

CPMK 07.1 · SEMESTER 5 & 7

Studi Kasus Supervised & Semi-supervised

PERTEMUAN

05

IF261220 3 SKS · Program Studi Informatika, Fakultas Ilmu Komputer

SEBELUM MULAI

Recap Pertemuan 3

Kelebihan & Kekurangan

Tiap paradigma punya trade-off sendiri antara kebutuhan data, biaya, dan evaluasi.

Studi Kasus Unsupervised

Clustering, deteksi anomali, dan asosiasi — tiga cara data tanpa label dimanfaatkan.

AGENDA PERTEMUAN 5

01

Recap singkat supervised & semi-supervised

02

Decision Tree & Self-training + latihan kelas

03

Empat kategori & tiga studi kasus mendalam

04

Ringkasan kategori & algoritma

05

Pengantar Latihan Soal (bobot 10%)

06

Rubrik, linimasa, dan cara pengumpulan

Mengingat Kembali Dua Paradigma



Supervised Learning

Belajar dari data berlabel untuk memprediksi label baru.

- + Akurasi tinggi, evaluasi objektif
- Butuh label yang mahal dan berkualitas



Semi-supervised Learning

Menggabungkan sedikit data berlabel dengan banyak data tanpa label.

- + Mengurangi kebutuhan pelabelan
- Performa bergantung pada asumsi struktur data

Hari ini, pertanyaannya bergeser dari “kapan dipakai?” menjadi “seperti apa penerapannya di dunia nyata?”

Detail kelebihan dan kekurangan lengkap kedua paradigma dibahas tuntas pada Pertemuan 3.

Empat Kategori Tugas Supervised Learning

01

Klasifikasi

Memprediksi kategori atau kelas dari data baru.

Decision Tree, Naive Bayes, Neural Network

02

Regresi

Memprediksi nilai numerik yang bersifat kontinu.

Linear Regression, Random Forest Regressor

03

Deteksi & Rekognisi

Mengenali objek, wajah, atau ucapan dari data mentah.

CNN, YOLO, Speech Recognition Models

04

Peramalan Deret Waktu

Memprediksi nilai masa depan dari data berurutan waktu.

ARIMA, LSTM, Prophet

Semi-supervised learning? Bisa diterapkan pada keempat kategori ini, kapan pun label yang tersedia terbatas.

Decision Tree

Dataset klasik: memprediksi apakah tim jadi bermain futsal berdasarkan empat kondisi cuaca hari itu.

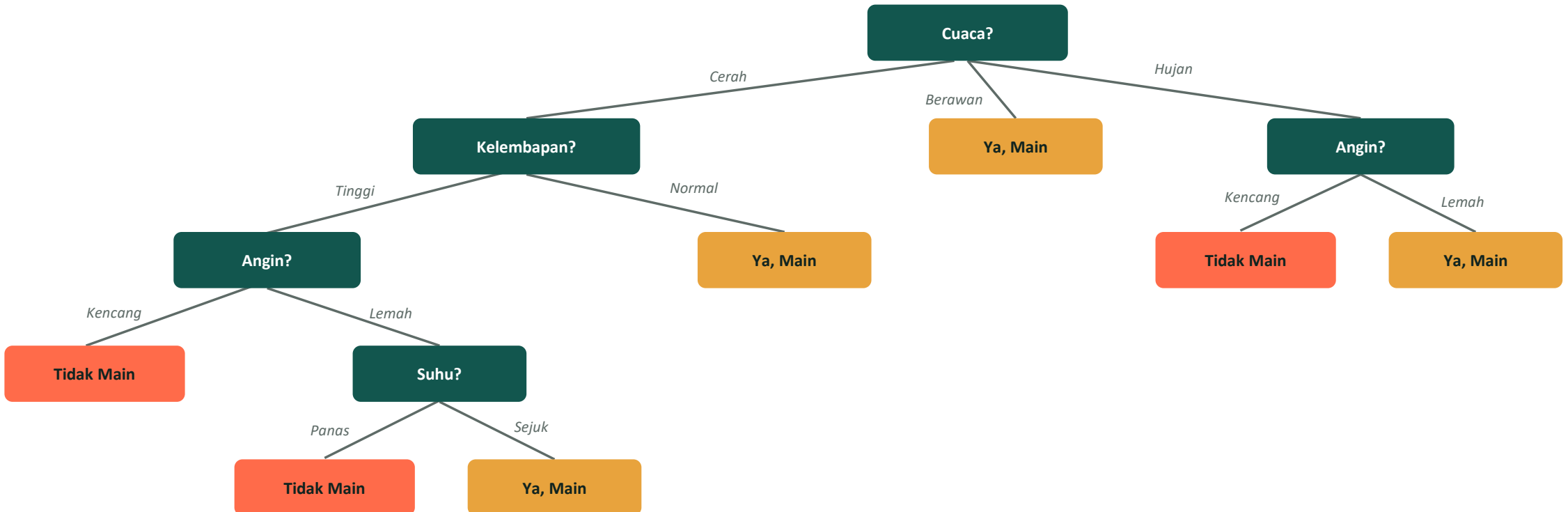
SEBAGIAN DATA PELATIHAN

Hari	Cuaca	Kelembapan	Angin	Suhu	Main?
1	Cerah	Tinggi	Lemah	Panas	Tidak
2	Cerah	Tinggi	Kencang	Panas	Tidak
3	Berawan	Tinggi	Lemah	Panas	Ya
4	Hujan	Normal	Lemah	Sejuk	Ya
5	Hujan	Normal	Kencang	Sejuk	Tidak
6	Cerah	Normal	Lemah	Sejuk	Ya
7	Cerah	Tinggi	Lemah	Sejuk	Ya

Catatan: algoritma memilih fitur akar (mis. Cuaca) berdasarkan seberapa besar fitur itu mengurangi ketidakpastian label. Proses ini diulang pada tiap cabang hingga datanya seragam, yang menghasilkan pohon dengan kedalaman berbeda-beda di tiap sisi.

Pohon Keputusan: Main Futsal atau Tidak?

Kedalaman hingga 5 tingkat dengan tiap cabang berhenti ketika datanya sudah seragam.



● Ya, Main ● Tidak Main ■ Simpul Keputusan (fitur)

Cabang terpanjang (Cerah → Tinggi → Lemah → Suhu) mencapai 5 tingkat, sementara cabang Berawan sudah berhenti di tingkat ke-2.

Self-training

Studi kasus: mengklasifikasikan topik artikel mahasiswa dengan hanya dua contoh berlabel.

Perhatikan bagaimana model sederhana bisa "mengajari dirinya sendiri" dari data tanpa label, lewat studi kasus klasifikasi topik artikel berikut.

DATA BERLABEL (AWAL)

Judul Artikel	Kata "Teknologi"	Kata "Olahraga"	Label
AI Ciptakan Robot Baru	3	0	Teknologi
Timnas Menang Liga	0	3	Olahraga

DATA TANPA LABEL — HASIL PREDIKSI ITERASI 1

Judul Artikel	Teknologi	Olahraga	Prediksi	Yakin?
Startup Kembangkan Aplikasi AI	2	0	Teknologi	Tinggi
Atlet Sabet Medali Emas	0	2	Olahraga	Tinggi
Robot Pemain Bola Ciptaan Mhs.	2	1	Teknologi	Rendah

Catatan: dua baris dengan keyakinan "Tinggi" digabungkan ke data latih untuk iterasi berikutnya; baris dengan skor tipis (2 vs 1) disimpan dulu sampai model lebih kuat.

Latihan Papan Tulis: Coba Sendiri!

Kerjakan iterasi pertama self-training untuk dataset baru ini bersama-sama di kelas.

DATA BERLABEL (AWAL)

Aduan Mahasiswa	Kata "Akademik"	Kata "Fasilitas"	Label
Nilai UAS Belum Keluar	2	0	Akademik
AC Kelas Mati Total	0	2	Fasilitas

DATA TANPA LABEL — ISI BAGIAN YANG KOSONG

Aduan Mahasiswa	Akademik	Fasilitas	Prediksi (Iterasi 1)	Yakin?
Dosen Telat Masuk Kelas	2	0	...	...
Wifi Perpustakaan Lambat	0	2	...	...
Ujian Susulan Belum Ada Jadwal	2	0	...	...
Parkir Motor Penuh Setiap Pagi	0	1	...	...
Dosen Minta Ganti Ruang, AC Rusak	1	1	...	...

Langkah kerja: (1) tentukan aturan dari data berlabel, (2) prediksi & tandai keyakinan kelima baris di atas, (3) sepakati mana yang digabung ke iterasi berikutnya dan diskusikan baris dengan skor imbang (1–1).

Diagnosis Penyakit

Model belajar memisahkan kelas 'sehat' dan 'sakit' dari data gejala berlabel.

Rumah sakit ingin membantu dokter menyaring pasien berisiko tinggi dari gejala dan hasil laboratorium yang sudah terkumpul.

PENDEKATAN

Data

Gejala, hasil lab, dan riwayat pasien berlabel sehat/sakit.

Pelatihan

Decision Tree atau Neural Network dilatih pada data berlabel.

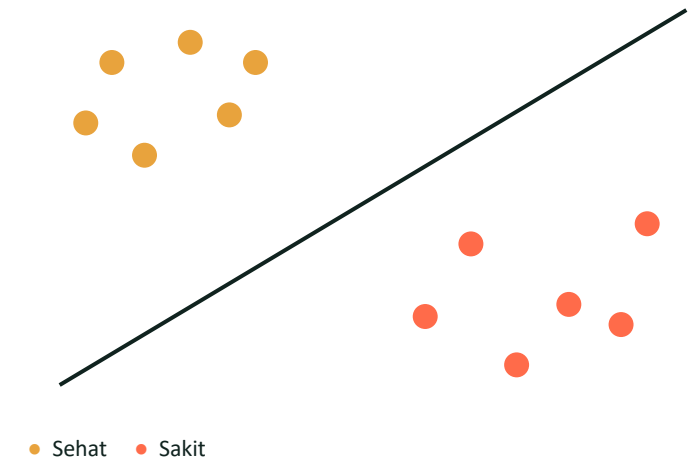
Batas Keputusan

Model membentuk batas pemisah antar-kelas.

Prediksi

Pasien baru diklasifikasikan berdasarkan batas tersebut.

BATAS KEPUTUSAN



Hasil: model membantu menyaring pasien berisiko tinggi untuk diprioritaskan pemeriksaan lanjutan oleh dokter.

Prediksi Harga Rumah

Model belajar hubungan antara luas rumah dan harga jualnya.

Agen properti ingin memperkirakan harga jual rumah baru berdasarkan luas bangunan dan riwayat transaksi rumah lain.

PENDEKATAN

Data

Luas bangunan dan harga jual dari transaksi sebelumnya.

Pelatihan

Linear Regression mencari garis yang paling pas dengan data.

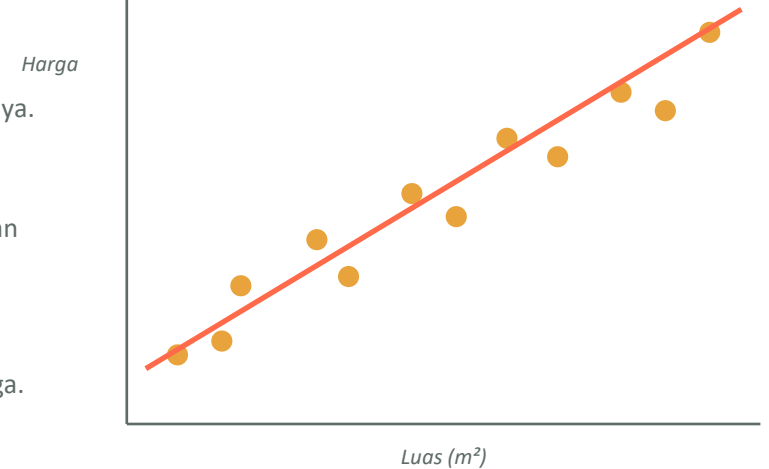
Garis Prediksi

Garis merepresentasikan tren hubungan luas dan harga.

Prediksi

Harga rumah baru diestimasi dari titik pada garis tersebut.

GARIS REGRESI



Hasil: estimasi harga membantu agen menyusun penawaran awal yang realistis sebelum negosiasi.

Klasifikasi Sentimen Ulasan

Sedikit ulasan berlabel memandu klasifikasi ribuan ulasan tanpa label.

Platform e-commerce ingin mengetahui sentimen ribuan ulasan produk, padahal hanya sebagian kecil yang sempat diberi label oleh tim.

PENDEKATAN

Data Awal

Sebagian kecil ulasan diberi label positif/negatif.

Pelatihan Awal

Model dilatih dari data berlabel yang tersedia.

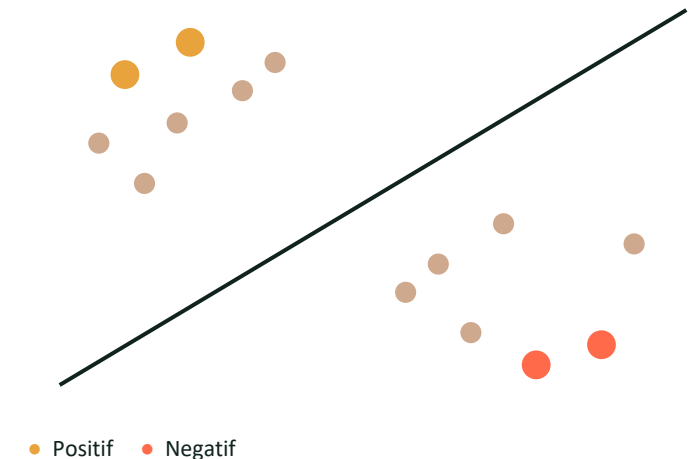
Self-training

Model memberi label perkiraan pada data tanpa label.

Pelatihan Ulang

Model dilatih ulang dengan data yang telah diperluas.

LABEL MENYEBAR



Hasil: seluruh ulasan mendapat estimasi sentimen tanpa perlu melabeli ribuan ulasan secara manual.

Kategori, Algoritma, dan Contoh Kasus

Kategori	Algoritma Umum	Contoh Kasus
Klasifikasi	Decision Tree, Naive Bayes, Neural Network	Diagnosis penyakit, deteksi spam
Regresi	Linear Regression, Random Forest Regressor	Prediksi harga rumah, prakiraan penjualan
Deteksi & Rekognisi	CNN, YOLO, Speech Recognition	Pengenalan wajah, transkripsi suara
Peramalan Deret Waktu	ARIMA, LSTM, Prophet	Prediksi permintaan, prakiraan cuaca
Semi-supervised	Self-training, Co-training, Label Propagation	Klasifikasi sentimen, kategorisasi dokumen

BAGIAN 02 • BOBOT 10%

Latihan Soal Minggu 5

Membuat program Python sederhana dan mereview satu jurnal terkait topik final project Anda.

02

Dua Komponen yang Dikumpulkan

01

Program Python Sederhana

Menerapkan salah satu algoritma yang telah dibahas pada dataset pilihan Anda sendiri.

02

Review Jurnal

Meringkas satu artikel machine learning yang relevan dengan topik final project Anda.

Keduanya dikumpulkan bersamaan melalui ilmu2.upnjatim.ac.id, dan bersama-sama membentuk 10% nilai akhir.

KOMPONEN 1

Program Python Sederhana

Menunjukkan bahwa Anda memahami alur kerja ML, bukan sekadar memanggil pustaka.

KETENTUAN

Pilih Satu Algoritma

Boleh dari supervised, unsupervised, atau semi-supervised yang sudah dibahas.

Pilih Dataset Sendiri

Dataset publik sederhana, atau data yang relevan dengan topik final project.

Ikuti Alur Kerja ML

Muat data → pra-pemrosesan → latih model → evaluasi hasil.

Sertakan Interpretasi

Jelaskan singkat apa arti hasil model dalam 3–5 kalimat.

FORMAT PENGUMPULAN

Berkas .py atau notebook .ipynb yang dapat dijalankan langsung, disertai komentar singkat pada tiap tahapan alur kerja.

Review Jurnal

Melatih kemampuan membaca kritis sebagai dasar penting untuk artikel ilmiah final project.

HAL YANG PERLU DITULISKAN

Masalah

Persoalan apa yang coba diselesaikan penulis jurnal?

Data & Metode

Dataset apa yang dipakai, dan algoritma atau pendekatan apa yang digunakan?

Hasil Utama

Bagaimana performa yang dilaporkan, dan metrik apa yang dipakai?

Relevansi

Mengapa jurnal ini relevan dengan topik final project Anda?

SUMBER JURNAL

Pilih dari basis data terindeks seperti Scopus, SINTA, atau IEEE Xplore. Panjang review sekitar setengah hingga satu halaman.

Rubrik Penilaian

Komponen	Kriteria	Bobot
Program berjalan tanpa error	Kode dapat dieksekusi ulang dari awal hingga akhir	30%
Alur kerja ML lengkap	Mencakup pra-pemrosesan, pelatihan, dan evaluasi	30%
Interpretasi hasil	Penjelasan singkat bermakna dan sesuai hasil model	20%
Kualitas review jurnal	Merangkum masalah, metode, hasil, dan relevansi dengan tepat	20%

Linimasa dan Cara Pengumpulan

01

Hari ini

Pilih algoritma, dataset, dan jurnal yang akan direview.

02

Sepanjang minggu

Kerjakan program Python dan tulis review jurnal secara paralel.

03

Sebelum deadline

Unggah kedua berkas ke ilmu2.upnjatim.ac.id dalam satu pengumpulan.

04

Setelah dinilai

Umpan balik dapat dipakai untuk menajamkan topik final project.

PENGUMPULAN

Unggah berkas `.py/.ipynb` dan review jurnal (`.pdf` atau `.docx`) sebagai satu paket ke ilmu2.upnjatim.ac.id sebelum deadline yang diumumkan di kelas.

Rangkuman dan Persiapan

SUDAH KITA PAHAM

- Decision Tree & Self-training
- Empat kategori tugas supervised learning
- Tiga studi kasus: klasifikasi, regresi, semi-supervised

LATIHAN SOAL

- Program Python + review jurnal, dikumpulkan bersamaan
- Empat kriteria penilaian dengan bobot merata
- Dikumpulkan lewat ilmu2.upnjatim.ac.id

PERSIAPAN PERTEMUAN 06

- Materi: topik dan jenis-jenis algoritma unsupervised learning
- Mulai praktik langsung membangun model dengan Python
- Siapkan daftar anggota kelompok untuk Final Project (maks. 5 orang)