

PENGEMBANGAN MODEL PREDIKSI STATUS KEBUNTINGAN SAPI MENGGUNAKAN ALGORITMA RANDOM FOREST DAN LOGISTIC REGRESSION

CHEISHA AMANDA



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@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 SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pengembangan Model Prediksi Status Kebuntingan Sapi Menggunakan Algoritma *Random Forest* dan *Logistic Regression*” 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 skripsi ini.

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

Bogor, Juli 2025

Cheisha Amanda
G6401211084



ABSTRAK

CHEISHA AMANDA. Pengembangan Model Prediksi Status Kebuntingan Sapi Menggunakan Algoritma *Random Forest* dan *Logistic Regression*. Dibimbing oleh DEAN APRIANA RAMADHAN dan MUHAMMAD AGIL.

Berdasarkan data dari Kementerian Pertanian, pada tahun 2024, produksi daging sapi mencapai 416,7 ribu ton, dengan defisit mencapai 291,3 ribu ton. Kondisi ini memerlukan peningkatan signifikan dalam produktivitas peternakan sapi, dimana rendahnya tingkat keberhasilan reproduksi menjadi tantangan utama. Penelitian ini bertujuan mengembangkan model prediksi status kebuntingan sapi menggunakan algoritma *Random Forest* dan *Logistic Regression* berdasarkan data historis sapi yang telah melalui tahapan *preprocessing*. Kedua model menunjukkan performa yang baik, dengan *Random Forest* mencatat akurasi 82%, dan *F1 Score* 0,87 setelah penyesuaian *threshold* optimal, mengungguli *Logistic Regression* (akurasi 80% dan *F1 Score* 0,86). Hasil prediksi probabilitas yang terkalibrasi ini dibagi ke dalam tiga kategori (observasi, evaluasi awalan, evaluasi lanjutan) untuk mendukung pengambilan keputusan yang lebih strategis di lapangan. Dengan demikian, penelitian ini berhasil mengembangkan alat prediktif yang akurat dan aplikatif sebagai sistem pendukung keputusan, yang berkontribusi dalam peningkatan produktivitas sapi.

Kata kunci: *logistic regression*, prediksi kebuntingan, produktivitas sapi, *random forest*, sistem pendukung keputusan

ABSTRACT

CHEISHA AMANDA. Development of a Cow Pregnancy Status Prediction Model Using Random Forest and Logistic Regression. Supervised by DEAN APRIANA RAMADHAN and MUHAMMAD AGIL.

Based on data from the Ministry of Agriculture, beef production in 2024 reached 416.7 thousand tons, with a deficit of 291.3 thousand tons, highlighting the need to boost cattle farm productivity. One major challenge is the low reproductive success rate. This study developed predictive models for cattle pregnancy using Random Forest and Logistic Regression. Both models demonstrated good performance, with Random Forest achieving an accuracy of 82% and an F1 Score of 0.87 after optimal threshold adjustment, outperforming Logistic Regression (80% accuracy and 0.86 F1 Score). The predicted probabilities were classified into three levels (observation, initial evaluation, and follow-up evaluation) to support strategic decisions, making the tool a practical, data-driven solution to improve cattle productivity.

Keywords: cattle productivity, decision support system, logistic regression, pregnancy prediction, random forest



- 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 2025
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.

**PENGEMBANGAN MODEL PREDIKSI STATUS
KEBUNTINGAN SAPI MENGGUNAKAN ALGORITMA
RANDOM FOREST DAN LOGISTIC REGRESSION**

CHEISHA AMANDA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@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.

Tim Penguji pada Ujian Skripsi:
Hari Agung Adrianto, SKom, MSi, Ph.D

Judul Skripsi : Pengembangan Model Prediksi Status Kebuntingan Sapi
Menggunakan Algoritma *Random Forest* dan *Logistic Regression*

Nama : Cheisha Amanda
NIM : G6401211084

Disetujui oleh

Pembimbing 1:
Dean Apriana Ramadhan, S.Komp., M.Kom.

Pembimbing 2:
Dr. drh. Muhammad Agil, M.Sc.Agr., Dipl.ACCM



Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Komp., M.Kom.
198108092008121002



Tanggal Ujian:
16 Juli 2025

Tanggal Lulus:



PRAKATA

Segala puji dan syukur penulis panjatkan ke hadirat Allah *subhanahu wa ta'ala* atas rahmat dan karunia-Nya, sehingga penulisan karya ilmiah ini dapat diselesaikan dengan baik. Penelitian ini telah dilakukan sejak Desember 2024 hingga Juli 2025 dengan judul “Pengembangan Model Prediksi Status Kebuntingan Sapi Menggunakan Algoritma *Random Forest* dan *Logistic Regression*.”

Penulis ingin menyampaikan apresiasi dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Kholid Tamjis dan Ibu Ery Yuni Murniasih, selaku orang tua penulis, serta kakak dan adik, atas doa, dukungan, dan semangat yang diberikan selama proses studi dan penyusunan karya ilmiah ini.
2. Dean Apriana Ramadhan, S.Kom., M.Kom., dan Dr. drh. Muhammad Agil, M.Sc.Agr., Dipl.ACCM, selaku dosen pembimbing yang senantiasa membimbing dengan penuh kesabaran dan memberikan masukan yang sangat berarti.
3. Hari Agung Adrianto, S.Kom., M.Si., Ph.D., selaku dosen penguji, atas kritik membangun dan saran yang sangat bermanfaat dalam penyempurnaan karya ini.
4. Direktur utama dan tim manajemen *breeding* PT. Lembu Jantan Perkasa yang telah mengizinkan untuk penggunaan data sapi yang digunakan dalam pengembangan model penelitian ini.
5. Rekan-rekan terdekat, atas kerja sama, dukungan moral, serta dorongan semangat dalam menyelesaikan penelitian ini.
6. Teman-teman Ilmu Komputer angkatan 58, atas kerja sama dan kebersamaan selama masa studi.
7. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan, baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa karya ilmiah ini masih memiliki keterbatasan dan kekurangan. Namun demikian, penulis berharap karya ini dapat memberikan kontribusi nyata dalam mendukung peningkatan efisiensi reproduksi ternak sapi melalui penerapan teknologi prediktif, serta menjadi referensi bagi pengembangan sistem pendukung keputusan di bidang peternakan dan *data science*.

Bogor, Juli 2025

Cheisha Amanda



DAFTAR ISI

DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
II TINJAUAN PUSTAKA	4
2.1 Reproduksi Sapi	4
2.2 <i>Random Forest</i>	7
2.3 <i>Logistic Regression</i>	9
2.4 Evaluasi Model	11
III METODE	13
3.1 Data Penelitian	13
3.2 Definisi Operasional	13
3.3 Tahapan Penelitian	14
3.4 Lingkungan Pengembangan	16
IV HASIL DAN PEMBAHASAN	17
4.1 Pengolahan Data Mentah	17
4.2 <i>Load Data dan Exploratory Data Analysis (EDA)</i>	19
4.3 <i>Data Preprocessing</i>	23
4.4 <i>Data Visualization</i>	26
4.5 <i>Build Model</i>	27
4.6 Evaluasi model	28
4.7 <i>Testing</i>	33
V SIMPULAN DAN SARAN	36
5.1 Simpulan	36
5.2 Saran	36
DAFTAR PUSTAKA	37
LAMPIRAN	40
RIWAYAT HIDUP	46



DAFTAR GAMBAR

1 Anatomi saluran reproduksi sapi betina	4
2 Anatomi saluran reproduksi sapi jantan	5
3 Struktur algoritma <i>random forest</i>	8
4 Struktur algoritma <i>logistic regression</i>	9
5 <i>Confusion matrix</i>	11
6 Alur diagram tahapan penelitian	14
7 Tampilan data mentah file Identitas Sapi	17
8 Tampilan data mentah file Sapi IB dan Semen	17
9 Tampilan data mentah file Sapi Partus dan Tanggal Partus	17
10 Tampilan data mentah file Sapi PKB	18
11 Tampilan dataset akhir	19
12 Tahap eksplorasi data: identifikasi variabel kategorikal	20
13 Tahap eksplorasi data: missing values pada variabel kategorikal	20
14 Tahap eksplorasi data: analisis deskriptif variabel kategorikal	21
15 Tahap eksplorasi data: visualisasi distribusi variabel kategorikal	22
16 Tahap eksplorasi data: identifikasi variabel numerik	22
17 Tahap eksplorasi data: missing values pada variabel numerik	22
18 Tahap eksplorasi data: analisis deskriptif variabel numerik	23
19 Tahap eksplorasi data: visualisasi distribusi variabel numerik	23
20 Contoh data hasil OneHotEncoder (kiri: sebelum vs kanan: setelah)	25
21 <i>Pipeline preprocessing</i> (kotak atas)	25
22 Tampilan (sebagian) data hasil <i>preprocessing</i>	26
23 Distribusi variabel target	26
24 <i>Pipeline preprocessing</i> dan <i>training</i> model (<i>logistic regression</i> dan <i>random forest</i>)	27
25 Evaluasi metrik utama hasil <i>training</i> model <i>random forest</i>	29
26 Evaluasi metrik utama setelah optimasi <i>threshold</i> model <i>random forest</i>	29
27 <i>Confusion matrix</i> hasil <i>training</i> dan setelah optimasi <i>threshold</i> model <i>random forest</i>	29
28 ROC curve model <i>random forest</i>	30
29 <i>Learning curve</i> model <i>random forest</i>	30
30 Evaluasi metrik utama hasil <i>training</i> model <i>logistic regression</i>	31
31 Evaluasi metrik utama setelah optimasi <i>threshold</i> model <i>logistic regression</i>	31
32 <i>Confusion matrix</i> hasil <i>training</i> dan setelah optimasi <i>threshold</i> model <i>logistic regression</i>	31
33 ROC curve model <i>logistic regression</i>	32
34 <i>Learning curve</i> model <i>logistic regression</i>	32
35 Tampilan 10 teratas dari hasil pengujian awal	33



36 KDE plot hasil keseluruhan pengujian awal di kedua model (<i>random forest</i> dan <i>logistic regression</i>)	34
37 Tampilan hasil prediksi menggunakan salah satu data uji	35

DAFTAR LAMPIRAN

1 Data <i>dictionary</i> identitas sapi	41
2 Data <i>dictionary</i> sapi ib	42
3 Data <i>dictionary</i> semen	43
4 Data <i>dictionary</i> tanggal partus	43
5 Data <i>dictionary</i> sapi partus	44
6 Data <i>dictionary</i> sapi pkb	45