9917 and a *recall* of 0,9762, and a runtime of 1,702 ms. The GI classification model for honey products from stingless bees, which achieved the best accuracy and performance, was complemented by the results of compiling honey quality information characteristics, dominant clinical test results, and spectral values that detect dominant clinical compounds.

This downstream-focused solution includes the implementation of the GI classification model in a prototype mobile application on the Android platform. This is expected to make a real and effective contribution to increasing public confidence in the quality of honey. Indirectly, the challenges faced by honey farmers will be reduced as consumption of this type of honey increases. Therefore, the government can respond by managing national honey purity quality standards.

Keywords: *fluorescence*, feature selection, *geographical indication*, *machine learning* and stingless bee honey.



@Hak cipta milik IPB University

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.

© Hak Cipta milik IPB, tahun 2026
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.



@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.

PENDEKATAN MODEL GENERATIF PADA *MACHINE LEARNING* UNTUK INDIKASI GEOGRAFIS (IG) MADU LEBAH *STINGLESS* BERBASIS FLUORESENSI

HATA MAULANA

Disertasi
sebagai salah satu syarat untuk memperoleh gelar
Doktor pada
Program Studi Ilmu Komputer

**PROGRAM STUDI DOKTOR ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



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 Tertutup Disertasi:

- 1 Dr.Eng. Annisa, S.Kom., M.Kom.
- 2 Prof. Dr. Indra Budi, S.Kom., M.Kom.

Promotor Luar Komisi Pembimbing pada Sidang Promosi Terbuka Disertasi:

- 1 Dr.Eng. Annisa S.Kom, M.Kom.
- 2 Prof. Dr. Indra Budi, S.Kom., M.Kom.



@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.



Judul Disertasi : Pendekatan Model Generatif pada *Machine Learning* untuk
Indikasi Geografis (IG) Madu Lebah *Stingless* Berbasis
Fluoresensi

Nama : Hata Maulana
NIM : G6601211009

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Yohanes Aris Purwanto M.Sc.

Pembimbing 2:
Dr. Sony Hartono Wijaya, S.Kom, M.Kom.

Pembimbing 3:
Dr.Eng. Heru Sukoco, S.Si., M.T.

Pembimbing 4:
Prof. Dr.Agr.Sc. Diding Suhandy, S.TP, M.Agr.

Diketahui oleh

Ketua Program Studi:
Prof. Dr. Imas Sukaesih Sitanggang, S.Si., M.Kom.
NIP 19750130 199802 2 001

Dekan Sekolah Sains Data, Matematika dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP 19660702 199302 1 001

Tanggal Ujian:
25 Juni 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Judul karya ilmiah "Pendekatan Model Generatif pada *Machine Learning* untuk Indikasi Geografis (IG) Madu Lebah *Stingless* Berbasis Fluoresensi". Dokumen ini merupakan disertasi yang digunakan sebagai satu syarat akademik untuk memperoleh gelar Doktor pada Program Studi Ilmu Komputer. Penulis sebagai manusia biasa menyadari sepenuhnya bahwa dalam penyelesaian penelitian dan penulisan dokumen ini tidak luput dari peran dan bantuan dari berbagai pihak. Oleh karena itu penulis dengan segala kerendahan hati menyampaikan terima kasih yang tulus dan sebesar-besarnya kepada:

1. Prof. Dr. Ir. Yohanes Aris Purwanto M.Sc., selaku Ketua Komisi Pembimbing yang telah meluangkan waktu dan tenaganya untuk memberikan arahan dan membimbing proses pengerjaan penelitian hingga akhir.
 2. Dr. Sony Hartono Wijaya, S.Kom., M.Kom., selaku Pembimbing Anggota, yang telah meluangkan waktu dan tenaganya untuk mengarahkan dan me-review dengan sabar dan teliti, khususnya pada pemodelan *machine learning*.
 3. Dr.Eng. Heru Sukoco, S.Si., M.T., selaku Pembimbing Anggota, yang telah meluangkan waktunya untuk mengingatkan dan me-review proses pengerjaan penelitian, khususnya pada bagian model dan pengembangan *mobile device*.
 4. Prof. Dr.Agr.Sc. Diding Suhandy, S.TP., M.Agr., selaku Pembimbing Anggota, yang telah meluangkan waktu dan tenaga dalam membimbing dan me-review dengan sabar dan teliti, serta memfasilitasi mahasiswa S1 dari UNILA dalam proses pengukuran data spektrum dengan komprehensif.
 5. Prof. Dr. Imas Sukaesih Sitanggang, S.Si., M.Kom., selaku Ketua Program Studi dan juga telah memberikan arahan dalam monitor dan evaluasi secara rutin setiap tengah dan akhir semester.
 6. Dr. Hendra Rahmawan, S.Kom., MT, selaku sekretaris Program Studi yang telah membantu memberikan arahan riset maupun administratif.
 7. Ibu Nining Rohaniah, A.Md sebagai staf administrasi Program Studi Doktor Ilmu Komputer, yang selalu tanggap dan solutif dalam membantu menyampaikan seluruh rangkaian proses administratif penyelesaian disertasi.
 8. Kedua orang tua, Istri, anak-anak dan seluruh saudara tercinta yang selalu memberikan dukungan dan doa terbaik untuk kelancaran proses studi.
 9. Seluruh kawan-kawan seperjuangan S3 Ilmu Komputer, baik angkatan ke-4 maupun tidak, yang selalu terbuka dalam diskusi selama proses penelitian.
 10. Segenap pimpinan dan rekan-rekan dosen Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta yang telah memberikan dukungan penuh.
- Ucapan terima kasih juga penulis sampaikan kepada PT. Suhita, CV. Nabila Natural dan CV. Nutrima Sehat Alami, serta bapak Ujang Bencoy yang telah membantu selama pengumpulan data. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2026

Hata Maulana



DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR ISTILAH	xvi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	5
1.4 Manfaat	5
1.5 Ruang Lingkup	5
1.6 Kebaruan	6
TINJAUAN PUSTAKA	7
2.1 Madu dan Standar Nasional Produk	7
2.2 Indikasi Geografis	8
2.3 UV-Vis spektra	8
2.3 Data Spektra	10
2.4 Pendekatan <i>Machine Learning</i>	11
2.5 Model Diskriminatif	11
2.6 Model Generatif	14
2.7 Analisis Korelasi	15
2.8 Evaluasi Model	16
2.9 Data Kosong	18
2.10 <i>Scrum</i>	18
2.11 Metode Perbandingan Model	19
2.12 <i>State of The Art</i> Penelitian	19
METODE	23
3.1 Waktu dan Tempat Penelitian	23
3.2 Alat, Bahan dan Data	23
3.3 Alur Kerja Penelitian	24
3.4 Arsitektur Sistem	28
3.5 Prosedur Kerja Peningkatan Algoritma	29
3.6 Tahapan Implementasi Algoritma	36
3.7 Analisis data	36
HASIL DAN PEMBAHASAN	38
4.1 Pra-proses dan Peningkatan Struktur Kluster Data	38
4.2 Model Klasifikasi Indikasi Geografis Madu Lebah <i>Stingless</i>	41
4.3 Confusion Matrik pada <i>Support Vector Machine</i>	43
4.4 Confusion Matrik pada <i>Random Forest</i>	43
4.5 Confusion Matrik pada <i>K-Nearest Neighbor</i>	44
4.6 Confusion Matrik pada <i>Naïve Bayes</i>	45
4.7 Evaluasi dan Performa Model Klasifikasi Indikasi Geografis	47
4.8 Algoritma Klasifikasi	49
4.9 Validasi Model (<i>Overfitting</i>)	52
4.10 Hasil Performa Model	52
4.11 Analisis Model	53
4.12 Arsitektur Aplikasi	53

4.13	Pengembangan Aplikasi	54
4.14	Realisasi Aplikasi	56
V	SIMPULAN DAN SARAN	59
5.1	Simpulan	59
5.2	Saran	59
	DAFTAR PUSTAKA	60
	LAMPIRAN	66
	RIWAYAT HIDUP	75

