OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 2, Februari 2024
ISSN 2828-2442 (media online)
Hal 375-380

Pembelajaran *Machine Learning*

Agung Wijoyo[1*], Asep Yudistira Saputra[1], Safitri Ristanti[1], Sultan Rafly Sya'Ban[1], Mila Amalia[1], Randi Febriansyah[1]

[1]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia
Email: [1*]dosen01671@unpam.ac.id, [2]a.yudistirasaputra98@gmail.com, [3]safitristantii@gmail.com, [4]sultanraflys15@gmail.com, [5]milaamelia.119@gmail.com, [6]loonatic1998@gmail.com
(* : coressponding author)

Abstrak−Machine learning adalah sistem yang dapat belajar membuat keputusan sendiri tanpa harus diprogram berulang kali oleh manusia sehingga komputer dapat menjadi lebih pintar dan belajar dari pengalamannya dengan data. Berdasarkan teknik pembelajarannya, supervised learning dapat dibedakan dengan menggunakan dataset berlabel (data training), sedangkan unsupervised learning menarik kesimpulan berdasarkan dataset. Input berupa dataset digunakan oleh machine learning untuk menghasilkan analisis yang tepat. Solusinya menggunakan Python yang menyediakan algoritme dan pustaka yang digunakan untuk membuat machine learning. Artificial intelligence (AI) kini bangkit kembali setelah puluhan tahun mengalami pasang surut. Kecerdasan buatan kembali populer dimana penerapannya dilakukan secara masive di aplikasi bisnis dan media sosial masa kini seperti Facebook, Twitter, Google, Amazon, bahkan berbagai aplikasi besar asal Indonesia seperti Go-jek, Tokopedia, dan lain sebagainya. Struktur pembahasan dalam buku ini meliputi 3 bagian besar, yaitu (1) Konsep Machine Learning dan Kecerdasan Buatan (2) Dasar-Dasar Pemrograman Python untuk Machine Learning serta (3) Contoh Penerapan Machine Learning Menggunakan Python dengan mengimplementasikan beberapa algoritma, baik Pembelajaran Supervised maupun Unsupervised Learning. Beberapa studi kasus dibahas secara lengkap mulai dari pemahaman algoritma, pengolahan dataset hingga training dan testing serta visualisasi hasil model machine learning yang dikembangkan.

Kata kunci: *Machine Learning*, Algoritma, *Supervised Learning, Unsupervised Learning*

Abstract− *Machine learning is a system that can learn to make decisions on its own without having to be programmed repeatedly by humans so that computers can become smarter and learn from their experiences with data. Based on the learning technique, supervised learning can be distinguished by using labeled datasets (training data), while unsupervised learning draws conclusions based on datasets. Input in the form of a dataset is used by machine learning to produce the right analysis. The solution uses Python which provides the algorithms and libraries used to create machine learning. Artificial intelligence (AI) is now rising again after decades of ups and downs. Artificial intelligence is back in popularity where its application is carried out massively in today's business and social media applications such as Facebook, Twitter, Google, Amazon, and even various large applications from Indonesia such as Go-jek, Tokopedia, and so on. The structure of the discussion in this book includes 3 major sections, namely (1) Concept of Machine Learning and Artificial Intelligence (2) Fundamentals of Python Programming for Machine Learning and (3) Examples of Application of Machine Learning Using Python by implementing several algorithms, both Supervised Learning and Unsupervised Learning. Several case studies are discussed in full starting from understanding algorithms, dataset processing to training and testing as well as visualizing the results of the machine learning models developed.*

Keywords: *Machine Learning, Algorithms, Supervised Learning, Unsupervised Learning.*

1. PENDAHULUAN

Seseorang selama hidupnya tidak pernah berhenti belajar. Ini terjadi secara tidak sadar dan alami. Namun secara mendalam sebuah mesin tentu saja bisa menjadi praktis. Seseorang belajar melalui pengalaman sehari-hari. Dari pengalaman inilah seseorang akan memperoleh pengetahuan. Untuk memperoleh ilmu bisa melalui berbagai cara. Cara paling sederhana adalah belajar menghafal atau menyimpan berita yang diklasifikasikan. Cara lain sudah ada dengan mendapatkan ilmu dari orang lain yang ahli. Seseorang juga dapat belajar melalui pengalaman memecahkan masalah yang dilakukannya. Setelah berhasil mengatasi dilema, seseorang akan mengingat struktur dan cara menyelesaikan masalah tadi. Jika manusia mengalami dilema serupa, maka seseorang dapat menyelesaikan masalah tersebut dengan lebih efisien. Dalam kehidupan sehari-hari, kita dapat melihat bahwa pembelajaran mesin berguna dalam kehidupan sehari-hari. Beberapa model yang terlihat sangat konkret adalah pengenalan suara, pengenalan sidik jari atau pengenalan tulisan

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 2, Februari 2024
ISSN 2828-2442 (media online)
Hal 375-380

tangan. Perkembangan machine learning saat ini sangat pesat. Berbagai penelitian dilakukan untuk membangun mesin yang lebih cerdas. Model terbaru yang bisa kita lihat adalah Asimo, robot cerdas dalam Honda, yang bisa mengenali pemiliknya sekaligus mengenali emosi. Untuk menghasilkan robot secerdas Asimo, beberapa teknik AI diterapkan padanya, seperti pengenalan suara untuk dapat berinteraksi, pengenalan gambar untuk dapat mengenali wajah pemilik dan mengenali ruangan dan masih banyak lagi. Pembelajaran mesin saat ini merupakan cabang ilmu yang populer. Sehingga untuk menghasilkan robot secerdas Asimo, beberapa teknik AI diterapkan padanya, seperti pengenalan suara untuk dapat berinteraksi, pengenalan gambar untuk dapat mengenali wajah pemilik dan mengenali ruangan dan masih banyak lagi. Machine learning saat ini menjadi salah satu cabang ilmu yang populer di media. Diciptakan sebagai cabang asal muasal kecerdasan buatan (artificial intelligence), hampir setiap orang pernah berinteraksi dengan menggunakan atau mendengar sistem komputer pribadi yang dibangun dengan menggunakan teknik pembelajaran mesin. Mulai dari melihat foto hashtag otomatis di Facebook, menggunakan rekomendasi pencarian di Google, mengklik rekomendasi produk homogen dalam belanja online, menikmati layanan email bebas spam, hingga mendengar berita AlphaGo yang mengalahkan pemain profesional top di game Go.

2. RUMUSAN MASALAH

1. Apa itu Pembelajaran Mesin?
2. Bagaimana cara kerja Pembelajaran Mesin?
3. Apa hubungan antara Bahasa Pemrograman Python dan Pembelajaran Mesin?

3. TUJUAN

1. Untuk menjelaskan apa yang dimaksud dengan Machine
2. Untuk menjelaskan cara kerja Machine Learning
3. Untuk menjelaskan hubungan antara Python dan Machine Learning

4. PEMBAHASAN

4.1 Pengertian Pembelajaran *Machine Learning*

Machine Learning atau Pembelajaran Mesin adalah teknik pendekatan dari Artificial Intelligent (AI) yang digunakan untuk meniru untuk menggantikan peran manusia dalam melakukan aktivitas untuk memecahkan masalah. Singkatnya, Machine Learning adalah sebuah mesin yang dibuat untuk dapat belajar dan melakukan pekerjaan tanpa arahan dari penggunanya. Menurut Arthur Samuel, seorang perintis Amerika di bidang permainan komputer dan kecerdasan buatan, AI menyatakan bahwa pembelajaran mesin adalah cabang ilmu yang mempelajari bagaimana memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit. Pembelajaran Mesin dapat melakukan ini jika didasarkan pada ide-ide yang diperoleh dari data sebelumnya dan mengidentifikasi pola dan membuat keputusan menggunakan sedikit intervensi manusia atau penggunanya. Jadi, kemampuan Machine Learning tidak terbatas saat mempelajarinya. Sejak pertama kali dikenal istilah Machine Learning banyak yang mengembangkannya, salah satunya yang paling terkenal adalah Deep Blue pada tahun 1996 yang dibuat oleh IBM, sebuah perusahaan Amerika Serikat yang memproduksi dan menjual hardware dan software komputer. Penggunaan Deep Blue yang paling terkenal adalah kemampuannya bermain catur. Deep Blue sangat populer dengan mengalahkan juara catur profesional. Sementara itu, Tom M. Mitchell mendefinisikan bahwa “Sebuah program komputer yang belajar dari pengalaman (experience) E dari tugas yang diberikan (Task) T dengan kinerja (performance) P-nya yang terukur. Misalnya, sebuah program komputer yang belajar menggunakan catur dapat mengembangkan kemampuannya P, yang diukur sebagai kemampuan untuk menang dalam tugas T bermain catur berdasarkan pengalaman yang diperoleh E dari bermain catur melawan dirinya sendiri.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 2, Februari 2024
ISSN 2828-2442 (media online)
Hal 375-380

4.2 Jenis Pembelajaran *Machine Learning*

Secara garis besar, algoritma machine learning dibagi menjadi tiga jenis, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Sebelum menyelesaikan masalah menggunakan machine learning, kita harus memahami ketiga jenis machine learning tersebut karena algoritma tersebut memiliki fungsi dan tujuannya masing-masing.

4.2.1 *Supervised Learning*

Supervised Learning mengadopsi konsep pendekatan fungsi, dimana pada dasarnya algoritma dilatih sehingga dapat memilih fungsi yang paling menggambarkan input dimana X tertentu membuat estimasi terbaik dari Y. Namun pada kenyataannya, tidak sedikit orang yang kesulitan menemukan fungsi yang paling cocok. Hal ini dikarenakan algoritma sebenarnya bergantung pada asumsi yang digunakan. Jika ada asumsi yang tidak terpenuhi, tidak jarang hasil pengolahan data menimbulkan bias. Oleh karena itu, algoritma ini memerlukan data pelatihan yang benar agar sistem dapat mempelajari pola dan regresi, klasifikasi, K-NN, Naive Bayes, Decision Tree, Regresi Linier, Support Vector Machine, dan jaringan saraf.

Algoritma supervised learning adalah jenis pembelajaran mesin yang paling umum. Dalam bahasa Indonesia, supervised learning diartikan sebagai pembelajaran. diawasi. Istilah "diawasi" muncul karena algoritme ini dirancang untuk belajar dengan contoh. Supervised learning memiliki beberapa kelebihan dan kekurangan. Keuntungan pertama, algoritma supervised learning adalah proses yang sederhana dan mudah dipahami. Selain itu, algoritma ini juga ampuh untuk klasifikasi. Data yang digunakan bukanlah data real time sehingga membutuhkan data baru untuk memprediksi hasil. Kelemahan dari algoritma ini adalah membutuhkan waktu komputasi yang cukup lama untuk pelatihan dan menggunakan algoritma yang lebih kompleks daripada algoritma unsupervised learning karena harus memberi label pada setiap masukan.

4.2.2 *Unsupervised Learning*

Algoritma unsupervised learning adalah algoritma yang tidak memerlukan data berlabel. Dalam pembelajaran tanpa pengawasan, algoritma tidak membutuhkan data pelatihan. Algoritma ini digunakan dalam pendeteksian pola dan pemodelan deskriptif yang tidak memerlukan kategori atau keluaran berlabel untuk membentuk dasar algoritma untuk menemukan model yang tepat. Algoritma ini digunakan untuk clustering dan aturan asosiasi. Keuntungan unsupervised learning adalah karena tidak memerlukan label, algoritme lebih fleksibel untuk mencari pola yang mungkin belum diketahui sebelumnya. Sedangkan kekurangan dari algoritma ini adalah sulitnya mencari informasi di dalam data karena tidak ada label dan lebih sulit membandingkan output dengan input. Setelah memahami supervised dan unsupervised learning, ternyata tidak semua algoritma dapat dikategorikan sebagai supervised atau unsupervised learning. Algoritma ini bisa disebut pembelajaran semi-supervised learning.

4.2.3 Pembelajaran *Semi Supervised* dan *Reinforcement Learning*

Seperti yang telah dijelaskan sebelumnya, semi-supervised learning adalah algoritma yang tidak dapat dimasukkan ke dalam pembelajaran supervised atau unsupervised learning. Algoritma ini cocok untuk sejumlah data besar yang dibagi menjadi dua bagian berlabel dan tidak berlabel. Keunggulan pembelajaran semi-supervised adalah biayanya lebih murah karena hanya separuh data yang diberi label dan tidak memerlukan ahli untuk mengolahnya.

Tujuan dari reinforcement learning adalah menggunakan pengamatan yang dikumpulkan dari interaksi dengan lingkungan untuk mengambil tindakan yang akan memaksimalkan hasil dan meminimalkan risiko. Algoritma ini akan terus belajar secara iteratif. Pada algoritma ini terdapat agen yang akan belajar dari interaksi dengan lingkungannya. Untuk menghasilkan sebuah model, penguatan algoritma pembelajaran melalui beberapa tahapan, diantaranya agen mengamati data masukan, setelah itu agen mengambil tindakan untuk mengambil keputusan. Setelah keputusan diambil, agen akan mendapatkan “reward” atau penguatan dari lingkungan. Kemudian mengamati input kembali dan proses pengambilan keputusan dilakukan kembali namun dengan tambahan penguatan dari lingkungan agar hasil keputusan yang diambil lebih akurat.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 2, Februari 2024
ISSN 2828-2442 (media online)
Hal 375-380

Machine learning dan data science adalah dua ilmu yang saling melengkapi. Kedua ilmu ini saling melengkapi dan biasanya diterapkan secara bersamaan. Data science adalah ilmu yang banyak dipelajari saat ini. Hal ini dikarenakan jumlah data yang terus meningkat dan profesi data scientist semakin diminati. Kedua ilmu ini bisa diterapkan di berbagai sektor, mulai dari keuangan, pendidikan, logistik, komunikasi, bahkan fashion. Kabar baiknya, data science bisa dipelajari oleh siapa saja dengan latar belakang pendidikan apapun.

4.3 Langkah – Langkah Implementasi *Machine Learning*

Adapun beberapa langkah dalam penerapan pembelajaran mesin (Machine Learning) adalah sebagai berikut:

a. Identifikasi Masalah Anda

Tahapan pertama dalam proses Machine Learning adalah harus mampu mendeteksi dan memilah masalah yang ada pada program. Apa permasalahan atau tujuan yang ingin dicapai dengan proses Machine Learning ini, termasuk saat mengumpulkan data. Karena tanpa data, machine learning tidak dapat menyelesaikan masalah. Data tersebut dapat berupa Excel, MS Access, atau file lainnya. Langkah ini merupakan bentuk dasar untuk pembelajaran selanjutnya. Dari soal-soal yang ada, Anda bisa memecahnya menjadi beberapa pertanyaan seperti:

1. Apa tujuanmu? Apa yang ingin Anda prediksi?
2. Fitur (berat) seperti apa yang akan Anda gunakan?
3. Jenis data apa yang ingin Anda masukkan? Apakah data itu tersedia?
4. Masalah apa yang kita hadapi? Klasifikasi Biner? Kekelompokan?
5. Perbaikan seperti apa yang diharapkan?
6. Bagaimana Anda mengukur fitur (bobot) dalam data?

b. Mengumpulkan dan Menyiapkan Data

Langkah selanjutnya dalam proses Machine Learning adalah mengumpulkan dan menyiapkan data. Semakin banyak dan semakin baik kualitas data yang Anda miliki, semakin baik kinerja Machine Learning. Contoh metode dalam pengumpulan data seperti web scraping dan data mining. Namun secara umum tahapan tersebut dapat digambarkan sebagai berikut.

Pertama, data yang akan digunakan "diajarkan" pada tahap pemilahan data dipisahkan menjadi tiga bagian, yaitu data latih (data yang akan dilatih), data validasi (data yang digunakan untuk validasi) dan data uji (data yang digunakan untuk eksperimen prediksi). Hal ini juga menegaskan bahwa Machine Learning adalah sesuatu yang sangat bergantung pada keberadaan data. Dari data ini, kita dapat memutuskan algoritma mana yang akan digunakan pada tahap selanjutnya.

c. Pemilihan Algoritma Untuk Solusi

Langkah selanjutnya adalah membuat model data. Tahap ini melibatkan pemilihan algoritma yang sesuai dan representasi data dalam bentuk model. Data yang telah disiapkan dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Data training digunakan untuk pengembangan model dan data testing digunakan sebagai referensi. Setelah itu dilakukan validasi model yaitu menguji model data dengan validasi data. Hal ini berguna untuk mendapatkan feedback dari input, proses dan output yang digunakan. Idealnya, untuk mendapatkan model terbaik, semua algoritma yang ada pada tools tersebut harus dicoba untuk mendapatkan model terbaik. Namun, hal ini seringkali tidak memungkinkan bahkan untuk Data Scientist yang cukup berpengalaman. Alasannya mungkin karena waktu yang tidak mencukupi atau pengetahuan yang tidak memadai tentang algoritma. Untuk pemula sebaiknya memilih beberapa model atau biasanya satu saja sudah cukup tapi yang sederhana dan mudah dipahami.

d. Evaluasi dan Perbaikan Model

Setelah memiliki model Machine Learning dan model ini dapat digunakan untuk memprediksi data uji, langkah terakhir adalah menganalisis kinerja model dan mengevaluasi model tersebut. Untuk menguji akurasi dan performa model digunakan data testing. Hal ini akan menentukan akurasi dari suatu algoritma yang telah dipilih pada tahap sebelumnya.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 2, Februari 2024
ISSN 2828-2442 (media online)
Hal 375-380

4.3 Menggunakan Python dalam Membuat Model *Machine Learning*

Alasan penggunaan bahasa pemrograman Python dalam pembuatan model di Machine Learning adalah:

a. Mudah Dipelajari

Python menawarkan kode yang pendek dan mudah dibaca. Meskipun algoritma yang rumit dan alur kerja yang fleksibel adalah kekuatan nyata di balik Pembelajaran Mesin dan AI, kesederhanaan Python memungkinkan pengembang dan Ilmuwan Data untuk menulis sistem yang dapat mereka percayai. Data Scientist dapat berfokus pada penyelesaian masalah ML daripada berfokus pada nuansa teknis bahasa pemrograman. Selain itu, Python menarik bagi banyak orang hanya karena mudah dipelajari. Kode python mudah dipahami manusia, yang membuat pembuatan model untuk pembelajaran mesin menjadi sangat mudah.

b. Varietas Library Tinggi dan Beragam

Menjalankan algoritma Machine Learning seringkali merupakan proses yang kompleks dan memakan waktu. Sangat penting untuk memiliki lingkungan digital yang terstruktur dan teruji dengan baik untuk memungkinkan Ilmuwan Data memberikan solusi pengkodean terbaik. Untuk menghemat waktu Ilmuwan Data, pemrogram memilih dari sejumlah kerangka kerja dan pustaka Python. Python memiliki banyak Library yang dapat digunakan untuk membuatnya menjadi mungkin. Ilmuwan Data atau Pengembang untuk menyelesaikan masalah yang sangat rumit dalam hitungan menit atau bahkan detik.

Beberapa contoh Library yang terdapat di dalam Python adalah sebagai berikut:

1. Keras, TensorFlow, dan Scikit-learn untuk pembelajaran mesin
2. NLTK dan spaCy untuk NLP (Pemrosesan bahasa alami)
3. SciPy untuk komputasi tingkat lanjut
4. NumPy untuk komputasi ilmiah dan analisis data berkinerja tinggi
5. OpenCV untuk Computer Vision
6. Panda untuk analisis Data Msnipulation
7. Seaborn dan Matplotloib untuk visualisasi data

c. Komunitas Luas dan Banyak Open Source

Banyaknya pengguna Python membuat Python memiliki komunitas yang luas dan solid. Banyak programmer senior Python yang bersedia membagikan ilmunya melalui berbagai platform, baik itu berupa artikel, video, atau menjawab pertanyaan-pertanyaan teknis yang sering dijumpai di berbagai platform seperti stackoverflow dan lain sebagainya. Kondisi ini membuatnya sangat membantu bagi siapa saja yang baru belajar Python. Kondisi ini juga memungkinkan para pemula python untuk belajar python secara mandiri.

5. KESIMPULAN

Machine Learning atau Pembelajaran Mesin adalah teknik pendekatan dari Artificial Intelligent (AI) yang digunakan untuk meniru untuk menggantikan peran manusia dalam melakukan aktivitas untuk memecahkan masalah. Singkatnya, Machine Learning adalah sebuah mesin yang dibuat untuk dapat belajar dan melakukan pekerjaan tanpa arahan dari penggunanya.

Secara garis besar, algoritma machine learning dibagi menjadi tiga jenis, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Adapun beberapa langkah dalam penerapan pembelajaran mesin (Machine Learning) adalah sebagai berikut:

1. Identifikasi Masalah Anda
2. Kumpulkan dan siapkan data
3. Pemilihan Algoritma Untuk Solusi
4. Evaluasi dan Perbaikan Model

Alasan penggunaan bahasa pemrograman Python dalam pembuatan model di Machine Learning adalah:

1. Mudah Dipelajari
2. Varietas Library Tinggi dan Beragam
3. Komunitas Luas dan Banyak Open Source

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 2, Februari 2024
ISSN 2828-2442 (media online)
Hal 375-380

REFERENCES

Tom M. Mitcell, Machine Learning, The McGraw-Hill, 1997

Sebastian raschka. 2015. Pyhton machine learning

Nurvinda Galuh, Jenis-jenis Machine learning. Gading, Serpong, Tanggerang, Banten. https://www.dqlab.id/contact

Darujati, C., & Gumlear, A. B (2012). Pemanfaatan Teknik Surpervisef Untuk Klasifikasi Teks Bahasa Indonesia. Jurnal Bandung Text Mining, 16(1), 5-1.