

Rangkuman Materi Perkuliahan Machine Learning Kelompok 9

Denvi Sekartazi Iskandar¹, Muhammad Ma'rufil Kurhi²,

^{1,2} Program Studi Teknik Informatika, STMIK Tazkia

³ E-mail :

241552010020.denvi@student.stmik.tazkia.ac.id¹

241552010002.ma'rufil@student.stmik.tazkia.ac.id²

Abstrak

Machine learning merupakan salah satu cabang dari Artificial Intelligence yang berfokus pada pengembangan sistem yang mampu belajar dari data dan meningkatkan performanya tanpa harus diprogram secara eksplisit. Materi dalam rangkuman ini membahas konsep machine learning secara menyeluruh, mulai dari pengantar dan landasan teoritis, implementasi di dunia nyata, hingga penerapan algoritma dan workflow pada machine learning dan deep learning. Pembahasan mencakup paradigma filosofis pembelajaran mesin, peran data dalam pengambilan keputusan, serta aspek matematis dan teknis dari algoritma seperti Decision Tree (ID3) dan Gaussian Mixture Model (GMM).

Selain itu, rangkuman ini juga menekankan pentingnya kualitas dan etika penggunaan data, validasi dataset, serta evaluasi performa model untuk menghasilkan sistem yang andal dan dapat dipercaya. Materi workflow machine learning dan deep learning disajikan untuk memberikan gambaran alur kerja end-to-end, mulai dari perumusan masalah, pengolahan data, pelatihan model, hingga deployment dan monitoring. Dengan pendekatan ini, diharapkan mahasiswa dapat memahami machine learning tidak hanya sebagai kumpulan algoritma, tetapi sebagai sebuah proses terstruktur yang mengintegrasikan teori, praktik, dan kebutuhan dunia nyata.

Kata kunci: Machine Learning, Artificial Intelligence, Data, Decision Tree, Gaussian Mixture Model, Deep Learning

**Titles is Bold with Capital Letters at The Beginning of Each Letter
Except for Conjunctions
(Maximum 3 Line Title: 14pt Bolt Times New Roman)**

Abstract

Machine learning is a branch of Artificial Intelligence that focuses on developing systems capable of learning from data and improving their performance without being explicitly programmed. This summary discusses machine learning concepts in a comprehensive manner, starting from introductory topics and theoretical foundations, real-world implementations, to the application of algorithms and workflows in machine learning and deep learning. The discussion covers the philosophical paradigms of machine

learning, the role of data in decision-making, as well as the mathematical and technical aspects of algorithms such as Decision Tree (ID3) and Gaussian Mixture Model (GMM).

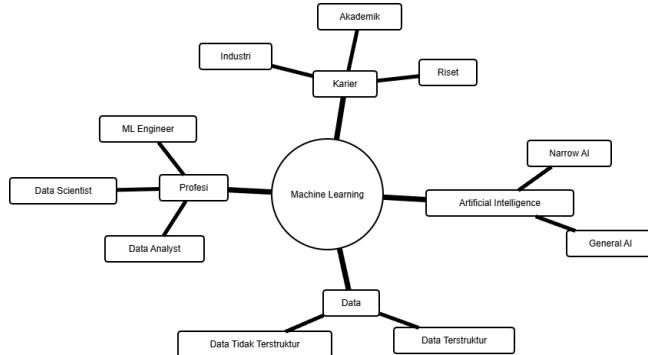
In addition, this summary emphasizes the importance of data quality and ethical data usage, dataset validation, and model performance evaluation to build reliable and trustworthy systems. The machine learning and deep learning workflow is presented to provide an end-to-end overview, starting from problem formulation, data processing, model training, evaluation, to deployment and monitoring. Through this approach, students are expected to understand machine learning not merely as a collection of algorithms, but as a structured process that integrates theory, practice, and real-world requirements.

Keywords: Machine Learning, Artificial Intelligence, Data, Decision Tree, Gaussian Mixture Model, Deep Learning

BAB 1 Pengantar Machine Learning dan Prospek karir

1. Deskripsi

Bab ini berfungsi sebagai landasan konseptual bagi mahasiswa untuk memahami posisi machine learning dalam ekosistem Artificial Intelligence (AI) serta relevansinya terhadap perkembangan teknologi dan dunia kerja. Machine learning tidak hanya dipahami sebagai teknik pemrograman, tetapi sebagai paradigma baru dalam membangun sistem cerdas yang mampu belajar dari data. Selain itu, bab ini juga memberikan gambaran prospek karir dan peran profesional yang muncul akibat pesatnya adopsi teknologi berbasis data dan AI.



1.1 Latar Belakang Machine Learning

Perkembangan machine learning didorong oleh meningkatnya ketersediaan data digital, kemajuan perangkat keras, serta kebutuhan sistem yang mampu beradaptasi secara otomatis. Sistem komputasi tradisional berbasis aturan (rule-based systems) memiliki keterbatasan ketika menghadapi data besar dan kompleks. Machine learning hadir sebagai solusi dengan memungkinkan sistem belajar langsung dari data historis tanpa harus diprogram secara eksplisit untuk setiap skenario.

1.2 Ruang Lingkup Artificial Intelligence

Artificial Intelligence merupakan bidang ilmu komputer yang berfokus pada penciptaan sistem yang mampu meniru kecerdasan manusia. Machine learning merupakan subbidang AI yang menekankan pembelajaran berbasis data, sedangkan deep learning merupakan bagian lebih spesifik yang menggunakan jaringan saraf berlapis. Hubungan ini penting untuk dipahami agar tidak terjadi kesalahan konseptual dalam penggunaan istilah AI di dunia akademik maupun industri.

1.3 Profesi dan Peran di Bidang Data dan AI

Munculnya machine learning melahirkan berbagai peran profesional baru, seperti Data Analyst yang berfokus pada eksplorasi dan visualisasi data, Data Scientist yang menggabungkan statistik dan pemodelan, serta Machine Learning Engineer yang bertanggung jawab pada implementasi dan deployment model. Setiap peran memiliki kompetensi teknis dan non-teknis yang berbeda, namun saling melengkapi dalam ekosistem AI.

1.4 Definisi Machine Learning Menurut Tom Mitchell

Tom Mitchell mendefinisikan machine learning sebagai berikut:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Definisi ini menekankan tiga komponen utama: **task**, **experience**, dan **performance**, yang menjadi dasar evaluasi seluruh sistem machine learning.

1.5 Filosofi Data dan Pengetahuan

Dalam konteks filsafat ilmu, data merupakan fakta mentah yang belum bermakna. Melalui proses pengolahan dan analisis, data berubah menjadi informasi, dan selanjutnya menjadi pengetahuan. Machine learning berperan sebagai jembatan otomatis dalam transformasi data menjadi pengetahuan yang dapat digunakan untuk pengambilan keputusan.

1.6 Tren Teknologi dan Prospek Karier

Tren global menunjukkan peningkatan signifikan dalam penggunaan AI di berbagai sektor seperti kesehatan, keuangan, pendidikan, dan manufaktur. Hal ini berdampak langsung pada meningkatnya kebutuhan tenaga kerja yang memahami data dan machine learning. Prospek karier di bidang ini bersifat jangka panjang dan lintas disiplin.

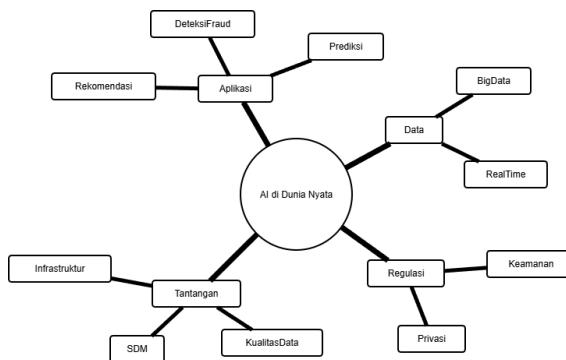
Pseudocode

```
experience = load_data()  
model = learn(experience)  
performance = evaluate(model)
```

BAB 2 Implementasi Artificial Intelligence di Dunia Nyata

2. Deskripsi

Bab ini membahas bagaimana konsep Artificial Intelligence (AI) dan machine learning benar-benar digunakan di dunia nyata, bukan cuma sebatas teori atau contoh di kelas. Pada praktiknya, AI banyak dimanfaatkan untuk membantu proses pengambilan keputusan, meningkatkan efisiensi kerja, dan mengolah data dalam jumlah besar. Selain itu, bab ini juga menyoroti tantangan yang sering muncul di industri, termasuk masalah regulasi, keamanan data, dan kesiapan sumber daya manusia..



2.1 Konsep Penerapan AI dalam Sistem Nyata

Dalam sistem nyata, AI biasanya tidak langsung menggantikan peran manusia, tetapi berfungsi sebagai alat bantu pengambilan keputusan. Sistem AI sering digunakan sebagai *decision support system*, di mana hasil prediksi atau analisis dari model dijadikan bahan pertimbangan oleh manusia. Pendekatan ini dianggap lebih aman dan realistik karena keputusan akhir tetap berada di tangan pengguna atau pihak yang bertanggung jawab. Selain itu, penerapan AI di dunia nyata juga harus mempertimbangkan faktor kecepatan, keandalan, dan kemudahan integrasi dengan sistem yang sudah ada. Model yang terlalu kompleks tetapi sulit digunakan justru jarang dipakai di industri.

2.2 Studi Kasus Penerapan Machine Learning

Contoh penerapan machine learning dapat dengan mudah ditemukan di kehidupan sehari-hari. Sistem rekomendasi pada e-commerce dan layanan streaming menggunakan machine learning untuk menyarankan produk atau konten berdasarkan riwayat pengguna. Di sektor keuangan, machine learning digunakan untuk mendeteksi transaksi mencurigakan atau penipuan secara otomatis. Studi kasus ini menunjukkan bahwa data historis memiliki peran penting dalam membangun model yang mampu memberikan prediksi yang bermakna dan relevan bagi kebutuhan bisnis.

2.3 Peran Data dalam Pengambilan Keputusan

Pengambilan keputusan berbasis data (*data-driven decision making*) semakin banyak diterapkan karena dianggap lebih objektif dibandingkan keputusan yang hanya mengandalkan intuisi. Dengan bantuan machine learning, organisasi dapat menganalisis data dalam jumlah besar dan menemukan pola yang sulit dilihat secara manual. Namun, keputusan berbasis data tetap perlu dikombinasikan dengan pemahaman konteks. Data yang salah atau bias dapat menghasilkan keputusan yang keliru meskipun model yang digunakan sudah canggih.

2.4 Regulasi, Keamanan, dan Kepatuhan Data

Dalam penerapan AI, aspek regulasi dan keamanan data tidak bisa diabaikan. Data yang digunakan sering kali mengandung informasi sensitif, seperti data pribadi pengguna. Oleh karena itu, organisasi harus mematuhi aturan perlindungan data dan memastikan sistem yang digunakan aman dari kebocoran. Penerapan AI yang baik biasanya didukung oleh kebijakan data governance yang jelas, mulai dari cara pengumpulan data hingga pengelolaan dan penyimpanannya.

2.5 Tantangan Implementasi AI di Industri

Meskipun AI menawarkan banyak keuntungan, implementasinya tidak selalu mudah. Tantangan yang sering dihadapi antara lain kualitas data yang belum baik, keterbatasan sumber daya manusia yang memahami AI, serta biaya infrastruktur yang cukup besar. Selain itu, tidak semua organisasi siap menerima perubahan teknologi, sehingga resistensi internal juga menjadi hambatan tersendiri.

Pseudocode

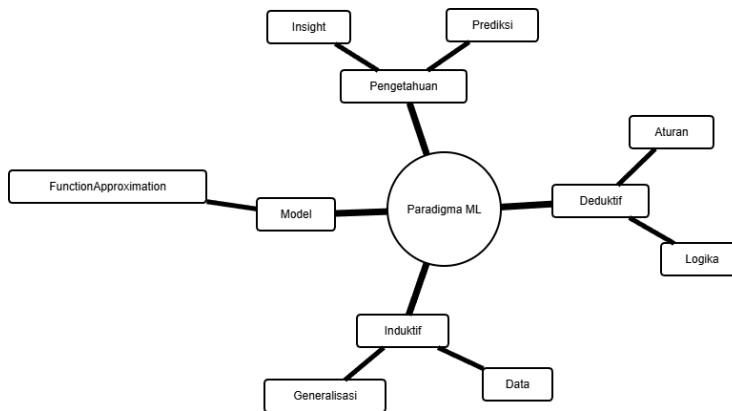
```
data = collect_data()  
  
model = train_model(data)  
  
prediction = model.predict(new_input)  
  
decision = support_decision(prediction)
```

BAB 3 Paradigma Filosofis Machine Learning

3. Deskripsi

Bab ini membahas cara berpikir atau landasan filosofis dibalik machine learning. Walaupun sering dianggap hanya soal coding dan algoritma, sebenarnya machine learning punya dasar pemikiran ilmiah

tentang bagaimana pengetahuan bisa dibentuk dari data. Dengan memahami hal ini, mahasiswa diharapkan tidak hanya “pakai algoritma”, tapi juga paham kenapa dan bagaimana model bisa belajar dari data.



3.1 Pendekatan Deduktif dan Induktif

Pendekatan deduktif adalah cara berpikir dari aturan umum ke kasus yang lebih spesifik. Contohnya, jika ada aturan “semua mahasiswa lulus jika nilai di atas 60”, maka kita tinggal mengecek nilainya. Pendekatan ini banyak digunakan pada sistem berbasis aturan. Sebaliknya, pendekatan induktif bekerja dari contoh-contoh data yang ada, lalu menarik kesimpulan umum. Machine learning lebih dekat dengan pendekatan ini karena model belajar dari data historis untuk menemukan pola, bukan dari aturan yang sudah ditentukan sejak awal.

3.2 Machine Learning sebagai Proses Induktif

Machine learning bisa dianggap sebagai proses belajar secara induktif. Model diberikan data, lalu secara bertahap mempelajari pola di dalamnya. Hasil dari proses ini bukan kebenaran mutlak, melainkan prediksi yang bersifat probabilistik. Artinya, selalu ada kemungkinan salah, terutama jika data yang digunakan kurang representatif.

3.3 Konsep Function Approximation

Dalam dunia nyata, hubungan antara input dan output biasanya sangat kompleks dan sulit dirumuskan secara matematis. Machine learning mencoba mendekati hubungan tersebut dengan membangun sebuah fungsi pendekatan. Model yang dilatih sebenarnya hanyalah perkiraan terbaik berdasarkan data yang tersedia.

3.4 Hubungan Data, Model, dan Pengetahuan

Data berperan sebagai bahan mentah, model berfungsi sebagai alat pengolah, dan pengetahuan merupakan hasil akhirnya. Jika data yang digunakan buruk atau tidak relevan, maka pengetahuan yang dihasilkan juga cenderung kurang akurat. Oleh karena itu, kualitas data dan pemilihan model sama-sama penting.

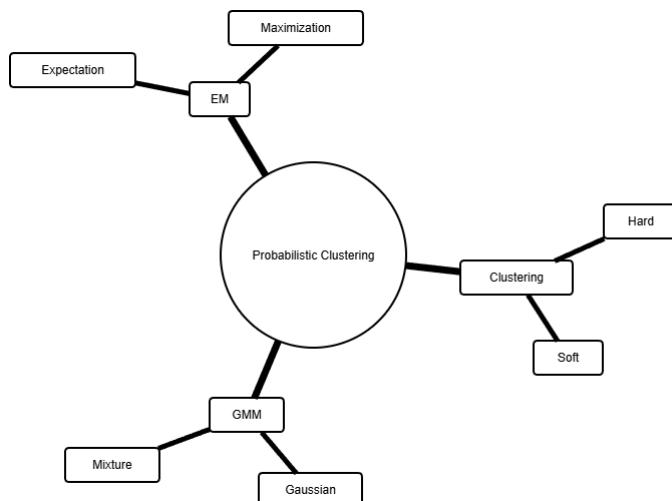
Pseudocode

```
data = collect_data()  
model = approximate_function(data)  
knowledge = extract_insight(model)
```

BAB 4 Probabilistic Clustering: Gaussian Mixture Model dan EM

4. Deskripsi

Bab ini membahas metode clustering berbasis probabilistik, khususnya Gaussian Mixture Model (GMM) dan algoritma Expectation-Maximization (EM). Pendekatan ini berbeda dari clustering biasa karena satu data tidak harus masuk ke satu cluster secara mutlak, tetapi memiliki peluang untuk menjadi bagian dari beberapa cluster sekaligus.



4.1 Konsep Dasar Clustering

Clustering bertujuan mengelompokkan data berdasarkan kemiripan tanpa menggunakan label. Pada clustering probabilistik, setiap data memiliki nilai probabilitas terhadap setiap cluster, sehingga lebih fleksibel dibandingkan clustering keras (hard clustering).

4.2 Model Probabilistik pada Clustering

Pendekatan probabilistik mengasumsikan bahwa data berasal dari distribusi tertentu. Dengan pendekatan ini, model mampu menangani data yang saling tumpang tindih antar cluster, yang sering terjadi pada data dunia nyata.

4.3 Gaussian Mixture Model (GMM)

GMM menganggap bahwa data dihasilkan dari gabungan beberapa distribusi Gaussian. Setiap distribusi Gaussian mewakili satu cluster dengan parameter seperti mean dan covariance. Model ini cocok digunakan ketika bentuk cluster tidak terlalu kaku.

4.4 Algoritma Expectation-Maximization

Algoritma EM bekerja secara iteratif melalui dua langkah utama. Pada tahap Expectation, model menghitung probabilitas setiap data terhadap masing-masing cluster. Pada tahap Maximization, parameter model diperbarui berdasarkan probabilitas tersebut. Proses ini diulang sampai hasilnya stabil.

4.5 Kriteria Konvergensi dan Interpretasi Model

Proses EM dikatakan konvergen ketika perubahan nilai log-likelihood sudah sangat kecil. Interpretasi model dilakukan dengan melihat distribusi Gaussian dan bagaimana data tersebar pada setiap cluster.

Pseudocode EM

```
initialize_parameters()
```

```
repeat:
```

```
    E_step()
```

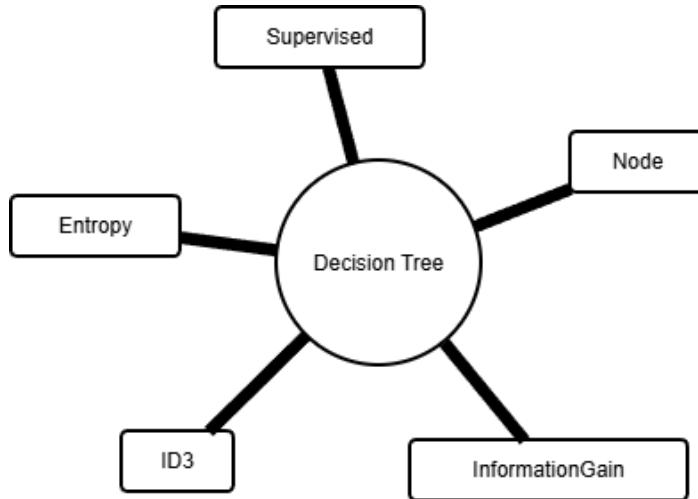
```
    M_step()
```

```
until convergence
```

BAB 5 – Decision Tree dan Algoritma ID3

5. Deskripsi

Bab ini membahas decision tree sebagai salah satu algoritma supervised learning yang mudah dipahami dan sering digunakan. Fokus utama diberikan pada algoritma ID3 yang memanfaatkan konsep entropy dan information gain untuk membangun pohon keputusan.



5.1 Konsep Supervised Learning

Supervised learning menggunakan data yang sudah memiliki label sebagai acuan pembelajaran. Model belajar dari pasangan input dan output untuk kemudian digunakan dalam melakukan prediksi pada data baru.

5.2 Struktur dan Komponen Decision Tree

Decision tree terdiri dari beberapa komponen utama, yaitu root node sebagai awal keputusan, internal node sebagai percabangan, dan leaf node sebagai hasil akhir. Struktur ini membuat decision tree mudah dipahami dan dijelaskan.

5.3 Entropy sebagai Ukuran Ketidakpastian

Entropy digunakan untuk mengukur seberapa acak atau tidak pastinya distribusi kelas dalam suatu dataset. Semakin tinggi nilai entropy, semakin sulit model menentukan keputusan.

5.4 Information Gain

Information gain digunakan untuk memilih atribut terbaik dalam membagi data. Atribut dengan information gain tertinggi akan dipilih karena mampu mengurangi ketidakpastian paling besar.

5.5 Algoritma ID3

Algoritma ID3 membangun decision tree secara bertahap dan rekursif. Pada setiap tahap, atribut dengan information gain terbesar dipilih sebagai node pemisah.

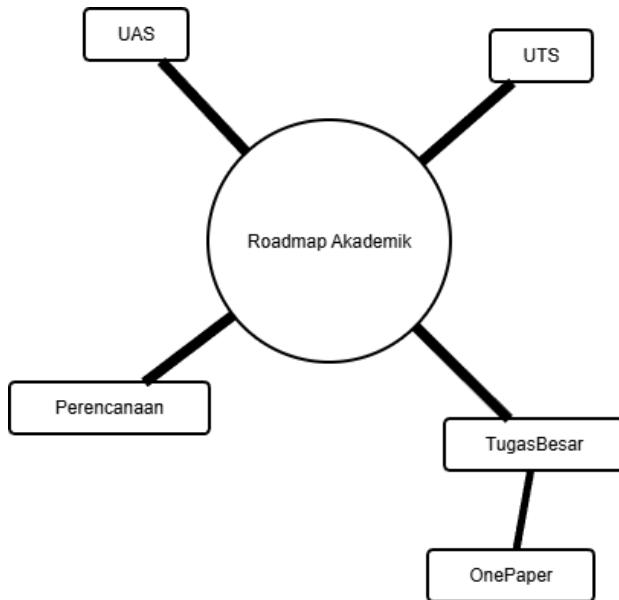
Pseudocode ID3

```
function ID3(data, attributes):  
    if data homogen:  
        return leaf  
    best_attr = select_highest_information_gain(attributes)  
    split_data(best_attr)
```

BAB 6 Roadmap Akademik dan Strategi Tugas Besar

6. Deskripsi

Bab ini membahas strategi akademik yang dapat membantu mahasiswa mengikuti mata kuliah machine learning dengan lebih terarah. Selain memahami materi, mahasiswa juga perlu memiliki strategi dalam menghadapi evaluasi seperti UTS, UAS, dan tugas besar.



6. Penjelasan Detail Materi

6. 1 Arsitektur Tugas Besar: Kontinuitas UTS ke UAS

Tugas besar dirancang sebagai proyek riset mini yang berlangsung dalam dua tahap yang saling

terhubung.

Fase UTS – Classical Machine Learning

Mahasiswa wajib menyelesaikan masalah menggunakan algoritma supervised learning klasik seperti

Decision Tree, SVM, Naive Bayes, atau regresi. Tujuannya adalah membangun baseline kinerja model,

yaitu tolok ukur awal seberapa baik masalah dapat diselesaikan dengan metode statistik standar.

Fase UAS – Deep Learning

Masalah yang sama sangat disarankan diselesaikan kembali menggunakan pendekatan deep learning

seperti neural network, CNN, atau RNN.

Mahasiswa harus menyertakan argumen ilmiah:

- Mengapa deep learning diperlukan?
- Apakah datanya cukup kompleks dan besar?
- Apa hipotesis peningkatan performa dibanding metode klasik?

Dengan pendekatan ini, mahasiswa belajar membandingkan metode dan menganalisis perbedaan hasil

secara ilmiah.

6.2 One-Paper Strategy (Strategi Efisiensi Riset)

Strategi agar mahasiswa tidak bekerja dua kali.

Skenario Ideal:

- Sejak UTS memilih dataset yang kompleks.
- Dataset yang sama digunakan untuk UTS dan UAS.
- Laporan hanya satu, tetapi berevolusi.
- Bab pendahuluan dan dataset tetap.
- Bab metodologi dan hasil diperluas dengan perbandingan ML klasik vs deep learning.

Skenario Buruk:

- Jika UTS menggunakan dataset terlalu sederhana (misalnya Iris atau Titanic), dataset tersebut tidak cocok untuk deep learning.

Gambar 9 Mind Map Roadmap Akademik dan Karier

- Akibatnya, saat UAS harus mulai dari nol: cari data baru, menulis ulang laporan, dan melakukan eksperimen ulang.

Strategi ini melatih mahasiswa untuk berpikir jangka panjang dalam perencanaan riset.

6.3 Standar Deliverables & Simulasi Software Engineering

Pengumpulan tugas mengikuti standar industri, bukan sekadar kirim file.

Manajemen Repositori GitHub:

- Wajib memiliki satu repositori utama.
- Struktur folder rapi:
 - /UTS
 - /UAS

Aset Wajib (Deliverables):

1. Proposal riset dengan hipotesis ilmiah.
2. Dataset (file mentah atau link sumber resmi).
3. Kode (Jupyter Notebook atau script Python) yang bisa dijalankan ulang.
4. Referensi jurnal ilmiah sebagai dasar teori.
5. Laporan akhir sesuai format jurnal komputer.

Mahasiswa dilatih menjadi peneliti dan software engineer, bukan sekadar pengguna library.

6.4 Hidden Curriculum: Networking dan Realita Industri

Ditekankan bahwa nilai akademik saja tidak cukup.

Kekuatan Komunitas

Banyak pekerjaan terbaik tidak diiklankan secara terbuka, melainkan melalui rekomendasi dari teman

atau relasi komunitas.

Ekosistem Open Source:

Komunitas seperti Python ID dan Java User Group sering mengadakan event gratis atau murah.

Mahasiswa dianjurkan hadir untuk:

- Belajar langsung dari praktisi.
- Membangun relasi dengan senior industri.

Relasi dipandang sebagai investasi karier jangka panjang.

6.5 Logistik dan Ekspektasi Kesulitan

Machine Learning adalah irisan dari:

- Matematika
- Probabilitas
- Pemrograman

Karena itu, mata kuliah ini berat dan tidak bisa dianggap remeh. Materi ini menjadi fondasi untuk mata

kuliah Artificial Intelligence di semester berikutnya. Jika dasar ML lemah, maka materi AI akan jauh lebih

sulit.

6.6 Sample Code / Pseudocode

Struktur Repositori Tugas Besar:

Project_ML/

|

|—— UTS/

| |—— data/

| |—— classical_model.ipynb

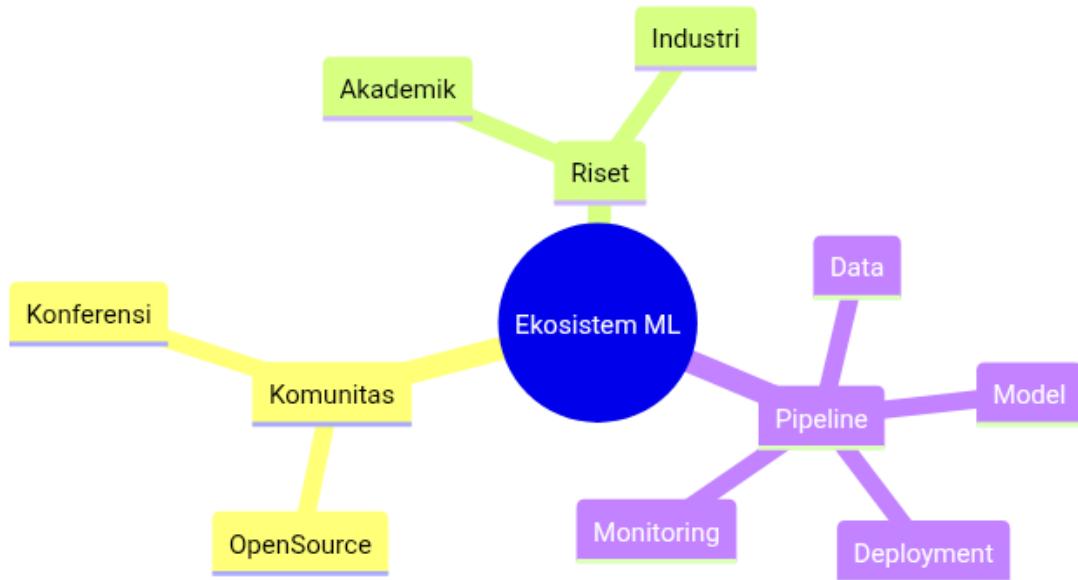
```
| └── proposal.pdf
|
└── UAS/
    ├── data/
    |   ├── deep_learning_model.ipynb
    |   ├── final_report.pdf
    |
    └── references/
        ├── paper1.pdf
        └── paper2.pdf
Contoh Implementasi Python (sketsa)
# Fase UTS
latih_model_klasik(dataset)
hitung_performa -> baseline_score
# Fase UAS
latih_model_deep_learning(dataset)
hitung_performa -> dl_score
bandingkan baseline_score dan dl_score
analisis hasil dan jelaskan secara ilmiah
```

BAB 7 Komunitas,Riset dan Pipeline Industri

7. Deskripsi

Perkembangan machine learning tidak terjadi secara terpisah, melainkan didukung oleh ekosistem yang kuat antara komunitas. Dunia riset dan industri. Banyak teknologi machine learning yang digunakan saat ini berawal dari penelitian akademik yang kemudian

dikembangkan dan diterapkan didunia industri. Oleh karena itu pemahaman mengenai ketiganya penting agar mahasiswa tidak hanya memahami teori, tetapi juga mengetahui bagaimana ilmu tersebut digunakan didunia nyata.



7.1 Peran Komunitas Dalam Pengembangan Keilmuan

Komunitas memiliki peran penting dalam perkembangan machine learning melalui publikasi riset, pengembangan perangkat lunak open-source, serta diskusi ilmiah. Platform seperti Github, arXiv, dan forum komunitas memungkinkan penyebarluasan pengetahuan secara global dan kolaboratif.

7.2 Hubungan Akademik dan Industri

Akademik berfokus pada pengembangan teori dan metode terbaru, sedangkan industri berfokus pada penerapan dan skalabilitas solusi. Kolaborasi antara keduanya menghasilkan inovasi yang relevan secara ilmiah dan bernilai praktis.

7.3 Machine Learning Pipeline

Machine learning pipeline mencakup tahapan pengumpulan data, preprocessing, pelatihan model, evaluasi, serta deployment. Pipeline yang terstruktur membantu memastikan konsistensi dan reproduksibilitas eksperimen.

7.4 Deployment dan Monitoring Model

Setelah model di-deploy, performanya harus dipantau secara berkala untuk mendeteksi degradasi performa atau perubahan pola data (*data drift*).

Pseudocode Pipeline

```
data = collect_data()  
model = train(data)  
deploy(model)  
monitor(model)
```

BAB 8 Disiplin Akademik dan Validasi Dataset

8. deskripsi

Bab ini menerapkan pentingnya disiplin akademik dan validasi dataset dalam proyek machine learning. Kesalahan pada tahap pemilihan atau validasi dataset dapat menyebabkan hasil penelitian yang menyesatkan dan tidak dapat direproduksi.



8.1 Etika Penggunaan Data

Etika penggunaan data mencakup perlindungan privasi, penghindaran bias, dan penggunaan data secara bertanggung jawab. Pelanggaran etika dapat merusak kredibilitas penelitian dan berdampak hukum.

8.2 Pemilihan dan Kualitas Dataset

Dataset harus relevan dengan permasalahan yang diteliti dan memiliki kualitas yang memadai. Dataset yang tidak representatif dapat menghasilkan model yang bias.

8.3 Validasi Target dan Fitur

Validasi dilakukan untuk memastikan bahwa target dan fitur tidak mengandung *data leakage* yang dapat meningkatkan performa model secara tidak realistik.

8.4 Perbedaan Supervised dan Unsupervised Learning

Supervised learning menggunakan label sebagai target pembelajaran, sedangkan unsupervised learning berfokus pada eksplorasi struktur data tanpa label.

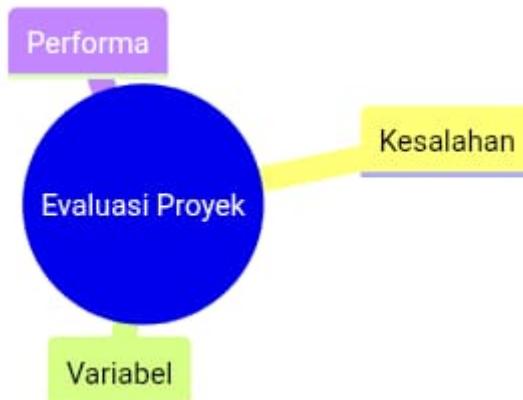
Pseudocode Validasi Dataset

```
check_bias(data)
validate_target(data)
validate_features(data)
```

BAB 9 Review Teknis dan Evaluasi Proyek

9. deskripsi

Bab ini membahas evaluasi teknis proyek machine learning secara menyeluruh, termasuk identifikasi kesalahan umum, analisis variabel, serta penilaian performa model. Evaluasi ini penting untuk memastikan kualitas ilmiah dan teknis proyek.



9.1 Kesalahan Umum dalam Proyek Machine Learning

Kesalahan umum meliputi overfitting, pemilihan metrik evaluasi yang tidak tepat, serta kesalahan preprocessing data. Identifikasi kesalahan ini membantu meningkatkan kualitas proyek.

9.2 Analisis Target dan Variabel

Analisis hubungan antara variabel input dan target membantu memahami kontribusi masing-masing fitur terhadap prediksi model.

9.3 Evaluasi Model dan Performa

Evaluasi performa model dilakukan menggunakan metrik yang sesuai dengan tujuan masalah, seperti akurasi, precision, recall, dan F1-score.

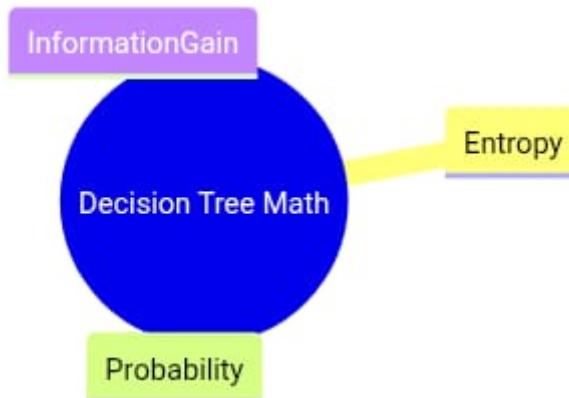
Pseudocode Evaluasi

```
pred = model.predict(X_test)  
evaluate(pred, y_test)
```

BAB 10 Matematika di Balik Decision Tree

10. deskripsi

Bab ini membahas dasar matematis dari algoritma decision tree, khususnya konsep entropi dan information gain. Pemahaman matematis ini penting agar mahasiswa tidak hanya menggunakan algoritma sebagai *black box*.



10.1 Konsep Entropi Secara Matematis

Entropi didefinisikan sebagai ukuran ketidak pastian dalam distribusi probabilitas suatu variabel acak. Nilai entropi maksimum terjadi ketika distribusi kelas merata.

10.2 Perhitungan Information Gain

Information gain dihitung sebagai selisih antara entropi awal dan entropi setelah data dibagi berdasarkan atribut tertentu.

10.3 Studi Kasus Play Tennis Dataset

Dataset Play Tennis digunakan sebagai contoh klasik untuk menunjukkan perhitungan entropy dan information gain secara manual.

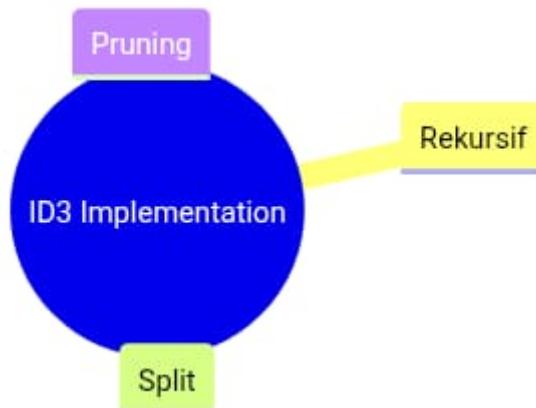
Pseudocode Perhitungan

```
entropy = calculate_entropy(data)
gain = entropy_before - entropy_after
```

BAB 11 Implementasi Teknis Algoritma ID3

11. deskripsi

Bab ini membahas bagaimana konsep teoritis algoritma ID3 diterjemahkan ke dalam implementasi teknis. Fokus diberikan pada proses rekursif dan teknik optimasi untuk meningkatkan generalisasi model.



11.1 Translasi Konsep Teori ke Implementasi

Implementasi ID3 membutuhkan penerjemahan konsep entropy dan information gain ke dalam code program.

11.2 Proses Rekursif pada Decision Tree

Decision tree dibangun secara rekursif hingga memenuhi kondisi penghentian tertentu.

11.3 Optimasi dan Pruning

Pruning digunakan mengurangi kompleksitas tree dan mencegah overfitting.

Pseudocode ID3 Lengkap

```
function build_tree(data):
    if stopping_condition:
        return leaf
    best_feature = select_feature(data)
```

```
split_data(best_feature)
```

BAB 12 Gaussian Mixture Model Lanjutan

12. deskripsi

Bab ini membahas aspek lanjutan dari Gaussian Mixture Model, termasuk interpretasi parameter, evaluasi konvergensi, dan analisis kualitas clustering.



12.1 Interpretasi Distribusi Gaussian

Setiap komponen Gaussian memiliki parameter mean dan covariance yang menggambarkan pusat dan bentuk cluster.

12.2 Analisis Parameter Model

Analisis parameter membuat memahami struktur data dan distribusi cluster.

12.3 Konvergensi dan Evaluasi Model

Evaluasi dilakukan menggunakan log-likelihood dan visualisasi hasil clustering.

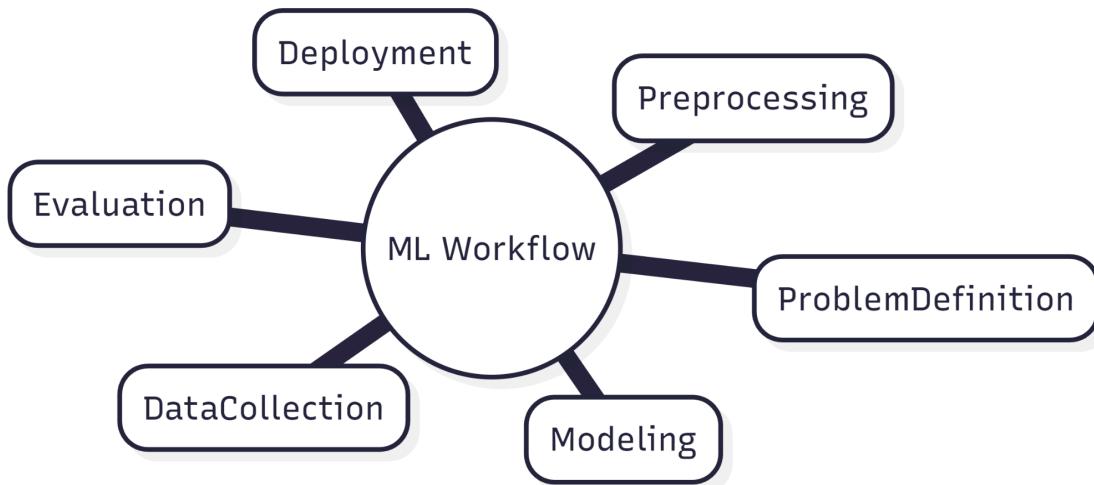
Pseudocode Evaluasi GMM

```
fit_gmm(data)
check_log_likelihood()
visualize_clusters()
```

BAB 13 Workflow dan Implementasi Machine Learning Secara End-to-End

13. deskripsi

Bab ini membahas alur kerja machine learning dari awal sampai akhir. Di pertemuan ini, mahasiswa diajak melihat machine learning bukan cuma soal algoritma, tapi sebagai sebuah proses yang utuh. Mulai dari memahami masalah, menyiapkan data, membangun model, sampai mengevaluasi hasilnya. Bab ini penting karena di dunia nyata, kegagalan proyek ML sering terjadi bukan karena algoritmanya, tapi karena workflownya kurang rapi.



13.1 Definisi Masalah dan Tujuan Model

Langkah pertama dalam workflow machine learning adalah memahami masalah yang ingin diselesaikan. Tanpa definisi masalah yang jelas, model yang dibangun bisa saja bagus secara teknis, tetapi tidak relevan secara bisnis atau kebutuhan pengguna.

Di tahap ini, mahasiswa perlu menentukan apakah masalahnya termasuk klasifikasi, regresi, atau clustering, serta metrik apa yang digunakan untuk menilai keberhasilan model.

13.2 Pengelola Data dan Feature Engineering

Setelah masalah jelas, fokus berpindah ke data. Data mentah jarang bisa langsung dipakai, sehingga perlu proses preprocessing seperti pembersihan data, normalisasi, dan encoding. Selain

itu, feature engineering berperan penting dalam meningkatkan performa model dengan mengekstrak informasi yang lebih relevan dari data.

Tahap ini sering memakan waktu paling lama dalam workflow machine learning.

13.3 Training, Evaluasi, dan Deployment

Model kemudian dilatih menggunakan data training dan diuji menggunakan data validasi atau testing. Evaluasi dilakukan untuk memastikan model tidak hanya bagus di data latihan, tetapi juga mampu bekerja dengan baik pada data baru.

Jika hasilnya sudah memuaskan, model dapat di *deploy* ke lingkungan produksi. Namun, deployment bukan akhir dari proses karena model tetap perlu dipantau performanya secara berkala.

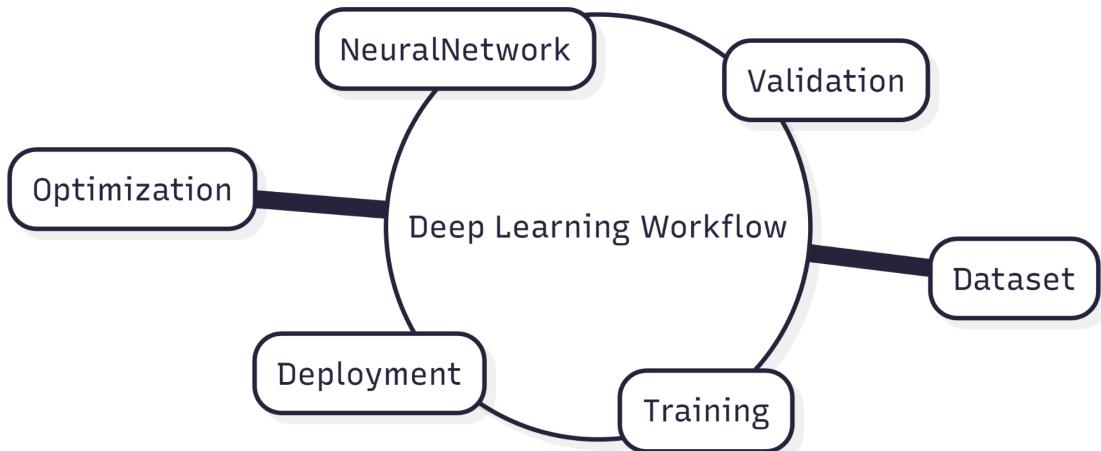
Pseudocode Workflow

```
define problem()  
prepare data()  
train_model()  
evaluate model()  
deploy model()
```

BAB 14 Workflow dan Implementasi Deep Learning

14. deskripsi

Bab terakhir ini membahas workflow khusus untuk deep learning. Walaupun secara umum mirip dengan machine learning tradisional, deep learning memiliki kompleksitas tambahan seperti arsitektur jaringan, tuning hyperparameter, dan kebutuhan komputasi yang lebih besar. Bab ini menjadi penutup yang menghubungkan seluruh materi sebelumnya dengan praktik deep learning di dunia nyata



14.1 Persiapan Dataset untuk Deep Learning

Deep learning sangat bergantung pada jumlah dan kualitas data. Dataset biasanya perlu dibagi menjadi training, validation, dan testing set. Selain itu, data augmentation sering digunakan untuk meningkatkan variasi data, terutama pada bidang seperti computer vision.

Persiapan dataset yang baik akan sangat mempengaruhi hasil akhir model.

14. 2 Training Model dan Optimasi

Training deep learning melibatkan proses forward propagation dan backpropagation secara berulang. Optimizer seperti SGD atau sumber Adam digunakan untuk memperbarui bobot jaringan. Proses ini membutuhkan waktu dan sumber daya komputasi yang cukup besar.

Mahasiswa juga perlu memahami konsep seperti learning rate, batch size, dan epoch agar training berjalan efektif.

14.4 Evaluasi, Deployment, dan Monitoring

Setelah training selesai, model dievaluasi menggunakan data testing. Jika performanya sudah sesuai target, model dapat digunakan dalam aplikasi nyata. Namun, seperti machine learning biasa, model deep learning juga perlu dimonitor karena performanya bisa menurun seiring perubahan data di dunia nyata.

Pseudocode Deep Learning Workflow

```
load_dataset()  
build_neural_network()  
train network()  
evaluate_network()  
deploy model()
```

Kesimpulan dari rangkuman :

berdasarkan rangkuman dari materi perkuliahan machine learning ini adalah dapat disimpulkan bahwa machine learning bukan sekadar kumpulan algoritma atau teknik pemrograman, tetapi merupakan sebuah pendekatan ilmiah dan sistematis untuk membangun sistem cerdas yang mampu belajar dari data. Seluruh bab yang dibahas saling terhubung dan membentuk pemahaman utuh, mulai dari konsep dasar, landasan filosofi, hingga implementasi teknis di dunia nyata.

Materi diawali dengan pengenalan machine learning dalam konteks Artificial Intelligence serta prospek kariernya, yang menunjukkan bahwa bidang ini memiliki peran penting dan relevan dengan kebutuhan industri saat ini. Selanjutnya, pembahasan implementasi AI di dunia nyata menegaskan bahwa keberhasilan machine learning tidak hanya ditentukan oleh algoritma, tetapi juga oleh kualitas data, pemahaman konteks, regulasi, dan kesiapan organisasi.

Aspek filosofi dan matematis yang dibahas, seperti pendekatan induktif, function approximation, entropy, dan information gain, membantu mahasiswa memahami bahwa model machine learning bekerja berdasarkan pendekatan probabilistik, bukan kebenaran mutlak. Hal ini diperkuat dengan pembahasan algoritma seperti Decision Tree(ID3) dan Gaussian Mixture Model yang menunjukkan bagaimana teori diterjemahkan ke dalam praktik.

Selain sisi teknis, materi ini juga menekankan pentingnya disiplin akademik, etika penggunaan data, validasi dataset, serta evaluasi proyek agar hasil yang diperoleh bersifat valid, dapat dipercaya, dan dapat direproduksikan. Pembahasan workflow machine learning dan deep learning di akhir materi menegaskan bahwa keberhasilan proyek ML sangat bergantung pada alur kerja yang rapi dan terstruktur, bukan hanya pada kompleksitas model.

Secara keseluruhan, rangkuman ini memberikan gambaran komprehensif tentang machine learning sebagai kombinasi antara teori, praktik, dan strategi akademik maupun industri. Pemahaman yang baik terhadap seluruh tahapan ini diharapkan dapat menjadi bekal mahasiswa untuk melanjutkan ke materi Artificial Intelligence dan Deep Learning yang lebih lanjut, serta mempersiapkan diri menghadapi tantangan di dunia riset maupun industri.