



# MODEL COMBINATORIAL *BASE LEARNER* STACKING DENGAN OPTIMASI *MODIFIED CUCKOO SEARCH* UNTUK KLASIFIKASI *MULTICLASS* SPEKTRUM FTIR KEJI BELING

I GUSTI NGURAH SENTANA PUTRA



PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2026



## PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Model *Combinatorial Base Learner Stacking* dengan Optimasi *Modified Cuckoo Search* untuk Klasifikasi *Multiclass* Spektrum FTIR Keji Beling” 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, Juni 2026

I Gusti Ngurah Sentana Putra  
M0501241019

## RINGKASAN

I GUSTI NGURAH SENTANA PUTRA. Model *Combinatorial Base Learner Stacking* dengan Optimasi *Modified Cuckoo Search* untuk Klasifikasi *Multiclass* Spektrum FTIR Keji Beling. Dibimbing oleh UTAMI DYAH SYAFITRI, ANANG KURNIA, dan MUHAMMAD RAFI.

Spektroskopi *Fourier Transform Infrared* (FTIR) merupakan teknik analisis kimia yang mampu mengkarakterisasi komposisi molekuler sampel secara cepat dan tidak destruktif, sehingga banyak digunakan dalam bidang farmasi, pangan, dan biologi. Salah satu tantangan utama dalam analisis data FTIR terletak pada karakteristiknya yang berdimensi sangat tinggi (*high dimensional*), di mana jumlah variabel spektral ( $p$ ) jauh melebihi jumlah sampel ( $n$ ), sehingga menimbulkan risiko multikolinearitas, *overfitting*, dan penurunan akurasi prediksi. Tantangan ini semakin kompleks ketika klasifikasi yang dihadapi bersifat *multiclass* dengan distribusi sampel yang tidak seimbang, seperti pada kasus autentikasi tanaman Keji Beling (*Strobilanthes crispus*) yang rentan dicampur dengan Sirih Hutan (*Piper aduncum*) akibat kemiripan morfologi dan selisih nilai ekonomi keduanya. Pendekatan *ensemble stacking* dipandang berpotensi mengatasi keterbatasan model tunggal dalam menangani data berdimensi tinggi, namun pemilihan kombinasi *base learner* yang tidak optimal dapat menurunkan performa model secara signifikan. Selain itu, metode *stacking* konvensional juga menghadapi keterbatasan dalam interpretasi kepentingan fitur akibat sifat *black-box* dari *meta learner*.

Penelitian ini bertujuan untuk menganalisis pengaruh prapemrosesan data terhadap performa klasifikasi spektrum FTIR, membandingkan performa metode *combinatorial base learner stacking* yang dioptimasi dengan algoritma *Modified Cuckoo Search* (MOCS), serta mengidentifikasi gugus fungsi spektral yang membedakan Keji Beling dan Sirih Hutan. Data yang digunakan merupakan data spektrum FTIR-ATR dari 223 amatan yang terdiri atas lima kelas, yaitu 100% Keji Beling, 100% Sirih Hutan, serta campuran Sirih Hutan dengan proporsi 5%, 25%, dan 50%. Prapemrosesan dilakukan dengan empat skenario, yaitu *Savitzky-Golay* (SG), *Standard normal variate* (SNV), serta kombinasi keduanya (SNV+SG dan SG+SNV). Pendekatan *combinatorial base learner stacking* mengeksplorasi semua subset kombinasi dari empat *base learner* dengan dua pilihan *meta learner*, yaitu XGBoost dan regresi logistik. Optimasi *hyperparameter* dilakukan menggunakan algoritma MOCS yang memodifikasi *Cuckoo Search* standar melalui penerapan *dynamic discovery probability* dan fungsi objektif *multi-objective*.

Evaluasi kinerja model dilakukan menggunakan empat metrik utama, yaitu akurasi, *F1-score*, *Balanced Accuracy* (BA), dan *G-mean*. Uji ANOVA empat faktor diterapkan untuk menilai pengaruh prapemrosesan, jenis model, konfigurasi *cuckoo*, dan rasio pembagian data secara simultan beserta interaksinya. Uji lanjut Tukey HSD digunakan untuk mengidentifikasi perbedaan antarperlakuan secara lebih rinci. Validitas dan konsistensi interpretasi kepentingan fitur antara pendekatan *base learner* dan *stacking feature importance aggregation* dinilai menggunakan *Intraclass Correlation Coefficient* (ICC) dan *Kendall's Coefficient of Concordance* (W).

Metode *bagging* dan *boosting* yang dioptimasi dengan MOCS menemukan bahwa prapemrosesan merupakan faktor penentu terhadap performa klasifikasi.



Data tanpa prapemrosesan menghasilkan performa yang relatif rendah pada seluruh algoritma, dengan akurasi terbaik hanya mencapai 0,7548 pada LightGBM di skema 80:20, yang mengindikasikan bahwa variabilitas spektral akibat pergeseran garis dasar dan *noise* belum teratasi. Sebaliknya, penerapan SNV dikombinasikan dengan SG menghasilkan peningkatan performa konsisten pada seluruh algoritma. Dalam perbandingan metode *boosting* dan *bagging*, Analisis komparatif antara MOCS dan *Cuckoo Search* standar (CSO) melalui uji ANOVA menunjukkan bahwa keduanya tidak berbeda signifikan pada seluruh metrik kinerja prediktif, namun MOCS secara signifikan lebih efisien secara komputasi dengan penghematan waktu rata-rata sebesar 347 detik atau 47,56% dibandingkan CSO, sebuah perbaikan yang merepresentasikan prinsip *Pareto improvement* karena kualitas prediksi setara diperoleh dengan biaya komputasi yang jauh lebih rendah. Variasi konfigurasi jumlah *nest* (5, 15, dan 25) tidak memberikan pengaruh signifikan terhadap metrik kinerja pada seluruh skenario, menegaskan bahwa mekanisme pembaruan solusi berbasis *Lévy flight* pada algoritma CS bersifat independen terhadap ukuran populasi dalam rentang yang diuji.

*Combinatorial base learner stacking* menunjukkan bahwa pendekatan ini secara konsisten dan signifikan mengungguli seluruh *base learner* individu maupun *stacking* konvensional pada hampir semua kombinasi prapemrosesan dan rasio pembagian data. Peningkatan performa paling signifikan diperoleh ketika *best combinatorial base learner stacking* berbasis *meta learner* XGBoost dipadukan dengan prapemrosesan SNV+SG atau SG+SNV, menghasilkan akurasi hingga 0,9889 dan BA sebesar 0,9821 pada skema 80:20, dengan deviasi standar yang sangat rendah yang menandakan stabilitas model antar ulangan. Analisis kombinasi *base learner* menunjukkan bahwa penggunaan seluruh *base learner* sekaligus tidak selalu menghasilkan performa tertinggi, yang menegaskan bahwa kualitas dan keberagaman antara model lebih menentukan performa dibandingkan jumlah model dalam *ensemble*. Hasil ANOVA empat faktor mengonfirmasi bahwa prapemrosesan dan jenis model merupakan faktor paling dominan terhadap seluruh metrik kinerja, interaksi keduanya signifikan, sementara konfigurasi *cuckoo* tidak berpengaruh signifikan terhadap metrik kinerja namun berpengaruh terhadap waktu komputasi. Pendekatan *stacking feature importance aggregation* yang menggabungkan nilai SHAP dari *base learner* dengan bobot koefisien *meta learner* berhasil meningkatkan konsistensi kepentingan fitur secara signifikan, dengan ICC meningkat sebesar 26,4% dan Kendall's W meningkat 7,7% dibandingkan pendekatan tanpa *stacking*, yang memperkuat validitas kimiawi interpretasi model. Dua wilayah bilangan gelombang paling kritis yang teridentifikasi sebagai *spectral marker* pembeda utama kedua tanaman adalah daerah aromatik ( $\sim 1602-1606\text{ cm}^{-1}$ ) yang mencerminkan dominasi senyawa fenolik/flavonoid kompleks pada Keji Beling, dan daerah C–O/C–N ( $\sim 1223-1225\text{ cm}^{-1}$ ) yang berkaitan dengan sidik jari alkaloid aromatik pada Sirih Hutan, sehingga gugus fungsi pada kedua wilayah tersebut dapat ditetapkan sebagai penanda spektral yang andal untuk keperluan autentikasi berbasis FTIR.

Kata kunci: *Cuckoo Search*, data spektrum FTIR, *ensemble stacking*, kepentingan fitur, klasifikasi *multiclass*, *combinatorial base learner*.

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.

## SUMMARY

I GUSTI NGURAH SENTANA PUTRA. Combinatorial Base Learner Stacking with Modified Cuckoo Search Optimization for Multiclass Classification of Keji Beling FTIR Spectra. Supervised by UTAMI DYAH SYAFITRI, ANANG KURNIA, and MUHAMMAD RAFI.

Fourier Transform Infrared (FTIR) spectroscopy is a chemical analysis technique capable of characterizing the molecular composition of samples rapidly and non-destructively, making it widely used in pharmaceuticals, food science, and biology. One of the main challenges in FTIR data analysis lies in its high-dimensional characteristics, where the number of spectral variables ( $p$ ) far exceeds the number of samples ( $n$ ), leading to risks of multicollinearity, overfitting, and reduced prediction accuracy. This challenge becomes even more complex when dealing with multiclass classification under imbalanced sample distribution, as in the case of authenticating Keji Beling (*Strobilanthes crispus*), which is prone to adulteration with Sirih Hutan (*Piper aduncum*) due to morphological similarities and differences in economic value. Ensemble stacking is considered a promising approach to overcome the limitations of single models in handling high-dimensional data; however, suboptimal selection of base learner combinations can significantly degrade model performance. Furthermore, conventional stacking methods also face limitations in feature importance interpretation due to the black-box nature of the meta-learner.

This study aims to analyze the effect of data preprocessing on FTIR spectrum classification performance, compare the performance of the combinatorial base learner stacking method optimized with the Modified Cuckoo Search (MOCS) algorithm, and identify the spectral functional groups that distinguish Keji Beling from Sirih Hutan. The data used are FTIR-ATR spectra from 223 samples consisting of five classes: 100% Keji Beling, 100% Sirih Hutan, and mixtures of Sirih Hutan at 5%, 25%, and 50% proportions. Preprocessing was performed using four scenarios: Savitzky-Golay (SG), Standard normal variate (SNV), and their combinations (SNV+SG and SG+SNV). The combinatorial base learner stacking approach explored all subset combinations of four base learners with two meta-learner options, XGBoost and logistic regression. Hyperparameter optimization was performed using the MOCS algorithm, which modifies the standard Cuckoo Search by applying dynamic discovery probability and a multi-objective function. Model performance evaluation was conducted using four main metrics: accuracy, F1-score, Balanced Accuracy (BA), and G-mean. Four-factor ANOVA was applied to assess the effects of preprocessing, model type, cuckoo nest, and split ratio simultaneously, along with their interactions. Tukey's HSD post-hoc test was used to identify differences between treatments in more detail. The validity and consistency of feature importance interpretation between the base learner and stacking feature importance aggregation approaches were assessed using the Intraclass Correlation Coefficient (ICC) and Kendall's Coefficient of Concordance (W).

The bagging and boosting methods optimized with MOCS found that preprocessing is a determining factor for classification performance. Data without preprocessing yielded relatively low performance across all algorithms, with the



best accuracy reaching only 0.7548 for LightGBM under the 80:20 split scheme, indicating that spectral variability due to baseline shift and noise remained problematic. In contrast, the application of SNV combined with SG consistently improved performance across all algorithms. In the comparative analysis between boosting and bagging methods, ANOVA tests showed that MOCS and standard Cuckoo Search (CSO) did not differ significantly across all predictive performance metrics; however, MOCS was significantly more computationally efficient, achieving an average time saving of 347 seconds or 47.56% compared to CSO an improvement representing a Pareto improvement principle, as equivalent prediction quality was obtained at a much lower computational cost. Variations in the number of nests (5, 15, and 25) did not significantly affect performance metrics across all scenarios, confirming that the Lévy flight-based solution update mechanism in the CS algorithm is independent of population size within the tested range.

Combinatorial base learner stacking showed that this approach consistently and significantly outperformed all individual base learners as well as conventional stacking across almost all preprocessing and split ratio combinations. The most significant performance improvement was achieved when the best combinatorial base learner stacking based on the XGBoost meta-learner was combined with SNV+SG or SG+SNV preprocessing, yielding accuracy up to 0.9889 and BA of 0.9821 under the 80:20 split scheme, with very low standard deviation indicating model stability across repetitions. Analysis of base learner combinations showed that using all base learners simultaneously did not always yield the highest performance, confirming that the quality and diversity among models are more determinant of performance than the number of models in the ensemble. Four-factor ANOVA results confirmed that preprocessing and model type are the most dominant factors affecting all performance metrics, with their interaction being significant, while cuckoo nest had no significant effect on performance metrics but did affect computational time. The stacking feature importance aggregation approach, which combines SHAP values from base learners with meta-learner coefficient weights, significantly improved feature importance consistency, with ICC increasing by 26.4% and Kendall's W by 7.7% compared to the non-stacking approach, thereby strengthening the chemical validity of model interpretation. The two most critical wavenumber regions identified as the main spectral markers distinguishing the two plants are the aromatic region ( $\sim 1602\text{--}1606\text{ cm}^{-1}$ ), reflecting the dominance of complex phenolic/flavonoid compounds in Keji Beling, and the C–O/C–N region ( $\sim 1223\text{--}1225\text{ cm}^{-1}$ ), associated with the aromatic alkaloid fingerprint in Sirih Hutan. Thus, the functional groups in these two regions can be established as reliable spectral markers for FTIR-based authentication purposes.

**Keywords:** Cuckoo Search, FTIR spectral data, ensemble stacking, feature importance, multiclass classification, combinatorial base learner.

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



@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.*



# **MODEL COMBINATORIAL *BASE LEARNER* STACKING DENGAN OPTIMASI *MODIFIED CUCKOO SEARCH* UNTUK KLASIFIKASI MULTICLASS SPEKTRUM FTIR KEJI BELING**

**I GUSTI NGURAH SENTANA PUTRA**

Tesis  
sebagai salah satu syarat untuk memperoleh gelar  
Magister pada  
Program Studi Statistika dan Sains Data

**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA  
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.

Penguji Luar Komisi Pembimbing pada Ujian Tesis:  
Dr. Bagus Sartono, S.Si, M.Si



Judul Tesis : Model *Combinatorial Base Learner Stacking* dengan Optimasi  
*Modified Cuckoo Search* untuk Klasifikasi *Multiclass* Spektrum  
FTIR Keji Beling  
Nama : I Gusti Ngurah Sentana Putra  
NIM : M0501241019

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:  
Dr. Utami Dyah Syafitri, S.Si., M.Si

\_\_\_\_\_

Pembimbing 2:  
Prof. Dr. Anang Kurnia, S.Si., M.Si

\_\_\_\_\_

Pembimbing 3:  
Prof. Dr. Mohamad Rafi, S.Si., M.Si

\_\_\_\_\_

Diketahui oleh

Ketua Program Studi:  
Dr. Agus Mohamad Soleh, S.Si, M.T.  
NIP 197503151999031004

\_\_\_\_\_

Dekan Sekolah Sains Data, Matematika, Dan Informatika :  
Prof. Dr. Ir. Agus Bueno, M.Si., M.Kom.  
NIP 196607021993021001

\_\_\_\_\_

Tanggal Ujian: 8 Mei 2026

Tanggal Lulus:



@Hak cipta milik IPB University

IPB University



IPB University  
— Bogor Indonesia —

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.



## PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala karunia-Nya sehingga karya ilmiah dengan judul “Model *Combinatorial Base Learner Stacking* dengan Optimasi *Modified Cuckoo Search* untuk Klasifikasi *Multiclass* Spektrum FTIR Keji Beling” berhasil diselesaikan. Karya ilmiah ini penulis susun sebagai salah satu syarat untuk menyelesaikan studi Magister Sains pada Program Studi Statistika dan Sains Data. Penulisan Karya ilmiah ini dapat terselesaikan tidak terlepas dari dukungan, bimbingan, dan bantuan berbagai pihak sehingga penulis menyampaikan ucapan terima kasih kepada:

1. Dr. Utami Dyah Syafitri, S.Si., M.Si, Prof. Dr. Anang Kurnia, S.Si., M.Si dan Prof. Dr. Mohamad Rafi, S.Si., M.Si selaku komisi pembimbing yang senantiasa bersedia meluangkan waktunya untuk memberikan bimbingan, masukan, dukungan, motivasi, dan arahan kepada penulis;
2. Dr. Agus M. Soleh, S.Si., M.T., Dr. Bagus Sartono, S.Si., M.Si., Dr. Septian Rahardianto, S.Stat., M.Si., dan Sachnaz Desta Oktarina, M.Agr., Ph.D., serta seluruh dosen pengajar dan staf tata usaha Program Studi Statistika dan Sains Data atas ilmu, bantuan, dan kerja samanya selama perkuliahan.
3. Terima kasih kepada LPDP RI atas dukungan moril dan material selama proses studi dan menyelesaikan tugas akhir ini
4. Terima kasih kepada penulis, I Gusti Ngurah Sentana Putra dengan penuh kesadaran, perjuangan keras dan resiliensi menyelesaikan tugas akhir ini, meskipun di tengah proses penulisan menghadapi berbagai tantangan dan hambatan.
5. Terima kasih kepada Ajik, Ibu dan Adik yang selalu memberikan doa & motivasi kepada penulis
6. Terima kasih kepada Ida Ayu Made Sinthya Dewi yang selalu menjadi teman bercerita & berdiskusi penulis serta memberikan dukungan dan selalu melimpahkan doa dan kasih sayangnya kepada penulis.
7. Rekan-rekan mahasiswa Pascasarjana Program Studi Statistika dan Sains Data angkatan 2024, Himpro Pascasarjana Statistika dan Kelurahan LPDP 12.0 terkhusus divisi sosial masyarakat yang telah memberikan semangat dan bantuan serta semua pihak yang telah banyak membantu dan tak sempat penulis sebutkan satu per satu.

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

Bogor, 2026

*I Gusti Ngurah Sentana Putra*

## DAFTAR ISI

|   |           |
|---|-----------|
| DAFTAR ISI  | xi        |
| DAFTAR TABEL  | xiii      |
| DAFTAR GAMBAR   | xiii      |
| DAFTAR LAMPIRAN   | xiv       |
| <b>I PENDAHULUAN</b>  | <b>1</b>  |
| 1.1. Latar Belakang   | 1         |
| 1.2. Rumusan Masalah  | 3         |
| 1.3. Tujuan   | 4         |
| 1.4. Manfaat  | 4         |
| <b>II TINJAUAN PUSTAKA</b>  | <b>5</b>  |
| 2.1. Keji Beling  | 5         |
| 2.2. Sirih Hutan  | 5         |
| 2.3. FTIR-ATR   | 6         |
| 2.4. Data Dimensi Tinggi  | 6         |
| 2.5. Prapemrosesan Data Spektrum  | 9         |
| 2.6. Principal Component Analysis   | 9         |
| 2.7. Metode <i>Ensemble</i>   | 11        |
| 2.8. <i>Boosting</i>  | 12        |
| 2.9. Bootstrap Aggregating ( <i>Bagging</i> )   | 15        |
| 2.10. Stacking  | 17        |
| 2.11. Optimization  | 18        |
| 2.12. <i>Metaheuristic</i>  | 19        |
| 2.13. <i>Cuckoo Search Optimization</i>   | 20        |
| 2.14. Ukuran performa model   | 23        |
| 2.15. Ukuran reliabilitas dan konsistensi   | 24        |
| 2.16. Metrik Evaluasi Clustering  | 25        |
| 2.17. SHapley Additive exPlanations (SHAP)  | 26        |
| <b>III METODE</b>   | <b>28</b> |
| 3.1. Data   | 28        |
| 3.2. <i>Modified Cuckoo Search Optimization</i>   | 28        |
| 3.3. <i>Combinatorial Base learner Stacking</i>   | 33        |
| 3.4. <i>Feature importance Aggregation</i>  | 35        |
| 3.5. Prosedur Analisis Data   | 37        |
| <b>IV MODIFIED CUCKOO SEARCH BAGGING DAN BOOSTING UNTUK<br/>AUTENTIKASI KEJI BELING BERBASIS FTIR</b> | <b>40</b> |
| 4.1. Eksplorasi Data  | 40        |
| 4.2. Kajian <i>Savitzky-Golay</i>   | 44        |
| 4.3. Arsitektur dan Konfigurasi   | 45        |
| 4.4. Perbandingan <i>Cuckoo Search</i> dengan <i>Modified Cuckoo Search</i>                           | 46        |
| 4.5. Performa <i>Bagging</i> dan <i>Boosting</i> dengan <i>Modified Cuckoo Search</i>                 | 47        |
| 4.6. Uji Statistik  | 52        |
| 4.7. <i>Feature importance</i>  | 53        |



|  |   |    |
|--|---|----|
| 4.8.   | Perbandingan Metode dan Limitasi Metode <i>Bagging Boosting</i> | 55 |
| V COMBINATORIAL <i>BASE LEARNER STACKING</i> UNTUK AUTENTIKASI KEJI BELING BERBASIS FTIR |   | 57 |
| 5.1.   | Evaluasi <i>Combinatorial Base learner Stacking</i>             | 57 |
| 5.2.   | Uji Statistik Model   | 65 |
| 5.3.   | <i>Feature importance</i>                                       | 68 |
| 5.4.   | Perbandingan <i>Feature importance</i>                          | 69 |
| VI SIMPULAN  |   | 71 |
| 6.1.   | Simpulan  | 71 |
| 6.2.   | Saran   | 72 |
| DAFTAR PUSTAKA   |   | 73 |
| LAMPIRAN   |   | 82 |
| RIWAYAT HIDUP  |   | 95 |

## DAFTAR TABEL

|    |  |    |
|----|--|----|
| 1  | <i>Confusion Matrix</i> Untuk Klasifikasi <i>Multiclass</i> (Blanchard <i>et al.</i> 2022)   | 23 |
| 2  | Ukuran Performa Model <i>Multiclass</i>  | 23 |
| 3  | Daftar Peubah Penelitian   | 28 |
| 4  | Proporsi Jumlah Sampel Penelitian  | 40 |
| 5  | Perbandingan Metrik Evaluasi Clustering pada Berbagai Metode Prapemrosesan Spektral  | 41 |
| 6  | Statistik Deskriptif SNR per Kelas Sampel  | 43 |
| 7  | Distribusi Tingkat Korelasi Antar <i>Feature</i> FTIR  | 44 |
| 8  | Kombinasi 5 terbaik parameter SG dengan metode <i>stacking</i>   | 45 |
| 9  | Ruang Pencarian <i>Hyperparameter</i> dan Konfigurasi Optimasi <i>Base learner</i>   | 46 |
| 10 | Perbandingan One-Way ANOVA MCSO dengan CSO   | 46 |
| 11 | Perbandingan Kinerja <i>Base Learner</i> dengan Rasio data Latih dan Uji 70:30 (Mean $\pm$ Standar Deviasi)                                    | 48 |
| 12 | Perbandingan Kinerja <i>Base Learner</i> dengan Rasio data Latih dan Uji 80:20 (Mean $\pm$ Standar Deviasi)                                    | 49 |
| 13 | Hasil Nilai <i>F-statistics</i> Uji ANOVA Model <i>Bagging</i> dan <i>Boosting</i> untuk Setiap Faktor, Interaksi, dan Metrik Evaluasi         | 52 |
| 14 | Hasil <i>feature importance</i> dan identifikasi gugus fungsi FTIR   | 54 |
| 15 | Kelebihan dan Kekurangan Metode <i>Bagging &amp; Boosting</i>  | 55 |
| 16 | Perbandingan Kinerja Model <i>Stacking</i> dengan Rasio Data Latih dan Uji 70:30 (Mean $\pm$ Standar Deviasi)                                  | 58 |
| 17 | Perbandingan Kinerja Model <i>Stacking</i> dengan Rasio Data Latih dan Uji 80:20 (Mean $\pm$ Standar Deviasi)                                  | 59 |
| 18 | Jumlah Kombinasi <i>Base Learner</i> Terbaik Berdasarkan <i>Meta Learner</i> serta Rasio Pembagian Data  | 61 |
| 19 | Hasil Nilai <i>F-statistics</i> Uji ANOVA Model <i>combinatorial base learner stacking</i> untuk Setiap Faktor, Interaksi, dan Metrik Evaluasi | 65 |
| 20 | Hasil Uji <i>Post Hoc Tukey HSD</i> untuk Metrik Evaluasi berdasarkan Faktor Praproses, Model Type, Nest Cuckoo, dan Rasio Pembagian Data      | 67 |
| 21 | Interpretasi FTIR Berdasarkan Area Senyawa Pembeda   | 68 |

## DAFTAR GAMBAR

|   |   |    |
|---|---|----|
| 1 | Perbandingan daun keji beling dan sirih hutan (Nurraihana dan Norfarizan-Hanoon 2013; Monzote <i>et al.</i> 2017) | 6  |
| 2 | Estimator 25-Tetangga Terdekat (Giraud 2021)  | 8  |
| 3 | Ilustrasi penambahan dimensi (Giraud 2021)  | 8  |
| 4 | Contoh Proses PCA (Gewers <i>et al.</i> 2022)   | 9  |
| 5 | Komponen dasar dalam transformasi PCA (Gewers <i>et al.</i> 2022)   | 10 |
| 6 | Metode <i>ensemble</i>  | 12 |
| 7 | Metode <i>Boosting</i>  | 13 |
| 8 | XGBoost   | 13 |



|    |  |    |
|----|--|----|
| 9  | LightGBM   | 14 |
| 10 | Metode <i>Bagging</i>  | 15 |
| 11 | Random Forest  | 16 |
| 12 | <i>ExtraTrees</i>  | 17 |
| 13 | Metode <i>Stacking</i> (Satofa dan Satofa 2024)  | 18 |
| 14 | Fungsi <i>Multimodal</i> (Giraud 2021)   | 19 |
| 15 | Alur <i>Modified Cuckoo Search</i>   | 32 |
| 16 | <i>Combinatorial base learner stacking</i>   | 35 |
| 17 | <i>Feature importance Aggregation</i>  | 36 |
| 18 | Diagram Alur Tahapan Analisis <i>Combinatorial Base learner Stacking</i>   | 39 |
| 19 | Visualisasi PCA Berdasarkan Metode Prapemrosesan   | 41 |
| 20 | Distribusi SNR per Kelas Sampel  | 42 |
| 21 | Profil <i>Noise</i> Rata-rata Semua Sampel per Region FTIR   | 43 |
| 22 | <i>Boxplot</i> perbandingan waktu MCSO dan CSO   | 47 |
| 23 | <i>Biplot</i> Evaluasi Kinerja Model <i>Base Learner</i> Berdasarkan Metrik Accuracy, F1, <i>Balanced Accuracy</i> , dan G-Mean pada Reduksi Dimensi PCA                                       | 51 |
| 24 | <i>Biplot</i> Evaluasi Kinerja Prapemrosesan <i>Base Learner</i> Berdasarkan Metrik Accuracy, F1, <i>Balanced Accuracy</i> , dan G-Mean pada Reduksi Dimensi PCA                               | 51 |
| 25 | 10 <i>Feature</i> terpenting setiap model  | 53 |
| 26 | Lokasi Region <i>Feature importance</i> LightGBM   | 54 |
| 27 | <i>Biplot</i> Evaluasi Kinerja Model <i>Base Learner</i> dan <i>Meta Learner</i> Berdasarkan Metrik Accuracy, F1 Score, <i>Balanced Accuracy</i> , dan G-Mean pada Reduksi Dimensi PCA         | 60 |
| 28 | <i>Biplot</i> Evaluasi Kinerja Prapemrosesan <i>Base Learner</i> dan <i>Meta Learner</i> Berdasarkan Metrik Accuracy, F1 Score, <i>Balanced Accuracy</i> , dan G-Mean pada Reduksi Dimensi PCA | 61 |
| 29 | <i>Boxplot</i> <i>Meta learner Stacking</i>  | 63 |
| 30 | <i>Boxplot</i> <i>Meta learner stacking</i> berdasarkan prapemrosesan  | 63 |
| 31 | <i>Boxplot</i> Perbandingan Kinerja Stacking antara <i>Meta Learner</i> pada Skenario Seluruh Kombinasi, 4 Model Terbaik, dan Kombinasi Optimal  | 64 |
| 32 | Region 10 <i>Feature importance Aggregation</i>  | 68 |
| 33 | Perbandingan ICC & Kendall's W tanpa dan dengan <i>stacking</i>  | 70 |
| 34 | Matriks korelasi spearman antar model <i>Feature importance</i>  | 70 |

## DAFTAR LAMPIRAN

|   |   |    |
|---|---|----|
| 1 | Lampiran 1 Perbandingan <i>Cuckoo Search</i> dengan <i>Modified Cuckoo Search</i> | 83 |
| 2 | Lampiran 2 Hasil hyperparameter CSO dan MCSO Extratrees                           | 84 |
| 3 | Lampiran 3 Kombinasi Setiap Ulangan Stacking                                      | 93 |

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