

Case Studies and Algorithms to Get You Started



Machine Learning

for Hackers

O'REILLY®

*Drew Conway &
John Myles White*

Secara definisi, *machine learning* atau pembelajaran mesin adalah ilmu atau studi yang mempelajari tentang algoritma dan model statistik yang digunakan oleh sistem komputer untuk melakukan *task* tertentu tanpa instruksi eksplisit. *Machine learning* bergantung pada pola dan kesimpulan. Untuk mendapatkan pola dan kesimpulan tersebut, algoritma *machine learning* menghasilkan model matematika yang didasari dari data sampel yang sering disebut dengan '*training data*'.

AI ini mengacu pada prosedur pemrograman komputer (machine) untuk mengambil suatu yang rasional. Apa itu rasional? Rasional adalah dasar dalam mengambil keputusan. Sebagai contoh, AI digunakan untuk memeriksa apakah parameter tertentu dalam sebuah program berperilaku Normal. Misalnya, mesin dapat menimbulkan alarm jika parameter mengatakan 'X' melintasi ambang batas tertentu yang pada gilirannya dapat mempengaruhi hasil proses terkait.

Machine Learning adalah subset dari AI dimana mesin dilatih untuk belajar dari pengalaman masa lalu. Pengalaman masa lalu dikembangkan melalui data yang dikumpulkan, kemudian menggabungkan dengan algoritma (seperti Naïve Bayes, Support Vector Machine (SVM)) untuk memberi hasil akhir.

Statistik adalah cabang matematika yang memanfaatkan data baik dari keseluruhan populasi atau sampel untuk melakukan analisis dan menyajikan kesimpulan. Beberapa teknik statistik yang digunakan adalah regresi, varians, standar deviasi, probabilitas bersyarat dan lainnya.

Mari kita pahami dari contoh berikut. Misalkan, saya perlu memisahkan kiriman di inbox email saya menjadi dua kategori, yaitu 'spam' dan 'penting'. Untuk mengidentifikasi email spam, saya dapat menggunakan algoritma Machine Learning yang dikenal sebagai Naïve Bayes yang akan memeriksa frekuensi kiriman spam masa lalu. Untuk mengidentifikasi email baru sebagai spam, Naïve Bayes menggunakan teori statistik Baye's Theorem (umumnya dikenal sebagai probabilitas bersyarat). Oleh karena itu, kita dapat mengatakan algoritma Machine Learning menggunakan konsep statistik untuk melakukan pembelajaran mesin.

Deep Learning dikaitkan dengan algoritma jaringan saraf tiruan – Artificial Neural Network (ANN) yang menggunakan konsep otak manusia untuk memudahkan pemodelan fungsi yang berubah-ubah. ANN membutuhkan sejumlah besar data dan

algoritma ini sangat fleksibel dalam hal menghasilkan banyak keluaran secara bersamaan. Baca artikel mengenai deep learning!

Data Mining digunakan untuk mencari informasi yang spesifik, sedangkan Machine Learning berkonsentrasi untuk melakukan tugas tertentu. Sebagai contoh untuk membantu perbedaan antara Machine Learning dan Data Mining, mengajar seorang cara menari adalah Machine Learning, sedangkan menggunakan seseorang untuk mencari pusat tarian terbaik di kota adalah Data Mining.

Machine Learning melibatkan proses struktural dimana setiap tahap membangun versi mesin yang lebih baik. Untuk penyederhanaan, proses Machine Learning bisa dibagi menjadi 3 bagian: Data as input, Abstracting the data, Generalization.

Langkah-langkah yang digunakan dalam Machine Learning yaitu: Mengumpulkan data, Mempersiapkan data, Melatih sebuah model, Mengevaluasi model, Meningkatkan kinerja. Mengumpulkan data, data mentah bisa berupa Excel, Ms Access, file teks dan lain-lain. Langkah ini membentuk dasar pembelajaran masa depan. Semakin banyak variasi, kepadatan dan volume data yang relevan, semakin baik prospek pembelajaran untuk mesin. Mempersiapkan data, setiap proses analitis berkembang dengan kualitas data yang digunakan. Kita perlu meluangkan waktu untuk menentukan kualitas data dan kemudian mengambil langkah-langkah untuk memperbaiki masalah seperti kehilangan data dan lainnya. Melatih sebuah model, langkah ini melibatkan pemilihan algoritma dan representasi data yang tepat dalam bentuk model. Data yang disiapkan dibagi menjadi dua bagian: train dan test. Bagian pertama (training data) digunakan untuk pengembangan model. Bagian kedua (data test), digunakan sebagai referensi. Mengevaluasi model, untuk menguji keakuratan, bagian kedua dari data (data test) digunakan. Langkah ini menentukan ketepatan dalam pemilihan algoritma berdasarkan hasil pengujian. Pengujian yang lebih baik untuk memeriksa ketepatan model adalah dengan melihat kinerjanya pada data yang tidak digunakan sama sekali selama pembuatan model. Meningkatkan kinerja, langkah ini mungkin melibatkan pemilihan model yang berbeda atau memperkenalkan lebih banyak variabel untuk meningkatkan efisiensi. Itulah sebabnya dibutuhkan banyak waktu untuk pengumpulan data dan persiapan data.

Jenis Algoritma Machine Learning terdiri dari supervised learning, unsupervised learning, dan reinforcement learning.

Model Supervised Learning / Predictive, model ini digunakan untuk memprediksi hasil masa depan berdasarkan data historis. Model prediktif biasanya diberi instruksi yang jelas sejak awal seperti apa yang perlu dipelajari dan bagaimana itu perlu dipelajari. Algoritma pembelajaran ini disebut Supervised Learning. Sebagai contoh: Supervised Learning digunakan saat perusahaan pemasaran mencoba untuk mengetahui pelanggan mana yang cenderung berpindah atau mencari supplier lain. Algoritma ini juga bisa digunakan untuk memprediksi kemungkinan terjadinya bahaya seperti gempa bumi, tornado dan lain-lain, dengan tujuan untuk mengetahui Total Nilai Asuransi. Beberapa contoh algoritma yang digunakan adalah: Nearest Neighbour, Naïve Bayes, Decision Tree, Regression, dan lain-lain.

Model Unsupervised Learning/Descriptive, model ini digunakan untuk melatih dimana tidak ada target yang ditetapkan dan tidak ada faktor yang penting dari yang lainnya. Sebagai contoh penggunaan model unsupervised learning ini, bila seorang penjual pengecer ingin mengetahui kombinasi produk apa yang cenderung lebih sering dibeli konsumen. Di industri farmasi, digunakan untuk memprediksi penyakit mana yang mungkin terjadi bersamaan dengan diabetes. Contoh algoritma yang digunakan di model ini: K-Means Clustering Algorithm.

Reinforcement Learning (RL), model ini adalah contoh pembelajaran mesin dimana mesin dilatih untuk mengambil keputusan spesifik berdasarkan kebutuhan bisnis dengan tujuan utama untuk memaksimalkan efisiensi (kinerja). Ide dari Reinforcement learning ini adalah mesin/perangkat lunak melatih dirinya secara terus menerus berdasarkan lingkungan yang dipengaruhinya, dan menerapkan pengetahuan yang diperkaya untuk memecahkan masalah bisnis. Proses belajar yang terus-menerus ini memastikan lebih sedikit keterlibatan manusia sehingga akan banyak menghemat waktu. Contoh algoritma yang digunakan dalam RL adalah Markov Decision Process. Untuk membedakan antara Supervised Learning dan Reinforcement Learning, dapat dicontohkan, sebuah mobil menggunakan Reinforcement learning untuk membuat keputusan rute mana yang harus ditempuh, kecepatan berapa yang harus dikemudikan, dimana beberapa pertanyaan tersebut diputuskan setelah berinteraksi dengan lingkungan. Sedangkan memperkirakan ongkos taksi dari satu tempat ke tempat lain adalah Supervised Learning.

Google dan Facebook adalah dua contoh perusahaan yang menggunakan Machine Learning secara ekstensif untuk mendorong iklan masing-masing ke pengguna yang relevan.

Pada layanan perbankan & keuangan, Machine Learning dapat digunakan untuk memprediksi pelanggan yang cenderung gagal membayar pinjaman atau tagihan kartu kredit. Ini sangat penting karena Machine Learning akan membantu bank untuk mengidentifikasi nasabah yang dapat diberikan pinjaman dan kartu kredit.

Kesehatan digunakan untuk mendiagnosis penyakit mematikan (misalnya kanker) berdasarkan gejala pasien dan menghitungnya dengan data terakhir dari jenis pasien yang sama.

Ritel digunakan untuk mengidentifikasi produk yang lebih sering dijual (bergerak cepat) dan produk yang lambat. Hal ini membantu memutuskan jenis produk yang akan ditampilkan atau dikeluarkan dari rak. Selain itu, algoritma Machine Learning dapat digunakan untuk menemukan dua atau lebih produk yang dijual bersama. Hal ini dilakukan untuk merangsang inisiatif loyalitas pelanggan yang pada gilirannya membantu para peritel untuk mengembangkan pelanggan setia.