

PRAKTIKUM *MACHINE LEARNING*

TUGAS PENDAHULUAN

TAHUN AJAR 2024/2025



NIM : Bambang Istijab
NAMA : 105222007

1. **Jelaskan apa itu *Machine Learning* (ML) dan bagaimana cara kerjanya dibandingkan dengan pemrograman konvensional. Berikan contoh yang menggambarkan perbedaannya. (20%)**

Jawaban:

Machine Learning (ML) adalah cabang kecerdasan buatan (AI) yang berfokus pada pengembangan algoritma yang memungkinkan sistem komputer belajar dari data dan meningkatkan kinerjanya tanpa instruksi eksplisit [1].

Machine Learning (ML) bekerja dengan memanfaatkan hubungan matematika yang ada diantara data input dan output [1]. Pendekatan konvensional dalam pemrograman biasanya melibatkan penulisan aturan dan logika secara langsung oleh programmer. Di sisi lain, Machine Learning (ML) memungkinkan sistem untuk menyesuaikan diri dengan pola dalam data, bahkan jika pola tersebut kompleks atau sulit dijabarkan secara eksplisit oleh manusia. Hal ini memungkinkan penyelesaian masalah yang lebih adaptif dan canggih [2].

Salah satu contoh yang menggambarkan perbedaan dari Machine Learning (ML) dengan pemrograman konvensional adalah program yang dapat mengidentifikasi spam pada email. Dengan pemrograman konvensional, kita harus menentukan semua aturan yang mendefinisikan email spam, seperti kata-kata tertentu dalam subjek atau isi email. Setiap aturan harus ditulis secara manual oleh programmer. Sedangkan dengan Machine Learning (ML) kita akan mengumpulkan sejumlah besar email yang sudah dikategorikan sebagai "spam" atau "bukan spam". Algoritma ML kemudian dilatih dengan data ini untuk mengenali pola yang membedakan antara spam dan non-spam. Setelah pelatihan, model dapat mengklasifikasikan email baru sebagai spam atau bukan berdasarkan pola yang telah dipelajarinya [2].

2. **Sebutkan dan jelaskan tiga kategori utama dalam *Machine Learning*. Berikan contoh penggunaan masing-masing kategori. (20%)**

Jawaban:

1. Supervised Learning

Supervised learning (pelajaran dengan arahan) adalah paradigma pembelajaran mesin di mana model diajarkan untuk memetakan input ke output, dengan menggunakan pasangan data latih yang terdiri dari input beserta label atau output yang sesuai [2]. Salah

satu contoh kasus supervised learning adalah Klasifikasi email spam. *Supervised learning* digunakan untuk membangun model klasifikasi yang dapat membedakan antara email spam dan bukan spam. Model ini memeriksa konten dan metadata email untuk menentukan apakah suatu email termasuk dalam kategori spam atau bukan [2].

- Fitur : Konten email, metadata.
- Label : Spam atau bukan

2. Unsupervised Learning

Unsupervised learning (pemelajaran tanpa arahan) adalah paradigma pemelajaran mesin di mana model diajarkan untuk menemukan pola atau struktur dalam data tanpa adanya label atau petunjuk pada output yang diharapkan. Tujuan utama *unsupervised learning* adalah menggali struktur intrinsik dari data, seperti kelompok-kelompok yang mungkin ada atau fitur-fitur yang saling terkait [2].

Salah satu contoh kasus unsupervised learning adalah K-Means untuk Clustering, *Unsupervised learning* dapat digunakan untuk melakukan clustering pada data. Sebagai contoh, algoritma K-Means dapat mengelompokkan data ke dalam cluster berdasarkan kemiripan fitur[2]. Contoh fiturnya adalah posisi geografis, preferensi pembelian pelanggan dan lainnya.

3. Reinforcement Learning

Reinforcement learning adalah paradigma pemelajaran mesin di mana agen belajar untuk membuat keputusan atau tindakan dalam sistem yang dinamis dengan maksud untuk memaksimalkan jumlah *reward* (imbalan) yang diterima dari lingkungannya[2].

Salah satu contoh kasus reinforcement learning adalah kendali robotika, Dimana agennya yaitu robot atau kendaraan otonom, sedangkan rewardnya adalah keberhasilan dalam menyelesaikan tugas.

3. Jelaskan lima tahap utama dalam alur kerja *Machine Learning*, dan berikan contoh atau ilustrasi yang relevan untuk setiap tahap. (20%)

Jawaban:

1. Pengumpulan Data

Dalam machine learning, langkah pertama adalah pengumpulan data. Kualitas dan kuantitas data yang kita kumpulkan akan secara langsung memengaruhi seberapa baik model prediktif kita nantinya. Misalnya, jika kita ingin membuat model yang dapat membedakan antara wine dan beer, kita perlu mengumpulkan data tentang warna dan kandungan alkohol dari kedua minuman ini. Data ini akan menjadi dasar pelatihan model yang akan kita buat [3].

2. Persiapan Data

Persiapan data ini termasuk menggabungkan semua data, mengacak urutan data, dan melakukan visualisasi data untuk memahami hubungan antara variabel yang berbeda. Selain itu, kita perlu membagi data menjadi dua bagian: data pelatihan dan data evaluasi.

3. Memilih Model

Ada banyak jenis model yang tersedia, seperti decision trees, support vector machines, dan neural networks. Pemilihan model harus didasarkan pada masalah yang sedang kita hadapi, sehingga model yang dipilih benar-benar sesuai dengan masalah tersebut [4]. Dalam contoh kita, dengan hanya dua fitur (warna dan persentase alkohol), kita dapat menggunakan model linear sederhana untuk membedakan antara wine dan beer [3].

4. Pelatihan Model

Setelah model terpilih, model tersebut harus dilatih dengan menggunakan data latih yang telah disiapkan sebelumnya. Proses pelatihan akan mengoptimalkan model agar dapat memecahkan masalah yang kita hadapi dengan baik [4]. Setelah model terpilih, model tersebut harus dilatih dengan menggunakan data latih yang telah disiapkan sebelumnya. Proses pelatihan akan mengoptimalkan model agar dapat memecahkan masalah yang kita hadapi dengan baik [3].

5. Evaluasi Model

Langkah selanjutnya adalah mengevaluasi model yang telah dilatih. Kita dapat menggunakan data uji untuk mengetahui seberapa baik performa model tersebut. Ada beberapa metrik yang dapat kita gunakan untuk mengevaluasi model, seperti akurasi atau mean squared error [4]. Evaluasi memberi kita gambaran tentang seberapa baik model Anda dapat berperforma di dunia nyata [3].

4. Apa yang dimaksud dengan "*training data*" dan "*testing data*"? Mengapa penting untuk memisahkan keduanya dalam proses pengembangan model *Machine Learning*? (20%)

Jawaban:

- Training data

Training data adalah data yang digunakan untuk membuat model *machine learning*. Dalam proses pembuatan model, data latih diinputkan ke dalam algoritma *machine learning* untuk mempelajari pola dan hubungan antara input dan output. Setelah algoritma selesai belajar, model yang dihasilkan akan digunakan untuk membuat prediksi pada data baru [5].

- Testing data

Testing data adalah data yang digunakan untuk mengevaluasi kinerja model *machine learning* yang telah dibuat. Data uji ini digunakan untuk melihat seberapa

baik model yang dibuat dapat menangani data yang belum pernah dilihat sebelumnya, atau untuk mengevaluasi seberapa baik model tersebut dapat mengatasi permasalahan yang dihadapi [5].

Untuk mencapai tujuan tersebut, machine harus diberi tahu mana “set data” yang harus dicapai / dilampaui, dan mana “set data” yang bisa digunakan untuk mencapai / melampaui goal ini tadi. “Set data” untuk dicapai inilah Test-Set, sedangkan “set data” untuk mencapainya disebut Training-Set. Training-Set ini nantinya akan digunakan untuk membuat model machine learning, sedangkan Test-Set ini akan digunakan untuk menguji performa dan kebenaran (terhadap korelasi) dalam model yang bersangkutan [6].

Memisahkan training data dan testing data juga adalah praktik standar dalam pengembangan model Machine Learning untuk memastikan bahwa model yang dihasilkan andal dan siap menghadapi data dunia nyata.

5. **Apa yang dimaksud dengan “*overfitting*” dan “*underfitting*” dalam konteks *Machine Learning*? Bagaimana cara mendeteksi dan mengatasi masalah ini? (20%)**

Jawaban:

- **Overfitting**

Overfitting adalah perilaku [machine learning](#) yang tidak diinginkan yang terjadi saat model machine learning memberikan prediksi akurat untuk data pelatihan tetapi tidak untuk data baru [7].

Metode terbaik untuk mendeteksi model overfit adalah dengan menguji model machine learning pada lebih banyak data dengan representasi komprehensif dari kemungkinan nilai dan jenis data input [7]. Salah satu metode pengujian untuk overfitting ini adalah “Validasi silang K-fold”, metode ini adalah salah satu metode pengujian yang digunakan dalam praktik. Dalam metode ini, ilmuwan data membagi set pelatihan menjadi K subset atau set sampel berukuran sama yang disebut fold [7].

Cara mengatasi overfitting ini dengan melakukan diversifikasi dan penskalaan set data pelatihan Anda atau menggunakan beberapa strategi ilmu data lainnya, seperti:

- Penghentian awal
- Pemangkasan
- Regularisasi
- Ensembling
- Augmentasi data

- **Underfitting**

Underfitting adalah jenis kesalahan lain yang terjadi saat model tidak dapat menentukan hubungan yang berarti antara data input dan output. Anda mendapatkan model

underfit jika model tersebut tidak dilatih dalam jangka waktu yang sesuai pada sejumlah besar titik data [7].

Salah satu cara umum untuk mendeteksi underfitting adalah dengan melihat kembali kurva pembelajaran model. Dalam model underfitting, kinerja set pelatihan dan set validasi akan buruk, dan kesenjangan di antara keduanya tidak akan berkurang meskipun lebih banyak data ditambahkan [8]. Untuk mencegahnya kita dapat melakukan beberapa cara berikut ini:

- Meningkatkan kompleksitas model
- Mengurangi regularisasi
- Menambahkan lebih banyak data pelatihan

PRAKTIKUM MACHINE LEARNING

TUGAS PENDAHULUAN

TAHUN AJAR 2024/2025



REFERENSI

- [1] H. H. Khairan, "Apa Itu Machine Learning? Arti, Contoh, Cara Kerja, Jenis, dan Keuntungannya," Lawencon, 9 Agustus 2024. [Online]. Available: <https://www.lawencon.com/machine-learning/>
- [2] Rumah Coding, "Pengertian Machine Learning," 1.1 tahun lalu. [Online]. Available: <https://rumahcoding.id/belajar/machine-learning-with-python/pengenalan-machine-learning/definisi-machine-learning/>.
- [3] Team Algoritma, "7 Langkah dalam Machine Learning: Dari Data hingga Prediksi," Algoritma Data Science School, 9 November 2023. [Online]. Available: <https://blog.algorit.ma/langkah-machine-learning/>.
- [4] U. Riswanto, "Dari Data hingga Implementasi: Alur Kerja Machine Learning yang Harus Anda Ketahui," Medium, 2.1 tahun lalu. [Online]. Available: <https://ujangriswanto08.medium.com/dari-data-hingga-implementasi-alur-kerja-machine-learning-yang-harus-anda-ketahui-632dad0714dd>
- [5] Administrator, "Jenis Data yang Dibutuhkan dalam Machine Learning," Ivosights, 27 Januari 2023. [Online]. Available: <https://ivosights.com/read/artikel/machine-learning-jenis-data-yang-dibutuhkan-dalam-machine-learning-19b45dd52f6d#:~:text=diajar%20secara%20eksplisit,-.Dalam%20machine%20learning%2C%20data%20yang%20digunakan%20dibagi%20menjadi%20dua%20jenis,kemampuan%20model%20yang%20telah%20dilatih.>
- [6] R. Kukuh, "Memisah Dataset Menjadi Training-Set dan Test-Set," *Machine Learning ID*, 9 Februari 2018. [Online]. Available: <https://medium.com/%40rkukuh/memisah-dataset-menjadi-training-set-dan-test-set-19b45dd52f6d>
- [7] Amazon Web Services, "Apa itu Overfitting? - Penjelasan tentang Overfitting di Machine Learning," 2025. [Online]. Available: https://aws.amazon.com/what-is/overfitting/?utm_source=chatgpt.com
- [8] [1] T. Babbar, "Are Your Machine Learning Models Making These Common Mistakes? Learn How to Avoid Overfitting and Underfitting," *AlliedOffsets Blog*, 14 April 2023. [Online]. Available: <https://blog.alliedoffsets.com/are-your-machine-learning-models-making-these-common-mistakes-learn-how-to-avoid-overfitting-and-underfitting>

PRAKTIKUM *MACHINE LEARNING*
TUGAS PENDAHULUAN
TAHUN AJAR 2024/2025



TAMBAHAN