MODEL KLASIFIKASI INDIKATIF TINGKAT KESEHATAN KOPERASI SIMPAN PINJAM BERBASIS DATA KEUANGAN MENGGUNAKAN RANDOM FOREST

(Studi pada KSP Aktif di Kabupaten Bandung)

Disusun Oleh:
CAHYA RAHMAWATI
2C2220011

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh Gelar Sarjana Sains Data

Konsentrasi
Digital Business Analytic

PROGRAM STUDI SARJANA SAINS DATA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS KOPERASI INDONESIA
2026

MODEL KLASIFIKASI INDIKATIF TINGKAT KESEHATAN KOPERASI SIMPAN PINJAM BERBASIS DATA KEUANGAN MENGGUNAKAN RANDOM FOREST

(Studi pada KSP Aktif di Kabupaten Bandung)

Disusun Oleh:
CAHYA RAHMAWATI
2C2220011

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh Gelar Sarjana Sains Data

Konsentrasi
Digital Business Analytic

Dosen Pembimbing:

Ir. H. Dady Nurpadi, M.P., CIIQA.

Mohamad Fahreza, S.E., M.B.A.



PROGRAM STUDI SARJANA SAINS DATA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS KOPERASI INDONESIA

2026

LEMBAR PENGESAHAN

Judul : Model Klasifikasi Indikatif Tingkat Kesehatan Koperasi Simpan Pinjam Berbasis Data Keuangan Menggunakan Random Forest (Studi pada KSP Aktif di Kabupaten Bandung)

Nama : Cahya Rahmawati

NIM : 2C2220011

Fakultas : Sains dan Teknologi

Program Studi : Sarjana Sains Data

Konsentrasi : Digital Business Analytic

Menyetujui dan mengesahkan

Pembimbing I	Pembimbing II
Ir. H. Dady Nurpadi, M.P., CIIQA.	Mohamad Fahreza, S.E., M.B.A.

Pjs. Dekan Fakultas Sains dan Teknologi	Pjs. Ketua Program Studi Sains Data
Ir. H. Dady Nurpadi, M.P., CIIQA.	Ir. H. Dady Nurpadi, M.P., CIIQA.

Dipindai dengan CamScanner

PERNYATAAN MENGENAI LAPORAN SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa laporan skripsi dengan judul "Model Klasifikasi Indikatif Tingkat Kesehatan Koperasi Simpan Pinjam Berbasis Data Keuangan Menggunakan Random Forest (Studi pada KSP Aktif di Kabupaten Bandung)" 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 laporan akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Universitas Koperasi Indonesia.

Sumedang, Oktober 2026

Cahya Rahmawati
2C2220011

LEMBAR PENGESAHAN

Judul : Model Klasifikasi Indikatif Tingkat Kesehatan Koperasi Simpan Pinjam Berbasis Data Keuangan Menggunakan Random Forest (Studi pada KSP Aktif di Kabupaten Bandung)
Nama : Cahya Rahmawati
NIM : 2C2220011
Fakultas : Sains dan Teknologi
Program Studi : Sarjana Sains Data
Konsentrasi : Digital Business Analytic

Menyetujui dan mengesahkan

Pembimbing I	Pembimbing II
Ir. H. Dady Nurpadi, M.P., CIIQA.	Mohamad Fahreza, S.E., M.B.A.

Pjs. Dekan Fakultas Sains dan Teknologi	Pjs. Ketua Program Studi Sains Data
Ir. H. Dady Nurpadi, M.P., CIIQA.	Ir. H. Dady Nurpadi, M.P., CIIQA.

RIWAYAT HIDUP



Penulis dilahirkan di Sumedang pada tanggal 23 Agustus 2004 sebagai anak ke-2 dari pasangan bapak Ali dan ibu Ade. Pendidikan sekolah menengah kejuruan ditempuh di SMK Yadika Tanjungsari, Sumedang, Jawa Barat, pada jurusan Teknik Komputer dan Jaringan, dan lulus pada tahun 2022. Pada tahun 2022, penulis diterima sebagai mahasiswa program sarjana (S-1) di Program Studi Sains Data di IKOPIN University. Selama mengikuti program S-1, penulis aktif menjadi:

1. Staf Magang Departemen Sosial dan Lingkungan Badan Eksekutif Mahasiswa (BEM) IKOPIN University Periode 2022 - 2023
2. Anggota Ikopin Badminton Club (IBC) Periode 2023
3. Anggota Tim Sosialisasi Humas HIMA Sains Data IKOPIN University Periode 2023 - 2024
4. Staf Departemen Hubungan Masyarakat dan Publikasi sekaligus Kepala Tim Kreatif HIMA Sains Data IKOPIN University Periode 2023 - 2025
5. Penerima Beasiswa BAZNAS Tahun 2024

ABSTRACT

CAHYA RAHMAWATI, 2026. *Indicative Health Classification Model for Savings and Loan Cooperatives Based on Financial Data Using Random Forest (A Study on Active Savings and Loan Cooperatives in Bandung Regency), under the guidance of H. Dady Nurpadi and Mohamad Fahreza.*

Savings and Loan Cooperatives (KSP) need to be monitored regularly, but financial reporting and health inspections do not yet cover all active KSPs. Between 2022 and 2025, the proportion of active KSPs in Bandung Regency submitting financial reports to the West Java Provincial Cooperative Office decreased from 58.06% to 51.15%. The available data are also incomplete for a comprehensive health assessment. This condition indicates the need for an approach that uses available financial data to support initial screening of KSP health conditions. This study aims to develop and evaluate an indicative health-status classification model for KSPs using Random Forest and to compare its performance with Decision Tree.

The unit of analysis is the KSP-year (cooperative-year), where one observation represents one KSP in one reporting year. The data consist of 287 eligible KSP-year observations from 2022-2025. The analysis used a group-aware train-test split, StratifiedGroupKFold cross-validation, Leakage-Controlledd testing, and a temporal holdout using 2025 data as the test period.

Random Forest achieved an accuracy of 0.6719, a balanced accuracy of 0.4629, and a Macro F1-Score of 0.4928, and performed better than Decision Tree on the study data. These scores show that the model's ability to distinguish health categories was uneven. In the temporal holdout, Random Forest achieved a balanced accuracy of 0.6320. These results indicate that Random Forest can be used as an initial screening tool, but the classification remains indicative and requires verification through the official cooperative health assessment mechanism.

Keywords: Savings and Loan Cooperative, indicative health-status classification, surrogate label, Random Forest, financial data, group-aware validation, temporal holdout

ABSTRAK

CAHYA RAHMAWATI, 2026. Model Klasifikasi Indikatif Tingkat Kesehatan Koperasi Simpan Pinjam Berbasis Data Keuangan Menggunakan *Random Forest* (Studi pada KSP Aktif di Kabupaten Bandung), di bawah bimbingan H. Dady Nurpadi dan Mohamad Fahreza.

Koperasi Simpan Pinjam (KSP) perlu dipantau secara berkala, tetapi pelaporan keuangan dan pemeriksaan kesehatan belum mencakup seluruh KSP aktif. Pada periode 2022-2025, proporsi KSP aktif di Kabupaten Bandung yang menyampaikan laporan keuangan kepada Dinas Koperasi dan Usaha Kecil Provinsi Jawa Barat menurun dari 58,06% menjadi 51,15%. Data yang tersedia juga belum lengkap untuk pemeriksaan kesehatan secara menyeluruh. Kondisi tersebut menunjukkan perlunya pendekatan yang memanfaatkan data keuangan yang tersedia untuk membantu *screening* awal kondisi kesehatan KSP. Penelitian ini bertujuan membangun dan mengevaluasi model klasifikasi tingkat kesehatan indikatif KSP menggunakan *Random Forest* serta membandingkan kinerjanya dengan *Decision Tree*.

Unit analisis penelitian adalah KSP-tahun (*cooperative-year*), yaitu satu observasi yang mewakili satu KSP pada satu tahun pelaporan. Data terdiri atas 287 observasi KSP-tahun periode 2022-2025 yang memenuhi kriteria kelayakan data. Analisis dilakukan menggunakan *group-aware train-test split, StratifiedGroupKFold cross-validation*, pengujian *Leakage-Controlledd*, dan *temporal holdout* dengan data tahun 2025 sebagai data uji.

Random Forest memperoleh *accuracy* sebesar 0,6719, *balanced accuracy* sebesar 0,4629, dan *Macro F1-Score* sebesar 0,4928, serta memiliki kinerja lebih baik dibandingkan *Decision Tree* pada data penelitian. Nilai tersebut menunjukkan bahwa kemampuan model dalam mengenali kategori kesehatan belum merata. Pada *temporal holdout*, *Random Forest* memperoleh *balanced accuracy* sebesar 0,6320. Hasil tersebut menunjukkan bahwa *Random Forest* dapat digunakan sebagai alat bantu *screening* awal, tetapi hasil klasifikasi bersifat indikatif dan tetap memerlukan verifikasi melalui mekanisme penilaian kesehatan koperasi secara resmi.

Kata Kunci: Koperasi Simpan Pinjam, klasifikasi tingkat kesehatan indikatif, *surrogate label, Random Forest,* data keuangan, *group-aware validation, temporal holdout*

KATA PENGANTAR

Puji syukur ke hadirat Allah Swt. atas segala rahmat, karunia, dan kemudahan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Model Klasifikasi Indikatif Tingkat Kesehatan Koperasi Simpan Pinjam Berbasis Data Keuangan Menggunakan Random Forest (Studi pada KSP Aktif di Kabupaten Bandung)" sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Data pada Program Studi S1 Sains Data, Fakultas Sains dan Teknologi, Universitas Koperasi Indonesia.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa perjalanan yang dilalui tidak selalu mudah. Oleh karena itu, penulis ingin menyampaikan terima kasih dan penghargaan yang tulus kepada semua pihak yang telah memberikan doa, dukungan, bantuan, serta semangat hingga skripsi ini dapat diselesaikan.

1. Kedua orang tua penulis, terutama Mamah Ade, yang telah membiayai dan mendukung penulis sejak lahir hingga saat ini. Terima kasih atas segala pengorbanan, kasih sayang, doa, dan perjuangan yang tidak pernah berhenti. Mamah adalah sosok ibu yang kuat dan selalu berusaha memberikan yang terbaik bagi penulis. You are a strong mother, and I am forever grateful to be your child.
2. Adik-adik penulis, Mega Aulia Risti dan M. Al Farizi. Semoga penulis dapat menjadi kakak yang membanggakan bagi kalian.
3. Dosen pembimbing penulis, Bapak Ir. H. Dady Nurpadi, M.P., CIIQA. selaku Dosen Pembimbing I dan Bapak Mohamad Fahreza, S.E., M.B.A. selaku Dosen Pembimbing II, terima kasih atas waktu, arahan, ilmu, dan kesabaran dalam membimbing penulis selama proses penyusunan skripsi ini.
4. Seluruh dosen Program Studi Sains Data dan staf sekretariat Fakultas Sains dan Teknologi Universitas Koperasi Indonesia, terima kasih atas ilmu, bantuan, dan pelayanan yang telah diberikan selama penulis menempuh pendidikan hingga menyelesaikan studi ini.

5. Seluruh teman sekelas Sains Data dan teman-teman Fakultas Sains dan Teknologi, terima kasih atas kebersamaan, pengalaman, dan kenangan yang telah mewarnai perjalanan penulis selama delapan semester. Semoga segala perjuangan yang telah dilalui membawa kita menuju keberhasilan masing-masing.

6. Kepada Rizki Fajar Ramadan, terima kasih telah bersedia menemani dan mengantar penulis mencari koperasi untuk keperluan penelitian, serta selalu berusaha hadir ketika penulis membutuhkan teman untuk bercerita. Terima kasih telah mendengarkan segala keluh kesah, rasa lelah, kekhawatiran, serta cerita yang datang silih berganti selama semester akhir ini. Bagi penulis, kehadiranmu bukan hanya sebagai pasangan, tetapi juga sebagai teman berbagi sedih, amarah, canda, dan tawa. Terima kasih atas setiap waktu, tenaga, perhatian, dan usaha yang telah diberikan. Penulis menyadari bahwa perjalanan kita tidak selalu berjalan mulus. Oleh karena itu, penulis memohon maaf atas segala kekurangan yang ada. Penulis berharap kita dapat terus belajar, bertumbuh, dan berproses bersama menjadi pribadi yang lebih baik. Terima kasih telah hadir dalam kehidupan penulis dan menjadi bagian dari perjalanan ini. Semoga doa-doa baik senantiasa menyertaimu.

7. Pada akhirnya, penulis menyadari bahwa skripsi yang baik adalah skripsi yang selesai. Terima kasih telah bertahan hingga sampai di titik ini.

Last but not least, I wanna thank me. I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for never quitting, I wanna thank me for always being a giver more than I receive, I wanna thank me for trying to do more right than wrong, I wanna thank me for just being me at all times.

Sumedang, Oktober 2026

Penulis

Cahya Rahmawati

DAFTAR ISI

RIWAYAT HIDUP ..... **i**

ABSTRACT ..... **ii**

ABSTRAK ..... **iii**

KATA PENGANTAR ..... **iv**

DAFTAR ISI ..... **vi**

DAFTAR TABEL ..... **x**

DAFTAR GAMBAR ..... **xii**

DAFTAR LAMPIRAN ..... **xiii**

BAB I PENDAHULUAN **1**

1.1 Latar Belakang Penelitian **1**

1.2 Identifikasi Masalah **8**

1.3 Maksud dan Tujuan Penelitian **9**

1.3.1 Maksud Penelitian 9

1.3.2 Tujuan Penelitian 9

1.4 Kegunaan Penelitian **10**

1.4.1 Kegunaan Teoretis 10

1.4.2 Kegunaan Praktis 11

BAB II PENDEKATAN MASALAH DAN METODE PENELITIAN **12**

2.1 Pendekatan Masalah **12**

2.1.1 Konsep Koperasi 12

2.1.2 Koperasi Simpan Pinjam (KSP) 13

2.1.3 Konsep dan Kriteria Kesehatan Koperasi 14

2.1.4 Jati Diri Koperasi 15

2.1.5 Evolusi Regulasi Penilaian/Pemeriksaan Kesehatan Koperasi 16

2.1.6 Sains Data dan *Machine Learning* 18

2.1.7 Kerangka Pemikiran 20

2.1.8 Kerangka Operasional Penelitian 22

2.2 Metode Penelitian 24

2.2.1 Metode Penelitian yang Digunakan 24

2.2.2 Data yang Diperlukan (Operasionalisasi Variabel) 25

2.2.3 Sumber Data dan Cara Menentukan Sumber Data 28

2.2.4 Teknik Pengumpulan Data 29

2.2.5 Analisis Data 30

2.2.6 Tempat/Lokasi Penelitian 35

2.2.7 Jadwal Penelitian 36

BAB III KEADAAN UMUM TEMPAT PENELITIAN 37

3.1 Keadaan Umum Organisasi 37

3.1.1 Profil Dinas Koperasi dan Usaha Kecil Provinsi Jawa Barat 37

3.1.2 Gambaran Umum KSP Aktif di Kabupaten Bandung 38

3.2 Keanggotaan dan Kelembagaan KSP 40

3.3 Kegiatan Usaha dan Kondisi Keuangan KSP 40

3.4 Implementasi Jati Diri Koperasi 42

3.4.1 Implementasi Definisi Koperasi..... 43

3.4.2 Implementasi Nilai-Nilai Koperasi 44

3.4.3 Implementasi Prinsip-Prinsip Koperasi 45

BAB IV HASIL DAN PEMBAHASAN 48

4.1 Deskripsi Data Penelitian 48

4.1.1 Sumber Data..... 48

4.1.2 Hasil Preprocessing..... 48

4.1.3 Distribusi Label..... 50

4.2 Pembentukan Label Kesehatan Indikatif 52

4.2.1 Dasar Pembentukan Label 52

4.2.2 Distribusi Label..... 53

4.3 Hasil Pemodelan dan Evaluasi..... 53

4.3.1 Pembagian Data Latih dan Data Uji (*Group-aware Train-Test Split)*. 53

4.3.2 *Majority Baseline* (Pembanding Paling Sederhana) 54

4.3.3 Random Forest Full Feature..... 55

4.3.4 Random Forest Leakage-Controlledd 59

4.3.5 Decision Tree 62

4.3.6 Perbandingan Model 64

4.4 Analisis *Feature Importance* (Variabel Paling Berpengaruh)..... 65

4.4.1 Feature Importance Random Forest 65

4.4.2 *Permutation Importance* (Pengecekan Silang) 67

4.4.3 Korelasi Antarvariabel 69

4.5 Validasi Temporal Holdout 70

4.6 Pembahasan 73

4.6.1 Kemampuan Random Forest dalam Mengklasifikasikan Kesehatan KSP 73

4.6.2 Keterbatasan Label Kesehatan yang Digunakan 73

4.6.3 Variabel Keuangan yang Paling Banyak Digunakan Model 74

4.6.4 Perbandingan *Random Forest* dengan *Decision Tree* 75

4.6.5 Kemampuan Model pada Data Tahun 2025 75

4.6.6 Implikasi Model Sebagai Alat Bantu Screening KSP 76

BAB V SIMPULAN DAN SARAN 79

5.1 Simpulan 79

5.2 Saran 81

DAFTAR PUSTAKA 83

LAMPIRAN 87

DAFTAR TABEL

Tabel 1.1 Jumlah KSP Aktif dan KSP yang Menyampaikan Laporan Keuangan di Kabupaten Bandung Periode 2022-2025 2

Tabel 2.1 Operasionalisasi Variabel Penelitian 26

Tabel 2.2 Ringkasan Penyaringan Observasi 30

Tabel 2.3 Tahapan Pembentukan Label Kesehatan Indikatif 31

Tabel 2.4 Pemetaan Kode Label (LabelEncoder) 32

Tabel 2.5 Distribusi Kategori pada Data Latih dan Data Uji 33

Tabel 2.6 Jadwal Penelitian 36

Tabel 3.1 Tren Status Pelaporan KSP Aktif di Kabupaten Bandung Tahun 2022-2025 39

Tabel 3.2 Statistik Deskriptif Variabel Keuangan per Tahun (2022-2025) 41

Tabel 4.1 Ringkasan Hasil Preprocessing Data 49

Tabel 4.2 Distribusi Label Kesehatan Indikatif 50

Tabel 4.3 Ringkasan Group-aware Train-Test Split 53

Tabel 4.4 Hasil Evaluasi Majority Baseline 54

Tabel 4.5 Hasil Evaluasi Random Forest Full Feature 56

Tabel 4.6 Classification Report Random Forest Full Feature 57

Tabel 4.7 Hasil Evaluasi Random Forest Leakage-Controlledd 59

Tabel 4.8 Classification Report Random Forest Leakage-Controlledd 60

Tabel 4.9 Hasil Evaluasi Decision Tree 62

Tabel 4.10 Classification Report Decision Tree 62

Tabel 4.11 Perbandingan Model Random Forest dan Decision Tree 64

Tabel 4.12 Feature Importance Random Forest Full Feature (MDI) 66

Tabel 4.13 Permutation Importance Random Forest Full Feature 68

Tabel 4.14 Ringkasan Data Temporal Holdout 70
Tabel 4.15 Hasil Evaluasi Temporal Holdout..... 71
Tabel 4.16 Classification Report Temporal Holdout..... 72

DAFTAR GAMBAR

Gambar 2.1 Kerangka Pemikiran Penelitian 21
Gambar 2.2 Kerangka Operasional Penelitian 23
Gambar 4.1 Distribusi Label Kesehatan Indikatif 50
Gambar 4.2 Confusion Matrix Majority Baseline 55
Gambar 4.3 Confusion Matrix Random Forest Full Feature 58
Gambar 4.4 Confusion Matrix Random Forest Leakage-Controlledd 61
Gambar 4.5 Perbandingan Random Forest Full Feature dan Leakage-Controlledd 61
Gambar 4.6 Confusion Matrix Decision Tree 63
Gambar 4.7 Perbandingan Model Random Forest dan Decision Tree 64
Gambar 4.8 Feature Importance Random Forest Full Feature (MDI) 65
Gambar 4.9 Feature Importance Random Forest Leakage-Controlledd (MDI) 66
Gambar 4.10 Permutation Importance Random Forest Full Feature 67
Gambar 4.11 Permutation Importance Random Forest Leakage-Controlled 68
Gambar 4.12 Korelasi Antar Variabel 69
Gambar 4.13 Confusion Matrix Temporal Holdout 72

DAFTAR LAMPIRAN

Lampiran 1 Hasil Wawancara 87

Lampiran 2 Hasil Klarifikasi Lanjutan dengan Dinas..... 89

Lampiran 3 Syntax Program Pengolahan dan Pemodelan Data..... 90