

# Pertemuan 9

## Konsep Pembelajaran Mesin

Alim Misbullah, S.Si., M.S.  
Jurusan Informatika  
Ganjil 2023/2024

# Topik Pembahasan

- Pengenalan Pembelajaran Mesin
- Konsep Utama dalam Pembelajaran Mesin
- Tipe-tipe Pembelajaran Mesin
- Pemodelan Data
- Algoritma Klasifikasi
- Algoritma Regresi

# Pengenalan Pembelajaran Mesin

# Definisi Pembelajaran Mesin

- **Pembelajaran Mesin**

- Subbidang dalam kecerdasan buatan yang berfokus pada pengembangan algoritma dan model komputer yang dapat belajar dari data dan membuat prediksi atau pengambilan keputusan tanpa perlu diprogram secara eksplisit.
- Pembelajaran mesin memungkinkan sistem untuk mengekstraksi pola dan pengetahuan dari data yang ada, dan dengan demikian, memungkinkan sistem untuk beradaptasi dengan data baru dan melakukan tugas yang lebih kompleks.

# Konsep Utama dalam Pembelajaran Mesin

# Konsep Utama dalam Pembelajaran Mesin

- **Data dan Dataset:**

- Data adalah bahan mentah dalam pembelajaran mesin. Dataset adalah kumpulan data yang digunakan untuk melatih dan menguji model pembelajaran mesin. Data dalam dataset dapat berupa atribut atau fitur yang menggambarkan entitas atau objek tertentu, seperti gambar, teks, suara, atau data numerik.

- **Latihan (Training):**

- Latihan adalah proses mengajarkan model pembelajaran mesin dengan menggunakan dataset yang telah diberikan. Model belajar dari data dan mencoba mengenali pola atau hubungan dalam dataset tersebut. Selama latihan, model disesuaikan dengan data sehingga dapat membuat prediksi yang lebih baik.

# Konsep Utama dalam Pembelajaran Mesin

- **Model:**

- Model adalah representasi matematis dari algoritma pembelajaran mesin. Ini adalah entitas yang belajar dari data selama latihan dan kemudian digunakan untuk membuat prediksi atau pengambilan keputusan. Model dapat berupa jaringan saraf tiruan, pohon keputusan, regresi linear, atau jenis lainnya tergantung pada tugas yang ingin diselesaikan.

- **Fitur (Features):**

- Fitur adalah atribut atau karakteristik dari data yang digunakan oleh model untuk membuat prediksi. Pemilihan fitur yang tepat sangat penting dalam pembelajaran mesin karena dapat memengaruhi kualitas model.

- **Label:**

- Label adalah nilai yang ingin diprediksi oleh model dalam tugas pembelajaran yang diawasi (supervised learning). Label ini digunakan selama latihan untuk mengajar model bagaimana membedakan dan mengklasifikasikan data.

# Konsep Utama dalam Pembelajaran Mesin

- **Pelabelan (Labeling):**

- Proses pelabelan melibatkan pemberian label atau kategori pada data yang ada. Ini adalah langkah yang penting dalam tugas pembelajaran yang diawasi, di mana model perlu mempelajari hubungan antara fitur dan label.

- **Pengujian (Testing) dan Validasi (Validation):**

- Setelah model dilatih dengan dataset, kemudian diuji dan divalidasi menggunakan data yang belum pernah dilihat sebelumnya. Pengujian digunakan untuk mengukur sejauh mana model mampu membuat prediksi yang akurat.

- **Pembelajaran yang Diawasi (Supervised Learning):**

- Ini adalah jenis pembelajaran mesin di mana model belajar dari dataset yang memiliki label atau keluaran yang benar. Tujuannya adalah memprediksi label yang benar untuk data baru.



# Konsep Utama dalam Pembelajaran Mesin

- **Pembelajaran yang Tidak Diawasi (Unsupervised Learning):**

- Ini adalah jenis pembelajaran mesin di mana model berusaha menemukan pola atau struktur dalam data tanpa label. Ini digunakan untuk tugas seperti pengelompokan (clustering) dan reduksi dimensi.

- **Pembelajaran yang Diperkuat (Reinforcement Learning):**

- Dalam pembelajaran yang diperkuat, model belajar melalui interaksi dengan lingkungan. Ini digunakan dalam tugas yang melibatkan pengambilan keputusan dan interaksi dinamis.

- **Overfitting dan Underfitting:**

- Overfitting terjadi ketika model terlalu rumit dan "menghafal" data pelatihan, tetapi tidak dapat menggeneralisasi dengan baik ke data baru. Underfitting terjadi ketika model terlalu sederhana dan tidak mampu mengambil pola dalam data pelatihan.

# Konsep Utama dalam Pembelajaran Mesin

- **Algoritma Pembelajaran Mesin:**

- Ada berbagai algoritma pembelajaran mesin yang digunakan untuk berbagai tugas, seperti regresi, klasifikasi, pengelompokan, dan lainnya. Contoh algoritma termasuk regresi linear, Naive Bayes, k-Nearest Neighbors (k-NN), pohon keputusan, dan jaringan saraf tiruan (neural networks).

- **Optimisasi Model:**

- Optimisasi model melibatkan penyetelan/konfigurasi parameter model untuk meningkatkan kinerja. Ini melibatkan teknik seperti validasi silang dan penalaan hiperparameter.

- **Evaluasi Model:**

- Evaluasi model adalah proses mengukur seberapa baik model bekerja dalam tugas tertentu. Ini melibatkan metrik seperti akurasi, presisi, recall, F1-score, dan lainnya, tergantung pada tugas.

- **Penerapan:**

- Setelah model dilatih dan dievaluasi, model tersebut dapat diterapkan dalam situasi dunia nyata untuk membuat prediksi atau pengambilan keputusan.

# Tipe-Tipe Pembelajaran Mesin

# Tipe-Tipe Pembelajaran Mesin

- **Pembelajaran yang Diawasi (Supervised Learning):**

- **Penjelasan:**

- Dalam pembelajaran yang diawasi, model belajar dari dataset yang memiliki label atau keluaran yang benar. Tujuannya adalah memprediksi label yang benar untuk data baru.

- **Contoh:**

- Klasifikasi gambar (seperti mengidentifikasi apakah gambar berisi kucing atau anjing), prediksi harga saham, deteksi spam email.

- **Penerapan:**

- Dalam pemrosesan bahasa alami, model dapat dilatih untuk mengklasifikasikan teks sebagai positif atau negatif (analisis sentimen). Model klasifikasi juga digunakan dalam deteksi penyakit berdasarkan data medis.

# Tipe-Tipe Pembelajaran Mesin

- **Pembelajaran yang Tidak Diawasi (Unsupervised Learning):**

- **Penjelasan:**

- Dalam pembelajaran yang tidak diawasi, model berusaha menemukan pola atau struktur dalam data tanpa label. Ini digunakan untuk tugas seperti pengelompokan dan reduksi dimensi.

- **Contoh:**

- Pengelompokan pelanggan berdasarkan perilaku pembelian, reduksi dimensi dalam analisis data.

- **Penerapan:**

- Dalam ilmu biologi, analisis ekspresi gen, tugas pengelompokan digunakan untuk mengelompokkan organisme berdasarkan karakteristik genetik mereka.

# Tipe-Tipe Pembelajaran Mesin

- **Pembelajaran yang Diperkuat (Reinforcement Learning):**

- **Penjelasan:**

- Dalam pembelajaran yang diperkuat, model belajar melalui interaksi dengan lingkungan. Ini digunakan dalam tugas yang melibatkan pengambilan keputusan dan interaksi dinamis.

- **Contoh:**

- Kendaraan otonom, permainan komputer (seperti permainan catur atau Go).

- **Penerapan:**

- Aplikasi kendaraan otonom menggunakan reinforcement learning untuk mengajarkan mobil bagaimana berperilaku dalam berbagai situasi lalu lintas.

# Tipe-Tipe Pembelajaran Mesin

- **Pembelajaran Semi-Supervised (Semi-Supervised Learning):**

- **Penjelasan:**

- Pembelajaran semi-terawasi adalah campuran antara pembelajaran terawasi dan tidak terawasi. Model dilatih dengan dataset yang sebagian besar tidak berlabel, tetapi juga mencakup beberapa data berlabel.

- **Contoh:**

- Klasifikasi dokumen di mana sebagian besar dokumen tidak memiliki label, tetapi beberapa di antaranya telah diberi label.

- **Penerapan:**

- Dalam pemrosesan bahasa alami, pemodelan bahasa menggunakan pembelajaran semi-terawasi untuk memahami makna kata dalam konteks.

# Tipe-Tipe Pembelajaran Mesin

- **Pembelajaran Penguatan (Self-Supervised Learning):**

- **Penjelasan:**

- Pembelajaran penguatan adalah tipe khusus dari pembelajaran mesin yang menggunakan informasi dari data itu sendiri untuk membuat label atau target pembelajaran.

- **Contoh:**

- Pretrained language models seperti BERT, GPT-3, dan tugas fill-in-the-blank.

- **Penerapan:**

- Dalam pemrosesan bahasa alami, model dapat dilatih untuk memahami konteks dalam teks tanpa label, yang dapat digunakan dalam tugas lain seperti analisis sentimen atau pemodelan bahasa.



# Tipe-Tipe Pembelajaran Mesin

- **Pembelajaran Transfer (Transfer Learning):**

- **Penjelasan:**

- Pembelajaran transfer melibatkan menggunakan model yang telah dilatih pada satu tugas untuk tugas lain yang mirip. Model tersebut kemudian disesuaikan ulang dengan dataset yang relevan.

- **Contoh:**

- Menggunakan model yang dilatih pada pengenalan gambar untuk tugas klasifikasi teks.

- **Penerapan:**

- Transfer learning banyak digunakan dalam pemrosesan bahasa alami, pengenalan wajah, dan berbagai aplikasi visi komputer.

- **Pembelajaran Penguatan (Ensemble Learning):**

- **Penjelasan:**

- Ensemble learning melibatkan penggabungan hasil beberapa model untuk meningkatkan kinerja prediksi. Ini dapat mencakup penggabungan model klasifikasi atau regresi.

- **Contoh:**

- Random Forest, Gradient Boosting, klasifikasi mayoritas.

- **Penerapan:**

- Ensemble learning digunakan dalam berbagai aplikasi, termasuk dalam tugas pengenalan wajah, pengenalan suara, dan pengenalan tulisan tangan.

Oemodelan Data dalam Pembelajaran Mesin

# Pemodelan Data

- **Pemodelan Data** dalam pembelajaran mesin adalah proses yang mencakup persiapan data, pemahaman data, pemilihan fitur, transformasi data, dan pembagian data menjadi subset latihan dan pengujian.
- Pemodelan data adalah langkah penting dalam siklus pembelajaran mesin dan memiliki dampak besar pada kualitas model yang dihasilkan.

# Langkah-Langkada dalam Pemodelan Data

- **Pengumpulan Data:**

- Tahap pertama dalam pemodelan data adalah mengumpulkan data yang relevan untuk masalah yang ingin diselesaikan. Data dapat berupa teks, gambar, suara, atau data numerik tergantung pada tugas.

- **Pemahaman Data:**

- Setelah data terkumpul, langkah berikutnya adalah memahami data dengan melakukan analisis eksploratif. Ini termasuk memeriksa statistik deskriptif, visualisasi data, dan pemahaman tentang karakteristik data, seperti apakah data mengandung outlier atau nilai yang hilang.

- **Pembersihan Data:**

- Data sering kali memerlukan pembersihan untuk mengatasi masalah seperti nilai yang hilang, outlier, atau inkonsistensi. Pembersihan data melibatkan pengisian nilai yang hilang, mengatasi outlier, dan menormalisasi data.

# Langkah-Langkada dalam Pemodelan Data

- **Pemilihan Fitur (Feature Selection):**

- Pemilihan fitur melibatkan pemilihan atribut atau fitur yang paling relevan untuk tugas yang ingin diselesaikan. Ini dapat mengurangi dimensi data dan meningkatkan kinerja model.

- **Transformasi Data:**

- Data mungkin perlu diubah atau diproses untuk meningkatkan pemahaman model. Transformasi data dapat mencakup normalisasi, pengkodean variabel kategori, ekstraksi fitur, atau transformasi skala.

- **Pembagian Data:**

- Data dibagi menjadi subset latihan dan pengujian untuk menguji kinerja model. Pembagian data umumnya dilakukan dengan metode acak. Anda dapat membagi data menjadi 70% untuk pelatihan dan 30% untuk pengujian, atau dengan rasio lain tergantung pada ukuran dataset.

# Langkah-Langkada dalam Pemodelan Data

- **Pengkodean Label (Label Encoding):**

- Jika Anda memiliki data kategori, Anda perlu mengkodekan label menjadi nilai numerik. Ini diperlukan karena sebagian besar algoritma pembelajaran mesin hanya dapat mengoperasikan data numerik.

- **Pengisian Data yang Hilang:**

- Jika ada nilai yang hilang dalam data, Anda perlu memutuskan cara mengisinya. Ini bisa melibatkan pengisian dengan nilai rata-rata, median, atau teknik imputasi lainnya.

- **Transformasi Skala (Scaling):**

- Beberapa algoritma pembelajaran mesin, seperti regresi logistik atau analisis komponen utama, memerlukan data yang telah diubah skala. Transformasi skala memastikan bahwa semua fitur memiliki pengaruh yang seimbang pada model.

# Algoritma Klasifikasi pada Pembelajaran Mesin

# Algoritma Klasifikasi

- **Algoritma Klasifikasi** adalah algoritma dalam pembelajaran mesin yang digunakan untuk mengklasifikasikan data ke dalam kategori atau kelas yang sudah ditentukan.
- Tujuan utama dari algoritma klasifikasi adalah membuat model yang mampu memprediksi kategori atau kelas yang benar berdasarkan atribut atau fitur data.



# Jenis Algoritma Klasifikasi

- **Regresi Logistik (Logistic Regression):**

- **Penjelasan:**

- Regresi logistik adalah algoritma klasifikasi yang digunakan untuk memprediksi probabilitas bahwa suatu entitas akan termasuk dalam salah satu dari dua kategori atau lebih (binary atau multiclass classification).

- **Contoh:**

- Prediksi apakah pelanggan akan membeli produk (Ya/Tidak), klasifikasi email sebagai spam atau bukan (binary classification).

- **Penerapan:**

- Regresi logistik banyak digunakan dalam pemasaran, analisis risiko kredit, dan banyak tugas klasifikasi lainnya.

# Jenis Algoritma Klasifikasi

- **Mesin Vector Pendukung (Support Vector Machine - SVM):**

- **Penjelasan:**

- SVM adalah algoritma klasifikasi yang mencari hyperplane terbaik untuk memisahkan data ke dalam kelas-kelas yang berbeda. Ini berusaha untuk mencapai pemisahan maksimum antara kelas-kelas tersebut.

- **Contoh:**

- Klasifikasi gambar berdasarkan fitur visual, seperti dalam pengenalan wajah atau deteksi objek.

- **Penerapan:**

- SVM digunakan dalam berbagai aplikasi, termasuk visi komputer, bioinformatika, dan pemrosesan teks.

# Jenis Algoritma Klasifikasi

- **Pohon Keputusan (Decision Tree):**

- **Penjelasan:**

- Pohon keputusan adalah algoritma klasifikasi yang menggunakan struktur berhirarki dari pohon untuk membuat keputusan klasifikasi berdasarkan fitur-fitur data.

- **Contoh:**

- Pengklasifikasian pelanggan berdasarkan atribut demografis seperti usia, pendapatan, dan lokasi geografis.

- **Penerapan:**

- Pohon keputusan digunakan dalam pengambilan keputusan bisnis, analisis data, dan di berbagai industri.

# Jenis Algoritma Klasifikasi

- **K-Nearest Neighbors (K-NN):**

- **Penjelasan:**

- K-NN adalah algoritma klasifikasi yang mengklasifikasikan data berdasarkan mayoritas kelas-kelas tetangga terdekatnya dalam ruang fitur.

- **Contoh:**

- Klasifikasi gambar berdasarkan atribut visual seperti warna, tekstur, atau bentuk.

- **Penerapan:**

- K-NN digunakan dalam pengenalan pola, klasifikasi teks, pengenalan wajah, dan pengenalan tulisan tangan.

# Jenis Algoritma Klasifikasi

- **Naive Bayes:**

- **Penjelasan:**

- Naive Bayes adalah algoritma klasifikasi yang didasarkan pada teorema Bayes dan menganggap bahwa atribut-atribut dalam data adalah independen satu sama lain (meskipun ini adalah asumsi sederhana).

- **Contoh:**

- Klasifikasi email sebagai spam atau bukan berdasarkan kata-kata yang muncul dalam pesan.

- **Penerapan:**

- Naive Bayes digunakan dalam pengenalan teks, analisis sentimen, dan berbagai aplikasi klasifikasi teks.

# Jenis Algoritma Klasifikasi

- **Random Forest:**

- **Penjelasan:**

- Random Forest adalah algoritma ensemble yang menggunakan sekumpulan pohon keputusan untuk klasifikasi. Ini memungkinkan pengambilan keputusan berdasarkan mayoritas suara dari pohon-pohon tersebut.

- **Contoh:**

- Pengklasifikasian spesies tanaman berdasarkan atribut botani seperti tinggi, lebar daun, dan tekstur batang.

- **Penerapan:**

- Random Forest digunakan dalam berbagai aplikasi seperti klasifikasi gambar, klasifikasi genetik, dan pemrosesan teks.

# Jenis Algoritma Klasifikasi

- **Jaringan Saraf Tiruan (Artificial Neural Networks - ANN):**

- **Penjelasan:**

- ANN adalah algoritma klasifikasi yang terinspirasi oleh struktur jaringan saraf manusia. Ini terdiri dari lapisan-lapisan neuron buatan yang melakukan pemrosesan data.

- **Contoh:**

- Pengenalan tulisan tangan, klasifikasi gambar, dan tugas pemrosesan bahasa alami.

- **Penerapan:**

- ANN digunakan dalam berbagai aplikasi yang melibatkan pemahaman data kompleks, termasuk pengenalan wajah dan bahasa alami.

# Algoritma Regresi pada Pembelajaran Mesin



# Algoritma Regresi

- **Regresi** dalam pembelajaran mesin adalah jenis algoritma yang digunakan untuk memodelkan hubungan antara variabel independen (fitur atau atribut) dan variabel dependen (target) dalam bentuk fungsi matematis.
- Tujuan utama regresi adalah untuk memahami dan memprediksi nilai numerik kontinu berdasarkan fitur-fitur yang ada.

# Jenis Algoritma Regresi

- **Regresi Linier (Linear Regression):**

- **Penjelasan:**

- Regresi linier adalah salah satu algoritma regresi paling sederhana yang digunakan untuk memodelkan hubungan linier antara variabel independen dan dependen. Dalam regresi linier sederhana, hanya ada satu variabel independen, sementara dalam regresi linier berganda, terdapat beberapa variabel independen.

- **Contoh:**

- Prediksi harga rumah berdasarkan luas tanah, jumlah kamar, dan usia bangunan.

- **Penerapan:**

- Regresi linier digunakan dalam berbagai aplikasi, seperti analisis ekonomi, ilmu sosial, dan ilmu keuangan.

# Jenis Algoritma Regresi

- **Regresi Logistik (Logistic Regression):**

- **Penjelasan:**

- Meskipun namanya mencerminkan "regresi," regresi logistik sebenarnya adalah algoritma klasifikasi yang digunakan untuk memprediksi probabilitas bahwa suatu entitas akan termasuk dalam salah satu dari dua kategori atau lebih. Ini memodelkan hubungan logistik antara variabel independen dan probabilitas kategori tertentu.

- **Contoh:**

- Prediksi probabilitas seseorang akan membeli produk (Ya/Tidak) berdasarkan faktor-faktor seperti usia, pendapatan, dan preferensi.

- **Penerapan:**

- Regresi logistik digunakan dalam analisis risiko kredit, pemodelan churn pelanggan, dan banyak tugas klasifikasi lainnya.

# Jenis Algoritma Regresi

- **Regresi Polinomial (Polynomial Regression):**

- **Penjelasan:**

- Regresi polinomial adalah varian dari regresi linier di mana hubungan antara variabel independen dan dependen dimodelkan sebagai polinom. Ini memungkinkan pemodelan hubungan nonlinier antara variabel.

- **Contoh:**

- Prediksi tingkat pertumbuhan pendapatan berdasarkan waktu dengan menggunakan model polinomial.

- **Penerapan:**

- Regresi polinomial berguna dalam pemodelan data yang memiliki tren nonlinier.

# Jenis Algoritma Regresi

- **Regresi Ridge dan Lasso:**

- **Penjelasan:**

- Regresi Ridge dan Lasso adalah teknik regularisasi yang digunakan dalam regresi linier untuk mengatasi masalah overfitting. Regresi Ridge memasukkan norma L2 pada fungsi tujuan, sementara Regresi Lasso memasukkan norma L1.

- **Contoh:**

- Prediksi harga rumah berdasarkan banyak fitur, termasuk beberapa yang mungkin tidak relevan.

- **Penerapan:**

- Regresi Ridge dan Lasso digunakan dalam pemodelan rumah, ekonomi, dan pemrosesan gambar.

# Jenis Algoritma Regresi

- **Regresi Nonparametrik (Nonparametric Regression):**

- **Penjelasan:**

- Regresi nonparametrik tidak mengikuti asumsi tentang bentuk fungsional dari hubungan antara variabel independen dan dependen. Ini memungkinkan pemodelan hubungan yang lebih fleksibel dan nonlinier.

- **Contoh:**

- Prediksi tingkat polusi udara berdasarkan waktu, suhu, dan kelembaban.

- **Penerapan:**

- Regresi nonparametrik digunakan dalam ilmu lingkungan, geokimia, dan pemodelan epidemiologi.

# Jenis Algoritma Regresi

- **Regresi Robust (Robust Regression):**

- **Penjelasan:**

- Regresi robust adalah teknik yang mengatasi data yang mungkin mengandung outlier atau gangguan. Ini memberi lebih sedikit bobot pada observasi yang tidak biasa.

- **Contoh:**

- Prediksi harga rumah dengan data yang mungkin mengandung outlier.

- **Penerapan:**

- Regresi robust digunakan dalam analisis data keuangan, pemodelan harga saham, dan analisis ekonomi.

# Jenis Algoritma Regresi

- **Regresi Bayes (Bayesian Regression):**

- **Penjelasan:**

- Regresi Bayes adalah pendekatan regresi yang menggunakan metode Bayes untuk mengevaluasi dan mengestimasi ketidakpastian dalam model.

- **Contoh:**

- Prediksi harga real estate dengan memperhitungkan ketidakpastian dalam model.

- **Penerapan:**

- Regresi Bayes digunakan dalam ilmu lingkungan, ilmu sosial, dan banyak tugas yang melibatkan estimasi yang tidak pasti.