



KLASIFIKASI DATA PUBLIKASI ILMIAH DOSEN SAINS DATA MATEMATIKA DAN INFORMATIKA IPB MENGUNAKAN SUPPORT VECTOR MACHINE

ADRI AULIA MAHRAN



**PROGRAM SARJANA 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 SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Klasifikasi Data Publikasi Ilmiah Dosen Sains Data, Matematika, dan Informatika IPB Menggunakan *Support Vector Machine*” 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, Agustus 2026

Adri Aulia Mahran
G64190050



ABSTRAK

ADRI AULIA MAHRAN. Klasifikasi Data Publikasi Ilmiah Dosen Sains Data, Matematika, dan Informatika IPB Menggunakan *Support Vector Machine*. Dibimbing oleh AHMAD RIDHA

Dengan data publikasi yang semakin banyak dan tersebar di internet, penelitian yang menggunakan *domain-specific dataset* semakin banyak ditemukan. Penelitian tersebut memiliki tujuan yang lebih sempit dengan ruang lingkup yang lebih terbatas untuk membantu menghasilkan kesimpulan yang menyesuaikan kebutuhan dan konteks yang lebih sempit. Penelitian ini meniru penelitian berikut dengan skala yang lebih sempit berupa pelatihan model klasifikasi label pada *dataset* publikasi dosen Sekolah Sains Data, Matematika, dan Informatika (SSMI) IPB. *Dataset* berisikan karya publikasi dari dosen fakultas tersebut serta data publikasi yang mensitasi karya tersebut. Penelitian berikut akan melakukan pembagian data dengan metode *holdout* yang dimodifikasi berupa *Stratified Shuffle Split* untuk mencegah tidak seimbangnnya pembagian data *training* dan *testing* dikarenakan label yang akan diklasifikasi adalah bidang kepakaran dosen pemilik karya publikasi tersebut. Hasil performa model klasifikasi akan menggunakan pendekatan *macro-average* dan direpresentasikan dalam *confusion matrix*.

Kata kunci: *Confusion Matrix, Stratified Shuffle Split, Support Vector Machine, TF-IDF.*

ABSTRACT

ADRI AULIA MAHRAN. Classification of IPB Data Science Mathematics and Informatics Lecturer Scientific publications using Support Vector Machine. Supervised by AHMAD RIDHA

With the ever-increasing volume of publication data available online, research utilizing domain-specific datasets has become more prevalent. Such research targets specific objectives within a focused scope to yield conclusions tailored to particular needs and contexts. This study replicates existing research on a more focused scale by training a label classification model using a dataset of publications by faculty members from the School of Data Science, Mathematics, and Informatics (SSMI) at IPB. The dataset comprises publications authored by these faculty members as well as data on publications citing those works. To address potential imbalances between training and testing sets, the study employs a modified holdout method known as Stratified Shuffle Split for data partitioning. Model performance is evaluated using a macro-averaging approach and visualized via a confusion matrix.

Keywords: *Confusion Matrix, Stratified Shuffle Split, Support Vector Machine, TF-IDF.*

KLASIFIKASI DATA PUBLIKASI ILMIAH DOSEN SAINS DATA MATEMATIKA DAN INFORMATIKA IPB MENGUNAKAN SUPPORT VECTOR MACHINE

ADRI AULIA MAHRAN

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
2026**

@Hak cipta milik IPB University

IPB University





@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:

- 1 Dr. Karlisa Priandana, S.T., M.Eng.
- 2 Dr. Ir. Sri Wahjuni, M.T.



@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.

Perpustakaan IPB University



Judul Skripsi : Klasifikasi Data Publikasi Ilmiah Dosen Sains Data, Matematika,
dan Informatika IPB Menggunakan *Support Vector Machine*

Nama : Adri Aulia Mahran
NIM : G64190050

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Ahmad Ridha, S.Kom., M.S.
NIP. 198005072005011001

Diketahui oleh

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

Tanggal Ujian:
16 Juli 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. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2026 sampai bulan Juli 2026 ini ialah *classification algorithm testing*, dengan judul “Klasifikasi Data Publikasi Ilmiah Dosen Sains Data, Matematika, dan Informatika IPB Menggunakan *Support Vector Machine*”.

Terima kasih penulis ucapkan kepada para pembimbing, Ahmad Ridha, S.Kom., M.S. yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada pembimbing akademik, moderator seminar, dan penguji luar komisi pembimbing. Di samping itu, penghargaan penulis sampaikan kepada Dr. Sony Hartono Wijaya, S.Kom., M.Kom., Endang Purnama Giri S.Kom., M.Kom. beserta staf administrasi Program Sarjana Ilmu Komputer IPB dan seterusnya yang telah membantu selama pelaksanaan penelitian. Ungkapan terima kasih juga disampaikan kepada ayah, ibu, serta adik dan kerabat-kerabat yang telah memberikan dukungan, doa, dan kasih sayangnya.

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

Bogor, Agustus 2026

Adri Aulia Mahran



DAFTAR ISI

DAFTAR GAMBAR	ix
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	1
1.4 Manfaat	2
1.5 Ruang Lingkup	2
TINJAUAN PUSTAKA	3
2.1 <i>Support Vector Machine</i>	3
2.2 <i>Confusion Matrix</i>	4
2.3 TF-IDF	5
2.4 <i>Stratified Shuffle Split</i>	6
III METODE	7
3.1 Data Penelitian	7
3.2 Penyiapan dan Praproses Data	7
3.3 Pemilihan Label dan <i>Feature Union</i>	9
3.4 <i>Data Splitting</i>	9
3.5 Pengembangan Model	10
3.6 Evaluasi Model	10
IV HASIL DAN PEMBAHASAN	11
4.1 Praproses Data	11
4.2 Pemberian Label dan <i>Feature Union</i>	11
4.3 Pengelompokan Data	11
4.4 Pelatihan Model	12
4.5 Hasil Evaluasi Model	12
V SIMPULAN DAN SARAN	14
5.1 Simpulan	14
5.2 Saran	14
DAFTAR PUSTAKA	15
RIWAYAT HIDUP	18



DAFTAR GAMBAR

1	Potongan kode yang pemetaan label klaster	11
2	Tabel persebaran data <i>training</i> dan <i>test</i> berdasarkan label	12
3	<i>Confusion matrix</i> dari pengujian model klasifikasi	13

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.