



**PEMANDANAN SKOR PROPENSITAS MENGGUNAKAN
RANDOM FOREST DALAM EVALUASI PENGARUH
KONSUMSI DAUN SUNGKAI TERHADAP KESEMBUHAN
COVID-19**

RAFIKA DAMAYANTI SURURIN NUFUS



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2025**

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 CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Pemadanan Skor Propensitas menggunakan *Random Forest* dalam Evaluasi Pengaruh Konsumsi Daun Sungkai Terhadap Kesembuhan Covid-19” 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, 13 Januari 2025

Rafika Damayanti Sururin Nufus
NIM G1501211041

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 mengikuti kepentingan yang wajar IPB University.



RINGKASAN

RAFIKA DAMAYANTI SURURIN NUFUS. Pemadanan Skor propensitas Menggunakan *Random Forest* dalam Evaluasi Pengaruh Konsumsi Daun Sungkai terhadap Kesembuhan Covid-19 Dibimbing oleh BUDI SUSETYO, BAGUS SARTONO, dan EFRIWATI.

Rancangan suatu percobaan idealnya harus memenuhi tiga prinsip utama. Prinsipnya adalah pengacakan, ulangan dan pengendalian lingkungan. Penerapan ketiga prinsip rancangan percobaan tersebut dilakukan untuk memenuhi asumsi penguji hipotesis agar penarikan kesimpulannya sah. Pada situasi seperti pandemic penerapan rancangan percobaan acak tidak dapat diterapkan karena kendala kode etik. Oleh karena itu, diperlukan metode alternatif untuk diterapkan pada penelitian ini.

Penelitian ini menggunakan pendekatan kuasi-eksperimen untuk mengevaluasi pengaruh konsumsi daun sungkai terhadap waktu pemulihan pasien Covid-19 Analisis skor propensitas diterapkan untuk mengurangi bias seleksi pada data observasi yang tidak teracak.

Pendekatan *quasi-experiment* menggunakan metode penelitian yang tidak melibatkan pengacakan kelompok perlakuan dan kontrol karena kendala etika atau keterbatasan sumber daya. Peneliti menentukan kelompok perlakuan dan kontrol berdasarkan data yang tersedia atau faktor lain yang tidak sepenuhnya dapat dikendalikan. Dalam penelitian ini, peneliti tidak dapat menentukan individu yang mengonsumsi daun sungkai secara acak karena keputusan konsumsi bergantung pada pilihan pribadi individu atau rekomendasi pihak lain seperti keluarga atau tenaga kesehatan.

Analisis skor propensitas digunakan untuk mengurangi bias seleksi akibat pengelompokan yang tidak acak. Skor propensitas dihitung menggunakan random forest, metode *machine learning* yang unggul dalam menangkap hubungan dan interaksi antar peubah. Data yang digunakan mencakup 723 individu di Sumatera Barat, dengan peubah meliputi umur, jenis kelamin, status vaksinasi, tingkat keparahan gejala, dan waktu pemulihan.

Bias seleksi terjadi ketika karakteristik awal kelompok perlakuan dan kontrol tidak sebanding. Sebagai contoh, individu yang mengonsumsi daun sungkai mungkin memiliki usia yang lebih muda atau status vaksinasi yang lebih lengkap dibandingkan individu yang tidak mengonsumsinya. Perbedaan ini memengaruhi validitas hasil karena baik usia maupun status vaksinasi dapat memengaruhi kecepatan pemulihan dari Covid-19. Peneliti menghitung skor propensitas menggunakan *random forest* untuk menangkap hubungan dan interaksi antar peubah. Peneliti menggunakan skor propensitas dalam metode *matching* untuk menciptakan keseimbangan karakteristik awal antar kelompok. Optimal Matching menghasilkan keseimbangan terbaik untuk memastikan hasil analisis yang valid dan terpercaya.

Pemadanan dilakukan dengan tiga metode utama, yaitu Nearest Neighbor Matching, Radius Matching, dan Optimal Matching. Evaluasi dilakukan menggunakan Standardized Mean Difference (SMD). Hasil evaluasi menunjukkan bahwa Optimal Matching merupakan metode terbaik, dengan pengurangan bias (Percent Bias Reduction/PBR) sebesar 41,68%. Nilai tersebut lebih tinggi dibandingkan Radius Matching (34,63%) dan Nearest Neighbor Matching (54,36%). Konsistensi hasil juga terlihat pada resampling sebanyak 10 kali yaitu Optimal Matching tetap menunjukkan nilai PBR terbaik.

Hasil analisis juga menunjukkan bahwa durasi konsumsi daun sungkai berhubungan signifikan dengan waktu pemulihan. Metode konsumsi seperti meminum air rebusan atau seduhan daun sungkai juga memberikan hasil yang signifikan, sementara jumlah konsumsi harian atau per seduhan tidak berpengaruh signifikan.



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

Penelitian ini menekankan pada keunggulan random forest dalam analisis skor propensitas, terutama dalam konteks data kategorik. Metode ini memberikan alternatif yang kuat untuk menggantikan regresi logistik tradisional, dengan hasil pendugaan yang lebih akurat dan valid. Selain itu, kombinasi random forest dengan optimal *matching* menunjukkan kemampuan analisis kuasi-eksperimen untuk menghasilkan kesimpulan yang valid meskipun tanpa pengacakan.

Sebagai implikasi, penelitian ini mendukung pemanfaatan daun sungkai sebagai alternatif herbal untuk mempercepat pemulihan pasien Covid-19, khususnya dalam konteks pengobatan tradisional di Indonesia. Dari perspektif statistika, penelitian ini menegaskan pentingnya pemilihan metode pemandangan yang tepat untuk mengurangi bias seleksi dan meningkatkan validitas hasil. Optimal *matching* menjadi metode yang paling efektif dalam menciptakan keseimbangan antar kelompok perlakuan dan kontrol.

Rekomendasi untuk penelitian selanjutnya meliputi peningkatan jumlah data dan peubah untuk analisis yang lebih mendalam. Eksplorasi metode machine learning lain, seperti catboost disarankan untuk meningkatkan akurasi pendugaan skor propensitas. Selain itu, penerapan teknik *matching* lain seperti dapat memperkaya hasil analisis.

Konsumsi daun sungkai memiliki pengaruh signifikan terhadap percepatan waktu pemulihan pasien Covid-19. Dengan memanfaatkan pendekatan kuasi-eksperimen, random forest, dan optimal *matching*, penelitian ini memberikan kontribusi signifikan pada metodologi statistika dan mendukung eksplorasi lebih lanjut terhadap tanaman herbal sebagai bagian dari pengobatan berbasis *science*.

Kata kunci: Covid-19, daun sungkai, kuasi-eksperimen, random forest, skor propensitas



SUMMARY

RAFIKA DAMAYANTI SURURIN NUFUS. Propensity Score Matching using Random Forest in Evaluating the Effect of Sungkai Leaf Consumption on Covid-19 Recovery. Supervised by BUDI SUSETYO, BAGUS SARTONO, and EFRIWATI.

The design of an experiment ideally should adhere to three main principles: randomization, replication, and environmental control. The application of these principles ensures that the assumptions for hypothesis testing are met, thereby making the conclusions valid. However, in situations such as a pandemic, implementing randomized experimental designs is not feasible due to ethical constraints. Therefore, alternative methods are necessary to conduct such research.

This study employed a quasi-experimental approach to evaluate the effect of sungkai leaf consumption on the recovery time of COVID-19 patients. Propensity score analysis was applied to reduce selection bias in non-randomized observational data.

The quasi-experimental approach involves research methods that do not use random assignment of treatment and control groups due to ethical constraints or resource limitations. Researchers determine the treatment and control groups based on available data or other factors that are not entirely controllable. In this study, researchers could not randomly assign individuals to consume sungkai leaves, as the decision depended on personal choice or recommendations from others, such as family or healthcare providers.

Propensity score analysis was used to minimize selection bias resulting from non-randomized groupings. Propensity scores were calculated using random forest, a machine learning method that excels at capturing relationships and interactions among variables. The data used consisted of 723 individuals in West Sumatra, with variables including age, gender, vaccination status, severity of symptoms, and recovery time.

Selection bias arises when the initial characteristics of the treatment and control groups are not comparable. For instance, individuals consuming sungkai leaves might be younger or have more complete vaccination statuses than those who do not. Such differences affect the validity of results, as both age and vaccination status can influence the recovery speed from COVID-19. Propensity scores, calculated using random forest, helped address these issues by balancing the initial characteristics of the groups. Matching methods were then applied to achieve balance, with Optimal Matching providing the best performance to ensure valid and reliable analysis results.

The matching was performed using three main methods, namely Nearest Neighbor Matching, Radius Matching, and Optimal Matching. The evaluation was conducted using the Standardized Mean Difference (SMD). The results indicated that Optimal Matching was the best method, achieving a Percent Bias Reduction (PBR) of 41.68 percent. This value was higher than the PBR achieved by Radius Matching, which was 34.63 percent, and Nearest Neighbor Matching, which was 54.36 percent. Consistent results were observed across 10 rounds of resampling, with Optimal Matching consistently delivering the highest PBR values.



The analysis also showed that the duration of sungkai leaf consumption significantly correlated with recovery time. Methods of consumption, such as drinking boiled water or brewed sungkai leaves, were also significant, while daily or per-serving consumption quantities were not.

This study highlights the advantages of random forest in propensity score analysis, particularly in the context of categorical data. This method offers a robust alternative to traditional logistic regression, providing more accurate and valid estimates. Furthermore, the combination of random forest with optimal matching demonstrates the capability of quasi-experimental analysis to yield valid conclusions, even without randomization.

As an implication, the study supports the use of sungkai leaves as an herbal alternative to accelerate the recovery of COVID-19 patients, especially in the context of traditional medicine in Indonesia. From a statistical perspective, the study underscores the importance of selecting appropriate matching methods to reduce selection bias and enhance the validity of results. Optimal matching proves to be the most effective method for creating balance between treatment and control groups.

Recommendations for future research include increasing the amount of data and variables for deeper analysis. Exploring other machine learning methods, such as CatBoost, is suggested to improve the accuracy of propensity score estimation. Additionally, applying other matching techniques could further enrich the analysis results.

Sungkai leaf consumption significantly affects the acceleration of recovery time in COVID-19 patients. By utilizing a quasi-experimental approach, random forest, and optimal matching, this study contributes significantly to statistical methodology and supports further exploration of herbal plants as part of science-based treatment.

Keywords: Covid-19, Sungkai leaves, quasi-experiment, random forest, propensity score.



©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 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.



**PEMANDANAN SKOR PROPENSITAS MENGGUNAKAN
RANDOM FOREST DALAM EVALUASI PENGARUH
KONSUMSI DAUN SUNGKAI TERHADAP KESEMBUHAN
COVID-19**

RAFIKA DAMAYANTI SURURIN NUFUS

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

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



IPB University

©Hak cipta milik IPB University

Pengaji Luar Komisi Pembimbing pada Ujian Tesis:
1 Dr. Ir. I Made Sumertajaya 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.



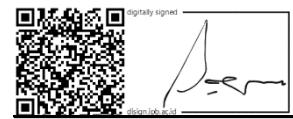


Judul Tesis : Pemadanan Skor propensitas Menggunakan Random Forest dalam Evaluasi Pengaruh Konsumsi Daun Sungkai Terhadap Kesembuhan Covid-19
Nama : Rafika Damayanti Sururin Nufus
NIM : G1501211041

Disetujui oleh



Pembimbing 1:
Dr. Budi Susetyo, M.S



Pembimbing 2:
Dr. Bagus Sartono, M.Si



Pembimbing 3:
Dr. Efriwati, M.Si

Diketahui oleh



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



Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Dr. Berry Juliandi, S.Si., M.Si
NIP. 19780723 2007011001

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.



IPB University

Bogor Indonesia



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanahu wa ta'ala atas segala karunia-Nya sehingga tesis ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini adalah "Pemadanan Skor propensitas Menggunakan Random Forest dalam Evaluasi Pengaruh Konsumsi Daun Sungkai Terhadap Kesembuhan Covid-19".

Terima kasih penulis ucapan kepada Dr. Budi Susetyo, M.S., Dr. Bagus Sartono, M.Si., dan Dr. Efriwati, M.Si., yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada moderator sidang yang telah memberikan saran untuk perbaikan. Di samping itu, penghargaan penulis sampaikan kepada BRIN yang telah membantu pengumpulan data. Ungkapan terima kasih juga disampaikan kepada Bapak, ibu, kakak, dan adik yang telah memberikan dukungan, doa, dan kasih sayangnya agar dapat menyelesaikan studi di IPB. Penulis juga mengucapkan terima kasih kepada seluruh dosen Statistika IPB dan seluruh karyawan Program Studi Statistika dan Sains Data IPB atas segala bantuan selama penulis mengikuti studi di IPB.

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

Bogor, Januari 2025

Rafika Damayanti Sururin Nufus

- 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 mengujikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

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.



IPB University

Bogor Indonesia



DAFTAR ISI

DAFTAR TABEL

x

DAFTAR GAMBAR

x

DAFTAR LAMPIRAN

x

I	PENDAHULUAN	1
	1.1 Latar Belakang	1
	1.2 Tujuan	3
II	TINJAUAN PUSTAKA	4
	2.1 Skor Propensitas	4
	2.2 Random Forest	4
	2.3 Pendugaan Skor Propensitas menggunakan <i>Random forest</i>	5
	2.4 Pemadanan berdasarkan Skor Propensitas	6
	2.5 Evaluasi Model	9
	2.6 Penghitungan Pengaruh Perlakuan	9
	2.7 <i>Percent Bias Reduction</i>	10
III	METODE	11
	3.1 Data Penelitian	11
	3.2 Tahapan Analisis	12
IV	HASIL DAN PEMBAHASAN	13
	4.1 Eksplorasi Data	13
	4.2 Hasil Pendugaan dan Uji Hipotesis	17
	4.3 Hasil Resampling menggunakan Tiga Metode pemadanan	23
	4.4 Visualisasi Menggunakan Hasil Pemadanan Terbaik	25
	4.5 Pengaruh Rentang Waktu Konsumsi terhadap Lamanya Sembuh	28
V	SIMPULAN DAN SARAN	30
	5.1 Simpulan	30
	5.2 Saran	30
	DAFTAR PUSTAKA	31
	LAMPIRAN	34
	RIWAYAT HIDUP	38

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.



Peubah yang digunakan	11
Peubah cara mengonsumsi sungkai terhadap lamanya waktu sembuh	12
Proporsi individu yang mengonsumsi ramuan sungkai	13
Proporsi padanan ketiga metode	22
Rentang umur yang terpadakan	22
<i>Percent Bias Reduction</i> berdasarkan pemanfaatan	22
Peubah bentuk konsumsi terhadap lamanya waktu sembuh	28
Analisis signifikansi <i>chi-square</i>	28
Peubah lama konsumsi yang signifikan	29

DAFTAR GAMBAR

1 Visualisasi Random Forest	5
2 Visualisasi <i>Nearest neighbor matching</i>	7
3 Visualisasi <i>Radius Matching</i>	7
4 Visualisasi <i>Optimal matching</i>	8
5 Hubungan umur terhadap konsumsi ramuan daun sungkai	13
6 Hubungan Jenis kelamin penyintas Covid-19 terhadap ramuan sungkai	14
7 Hubungan status vaksin terhadap konsumsi amuan sungkai	14
8 Hubungan terkonfirmasi Covid-19 terhadap konsumsi ramuan sungkai	15
9 Hubungan Tempat isolasi terhadap konsumsi ramuan sungkai	15
10 Waktu terkonfirmasi Covid-19 terhadap konsumsi ramuan sungkai	16
11 Gejala penyintas Covid-19 terhadap ramuan sungkai	16
12 Hasil skor propensitas menggunakan <i>Random forest</i>	17
13 Nilai SMD pemanfaatan <i>nearest neighbor</i>	18
14 Distribusi skor propensitas setelah pemanfaatan Nearest Neighbor	18
15 Nilai SMD pemanfaatan optimal	19
16 Distribusi skor propensitas setelah pemanfaatan optimal	20
17 Nilai SMD pemanfaatan radius	20
18 Distribusi skor propensitas pada <i>radius matching</i>	21
19 Nilai ATT dari metode terbaik	23
20 Visualisasi PBR optimal matching setelah resampling	24
21 Visualisasi PBR radius matching setelah resampling 10 kali	24
22 Hasil PBR Nearest Neighbor Matching	25
23 Proporsi jenis kelamin setelah dipadankan	25
24 Proporsi status vaksin setelah dipadankan	26
25 Proporsi terkonfirmasi Covid-19 setelah dipadankan	26
26 Proporsi tindakan perawatan setelah dipadankan	27
27 Proporsi gejala setelah dipadankan	27

DAFTAR LAMPIRAN

Tabel hasil perhitungan skor propensitas menggunakan random forest	35
--	----