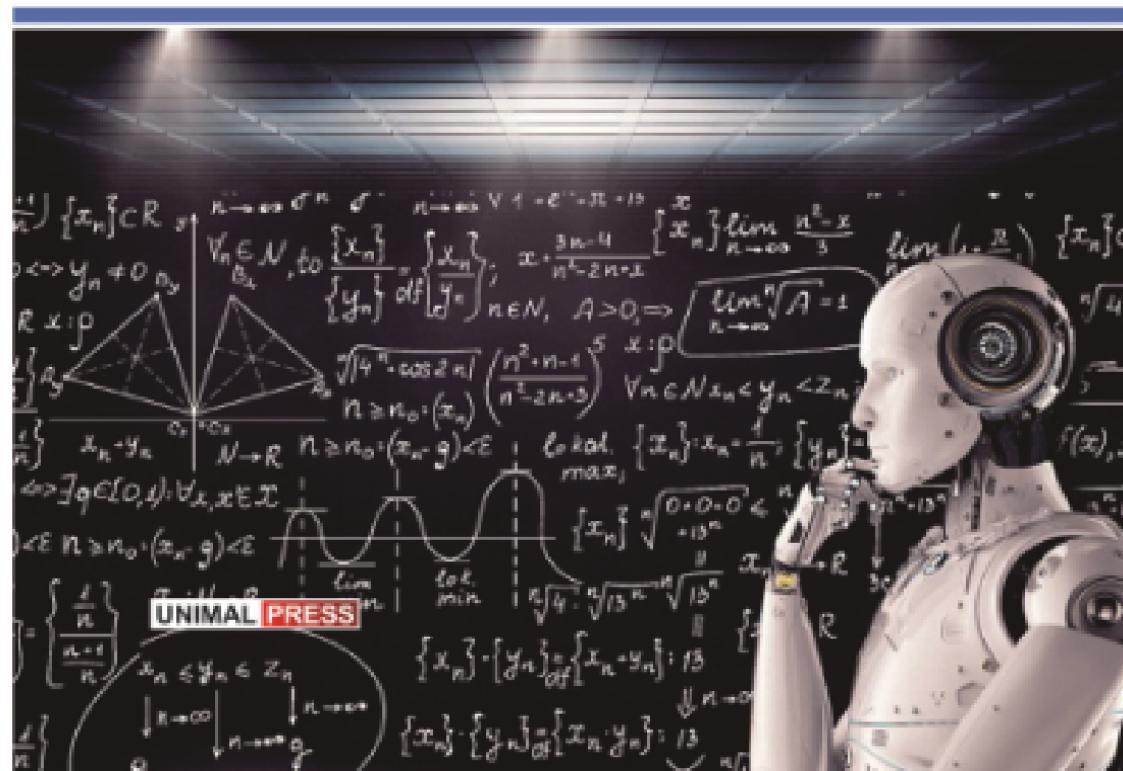


MACHINE LEARNING

Panduan Memahami Data Science, Supervised Learning,
Unsupervised Learning dan Reinforcement Learning



MACHINE LEARNING



universitas
MALIKUSSALEH

**ROZZI KESUMA DINATA
NOVIA HASDYNA**

MACHINE LEARNING

Editor:
DR. FAJRIANA, S.Si., M.Si.

UNIMAL PRESS

Judul: **MACHINE LEARNING**

x + 156 hal., 15 cm x 23 cm

Cetakan Pertama: Juli, 2020

Hak Cipta © dilindungi Undang-undang. All Rights Reserved

Penulis:

ROZZI KESUMA DINATA

NOVIA HASDYNA

Editor:

DR. FAJRIANA, S.Si., M.Si.

Penata Letak: **Eriyanto**

Pracetak dan Produksi: **Unimal Press**

Penerbit:

ISBN 978-602-464-096-5

UNIMAL PRESS

Unimal Press

Jl. Sulawesi No.1-2

Kampus Bukit Indah Lhokseumawe 24351

PO.Box. 141. Telp. 0645-41373. Fax. 0645-44450

Laman: www.unimal.ac.id/unimalpress.

Email: unimalpress@gmail.com



ISBN:

978 – 602 –464 – 096-5

Dilarang keras memfotocopy atau memperbanyak sebahagian atau seluruh buku ini tanpa seizin tertulis dari Penerbit

Kata Pengantar

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Penyayang, kita panjatkan puji syukur atas kehadiran-Nya, yang telah melimpahkan rahmat, hidayah, dan inayah-Nya kepada kita, sehingga penulis dapat menyelesaikan buku yang berjudul "**Machine Learning**". Shalawat beserta salam kepada Baginda Rasulullah Muhammad, SAW yang telah membawa pedoman kepada sahabat serta seluruh ummat.

Buku ini dimaksudkan untuk diajarkan baik di tingkat S1 maupun di tingkat S2. Buku ini membahas tentang teori, konsep dan algoritma pembelajaran mesin. *Machine learning* atau pembelajaran mesin terdiri dari: *supervised learning*, *unsupervised learning* dan *reinforcement learning*.

Untuk membantu pembaca memahami isi buku dengan lebih baik, diakhir setiap bab disajikan istilah-istilah penting dan pertanyaan-pertanyaan yang dibahas di bab bersangkutan. Dengan menjawab pertanyaan-pertanyaan tersebut, maka akan diketahui bagian mana yang sudah dipahami dan yang belum dipahami.

Materi dibuku ini dimaksudkan untuk diberikan selama 14 kali pertemuan tatap muka dengan silabus sebagai berikut ini.

Pertemuan Ke-	Materi
1	Overview dan Penjelasan Silabus
2	Bab 1 - <i>Machine Learning</i>
3	Bab 2 - Algoritma <i>Machine Learning</i>
4	Bab 3 - <i>Linier Regression</i>
5	Bab 4 - <i>K Nearest Neighbor</i>
6	Bab 5 - <i>K Means</i>
7	Bab 6 - <i>Naïve Bayes</i>
8	Bab 7 - <i>K Medoids</i>
9	Bab 8 - <i>Genetic Algorithm</i> Bagian 1
10	Bab 8 - <i>Genetic Algorithm</i> Bagian 2
11	Bab 9 - <i>Support Vector Machine</i>
12	Bab 10 - <i>Fuzzy C -Means</i>
13	Bab 11 - <i>Confusion Matrix</i>
14	Bab 12 - <i>K Fold Cross Validation</i>

Pada kesempatan ini penulis ingin mengucapkan terima kasih sepenuhnya kepada istri tercinta, Novia Hasdyna, S.T., M.Kom yang telah banyak memberikan bantuan berupa ide, masukan dan saran dalam proses pembuatan buku ini. Anak tersayang, Furqan Siddiq Mutawakkil yang selalu mendukung ayahanda dan ibunda. Juga untuk kedua orang tua penulis yang selalu setia mendoakan dan memberikan semangat kepada penulis.

Penulis menyadari bahwa dalam penulisan ini masih mempunyai kekurangan. Kesalahan-kesalahan buku ini sepenuhnya ada di penulis. Untuk perbaikan-perbaikan di edisi-edisi selanjutnya, saran dan kritik membangun dari pembaca sangat diharapkan.

Lhokseumawe, Juli 2020

Penulis.

Daftar Isi

KATA PENGANTAR	V
DAFTAR ISI	VII
DAFTAR TABEL	IX
DAFTAR GAMBAR	X
BAB 1 MACHINE LEARNING.....	1
1.1 PENGANTAR MACHINE LEARNING	1
1.2 PERBEDAAN ANTARA ARTIFICIAL INTELLIGENCE, MACHINE LEARNING , DATA MINING dan DATA SCIENCE.....	1
1.2.1 <i>Artificial Intelligence</i>	2
1.2.2 Machine Learning.....	6
1.2.3 Data Mining.....	7
1.2.4 Data Science	8
BAB 2 ALGORITMA MACHINE LEARNING.....	11
2.1 PENGANTAR ALGORITMA MACHINE LEARNING.....	11
2.2 DATA PREPARATION.....	12
2.3 PEMILIHAN VARIABLE	14
2.4. SUPERVISED LEARNING	14
2.5 UNSUPERVISED LEARNING	15
2.7 DEEP LEARNING	17
2.8 KLASIFIKASI	18
2.9 CLUSTERING	18
BAB 3 LINIER REGRESSION	21
3.1 PENGANTAR LINIER REGRESSION	21
3.2 SIMPLE LINIER REGRESSION.....	22
3.3 SOAL DAN PENYELESAIAN ANALISIS REGRESI LINIER SEDERHANA	24
3.4 MULTIPLE LINIER REGRESSION	30
BAB 4 K- NEAREST NEIGHBOR.....	37
4.1 PENGANTAR K-NEAREST NEIGHBOR.....	37
4.2. PROSES KERJA ALGORITMA K-NEAREST NEIGHBOR (KNN).....	38
4.3 FORMULA DISTANCE PADA ALGORITMA K-NEAREST NEIGHBOR (KNN).....	40
4.4 PERHITUNGAN ALGORITMA K-NEAREST NEIGHBOR (KNN)	41
4.4.1 Menghitung parameter k	42
4.4.2 Menghitung Euclidean Distance	42
BAB 5 K- MEANS.....	51
5.1 PENGANTAR K-MEANS.....	51
5.2 PROSES KERJA ALGORITMA K-MEANS	53
5.3 FORMULA DISTANCE PADA ALGORITMA K-MEANS	54
BAB 6 NAÏVE BAYES.....	69

6.1	PENGANTAR NAÏVE BAYES	69
6.2	PROSES KERJA ALGORITMA NAÏVE BAYES.....	71
6.3	PERHITUNGAN ALGORITMA NAÏVE BAYES.....	73
BAB 7	K-MEDOIDS	81
7.1	PENGANTAR K-MEDOIDS.....	81
7.2	PROSES KERJA ALGORITMA K-MEDOIDS	82
7.3	PERHITUNGAN ALGORITMA K-MEDOIDS	83
BAB 8	GENETIC ALGORITHM.....	91
8.1	PENGANTAR ALGORITMA GENETIKA	91
8.2	PROSES KERJA ALGORITMA GENETIKA	93
8.3	DEFINISI PENTING DALAM ALGORITMA GENETIKA	94
8.4	JENIS-JENIS SELECTION	95
8.4.1	Roulette Wheel Selection	95
8.4.2	Rank Based Selection	96
8.4.3	Stochastic Universal Sampling	98
8.4.4	Tournament Selection (Seleksi Turnamen).....	98
8.5	JENIS-JENIS CROSSOVER.....	100
8.5.1	One Point Crossover.....	100
8.5.2	Two-point crossover.....	100
8.5.3	Uniform Crossover.....	100
8.5.4	Partially Mapped Crossover (PMX)	101
8.5.5	Order Crossover (OX)	102
8.6	PERHITUNGAN ALGORITMA GENETIKA.....	103
BAB 9	SUPPORT VECTOR MACHINE.....	113
9.1	PENGANTAR SUPPORT VECTOR MACHINE	113
9.2	PROSES KERJA ALGORITMA SVM	114
9.3	PERHITUNGAN ALGORITMA SVM	117
BAB 10	FUZZY C -MEANS.....	125
10.1	PENGANTAR FUZZY C- MEANS	125
10.2	PROSES KERJA ALGORITMA FUZZY C-MEANS.....	128
10.3	PERHITUNGAN ALGORITMA FUZZY C-MEANS.....	131
BAB 11	CONFUSION MATRIX.....	135
11.1	PENGANTAR CONFUSION MATRIX	135
11.2	PENGUKURAN KINERJA ALGORITMA SUPERVISED LEARNING DENGAN CONFUSION MATRIX.....	136
11.3	PERHITUNGAN CONFUSION MATRIX.....	139
BAB 12	K-FOLD CROSS VALIDATION.....	147
12.1	PENGANTAR K- FOLD CROS VALIDATION.....	147
12.1	CARA KERJA K-FOLD CROSS VALIDATION.....	149
12.3	PERHITUNGAN K-FOLD CROSS VALIDATION.....	149
DAFTAR PUSTAKA.....	155	
BIODATA PENULIS	156	

Daftar Tabel

TABEL 2.1	Dataset Transaksi Penjualan.....	12
TABEL 2.2	Supervised, Unsupervised, Reinforcement Learning Algorithm, Deep Learning.....	15
TABEL 3.1	Hubungan Antara Suhu Ruangan Dengan Jumlah Cacat Produksi.....	25
TABEL 3.2	Hasil Perhitungan X^2 , Y^2 , XY	26
TABEL 3.3	Data Nilai X dan Y.....	28
TABEL 3.4	Hasil Perhitungan X^2 , Y^2 , XY	29
TABEL 3.4	Contoh Soal MLP.....	31
TABEL 4.2	Sampel Data Testing I.....	42
TABEL 4.3	Hasil Perhitungan Jarak Pada Sampel Data Testing Ke-1	45
TABEL 4.4	Hasil Klasifikasi Dataset Testing Ke-1 Dengan K=3.....	46
TABEL 4.5	Data Training Contoh Ii.....	47
TABEL 4.6	Hasil Perhitungan Jarak Z Dengan <i>Euclidean Distance</i>	48
TABEL 11.1	Confusion Matrix.....	140

Daftar Gambar

GAMBAR 1.1	Proses Berfikir dan <i>Reasoning</i>	5
GAMBAR 1.2	Ruang Lingkup Data Science.....	8
GAMBAR 2.1	Ruang Lingkup Algoritma Machine Learning.....	13
GAMBAR 2.2	Supervised Learning.....	14
GAMBAR 2.3	Unsupervised Learning.....	16
GAMBAR 2.4	<i>Reinforcement Learning</i>	16
GAMBAR 2.5	Ai, ML dan DL.....	17
GAMBAR 2.6	Algoritma Deep Learning.....	18
GAMBAR 3.1	Regresi Linear Sederhana.....	22
GAMBAR 4.1	K-Nearest Neighbor.....	38
GAMBAR 4.2	Proses Kerja K-Nearest Neighbor.....	40
GAMBAR 5.1	Ilustrasi K-Means	51
GAMBAR 6.1	Reverend Thomas Bayes.....	69
GAMBAR 6.2	Persamaan Teorema Bayes.....	71
GAMBAR 7.1	K-Medoids Dan K-Means.....	82
GAMBAR 8.1	Siklus Algoritma Genetika	93
GAMBAR 8.2	Individu.....	94
GAMBAR 8.3	Ilustrasi Penyelesaian Permasalahan Dalam Algoritma Genetika	95
GAMBAR 8.4	Roulete Wheel Selection.....	96
GAMBAR 8.5	<i>Rank Based Selectio</i>	97
GAMBAR 8.6	<i>Stochastic Universal Sampling</i>	98
GAMBAR 8.7	<i>Tournament Selection</i>	99
GAMBAR 8.11	Partially Mapped Crossover (PMX)	101
GAMBAR 9.1	Hyperplane Yang Memisahkan Dua Kelas Positif (+1) dan Negatif(-1)	113
GAMBAR 9.2	SVM.....	114
GAMBAR 9.3	Vector SVM.....	115
GAMBAR 9.4	Vektor Positif.....	116
GAMBAR 9.5	Vektor Negatif	116
GAMBAR 9.6	Vektor Antara Batas Keputusan dan Vektor Pendukung.....	117
GAMBAR 9.7	Plot Contoh Data SVM	118
GAMBAR 9.8	Menemukan Hyperplane	120
GAMBAR 9.9	Soft Margin.....	121
GAMBAR 9.10	Klasifikasi And	123
GAMBAR 10.1	Ilustrasi Fuzzy C - Means	125
GAMBAR 10.2	Fuzzy C - Means	126
GAMBAR 11.1	Confusion Matrix.....	136
GAMBAR 11.2	Ilustrasi Confusion Matrix	138
GAMBAR 11.3	False Positive dan False Negatif.....	139
GAMBAR 12.1	K-Fold Cross Validaton.....	148
GAMBAR 12.2	Ilustrasi K Fold Cross Validation.....	148

BAB 1

MACHINE LEARNING

1.1 PENGANTAR MACHINE LEARNING

Machine learning atau dikenal dengan pembelajaran mesin adalah ilmu komputer yang bisa bekerja tanpa diprogram secara eksplisit. Banyak peneliti berpikir bagaimana cara untuk membuat kemajuan menuju AI terhadap tingkat manusia. *Machine learning* ini merupakan kecerdasan buatan yang mempelajari bagaimana membuat data. *Machine learning* ini biasa disingkat dengan ML. Ini dibutuhkan untuk menerapkan teknik yang cepat dan kuat dalam menemukan masalah baru.

Secara definisi, *machine learning* merupakan 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*.’

Pemakaian teknik ini berkaitan dengan pembelajaran mesin dan AI. Mesin ini membuktikan kepada algoritma atau program yang berjalan di komputer. Oleh karena itu, jika kita ingin belajar *machine learning*, pastikan terus berinteraksi dengan data. Semua pengetahuan *machine learning* pasti akan melibatkan data.

1.2 PERBEDAAN ANTARA ARTIFICIAL INTELLIGENCE, MACHINE LEARNING , DATA MINING dan DATA SCIENCE

Setelah penulis memperdalam lebih lanjut masing-masing bidang, istilah-istilah diatas berbeda, namun memang banyak beririsan. Dengan pemahaman yang benar, ketiga ranah tersebut cukup mudah dibedakan.

1.2.1 Artificial Intelligence

Artificial Intelligence atau disingkat AI merupakan sebuah bidang ilmu yang dibangun berdasarkan pemikiran berabad-abad, dan menjadi disiplin ilmu yang terkenal selama lebih dari 50 tahun (Poole & MacWorth (2010). AI telah tumbuh pada peta ilmu pengetahuan dan berada di antara psikologi dan ilmu komputer. Para peneliti terpaut pada tugas membuat komputer melakukan hal-hal yang memerlukan kecerdasan apabila dilakukan oleh manusia. Kecerdasan buatan adalah sebuah bidang ilmu yang mencakup teknik-teknik komputasi untuk melaksanakan tugas-tugas yang memerlukan kecerdasan apabila dilaksanakan oleh manusia.

Isu-isu mendasar dari AI adalah representasi pengetahuan, pencarian, persepsi, dan inferensi. Inferensi adalah proses penciptaan representasi eksplisit pengetahuan dari keadaan implisit. Pengetahuan akan tersedia sebagai kumpulan dari pernyataan logika, heuristik, aturan-aturan, prosedur, korelasi statistik, dan lain sebagainya.

Adapun kelas objek yang dianggap cerdas adalah seperti: semut, burung, kucing, anjing, manusia. Benda-benda yang tidak cerdas adalah: Meja, kursi, komponen-komponen hardware dari komputer.

AI dapat berarti simulasi dari perilaku manusia dan proses kognitif pada computer dan studi tentang keadaan dari seluruh ruang fikiran cerdas.

Berikut analisis logika tentang kecerdasan:

Pernyataan: "Badu anak yang cerdas" memiliki banyak interpretasi:

- Badu tahu banyak hal
- Badu berfikir cepat
- Badu selalu mempertimbangkan akibat yang timbul dari tindakan yang diambilnya.

Seseorang boleh jadi memiliki banyak pengetahuan tetapi mungkin tidak mampu mengorganisasikan pengetahuan ini dengan cara yang kreatif. Beberapa aspek penting kecerdasan manusia:

- ❖ Penggunaan intuisi

- ❖ Akal sehat (*commonsense*)
- ❖ Pertimbangan (*judgement*)
- ❖ Kreativitas (*creativity*)
- ❖ Pengarahan sasaran (*goal directedness*)
- ❖ Alasan yang masuk akal (*plausible reasoning*)
- ❖ Pengetahuan (*knowledge*)
- ❖ Kepercayaan (*beliefs*)

Meskipun kecerdasan manusia itu hebat tetapi memiliki beberapa kekurangan:

- ❖ Intelektual manusia tidak sempurna (*fallible*)
- ❖ Memiliki basis pengetahuan yang terbatas
- ❖ Pemrosesan informasi terhadap sederetan kejadian alam berlangsung sangat lambat di otak apabila dibandingkan dengan computer.

Arti kecerdasan bukan semata-mata kemampuan pemrosesan informasi oleh otak tetapi kemampuan manusia untuk mendemonstrasikan kecerdasannya dengan berkomunikasi secara efektif dan dengan pembelajaran. Adapun pengetahuan dapat diperoleh dari pengalaman dan kemudian didemonstrasikan dengan mengkomunikasikan pengetahuan yang telah diperoleh itu. Komunikasi yang baik memerlukan keahlian untuk menganalisa dan mensintesa pesan-pesan. Untuk mengkomunikasikan sesuatu secara efektif, kita harus mampu menyebarkan pesannya baik melalui surat, lisan atau bentuk komunikasi lain sehingga maksudnya dapat dipahami dengan jelas oleh penerimanya.

Pembelajaran adalah kemampuan belajar atau mengadaptasikan perilaku terhadap situasi yang baru dianggap sebagai komponen kecerdasan yang penting. Mengorganisasikan pengetahuan adalah hal penting dalam proses pembelajaran.

Komponen yang penting dari kecerdasan adalah kemampuan untuk belajar atau beradaptasi pada situasi baru. Adapun langkah-langkah pembelajaran adalah sebagai berikut:

- ❖ Menangkap isu utama dalam konteks yang baru.

- ❖ Mencari unit-unit deskriptif dasar dalam sebuah situasi tertentu
- ❖ Mengorganisasikan pengetahuan adalah hal penting dalam proses pembelajaran.
- ❖ Fakta-fakta haruslah dapat diakses apabila diperlukan.
- ❖ Keahlian (*skill*) harus pula berperan apabila berada pada situasi yang sesuai.
- ❖ Pengetahuan hendaklah terstruktur sedemikian hingga pembelajaran lebih lanjut dapat berlangsung.
- ❖ Pengetahuan juga mengandung sebuah framework dimana berbagai fakta dan aspek-aspek pengalaman dapat disimpan.
- ❖ Pengetahuan yang disimpan dalam framework tersebut mengandung baik fakta-fakta spesifik maupun aturan-aturan umum.

Tujuan terpenting AI adalah untuk menambah pemahaman manusia tentang reasoning, pembelajaran, dan persepsi, yakni untuk membangun perangkat-perangkat suatu pandangan yang lebih matang mengenai kecerdasan manusia daripada yang ada saat ini.

Pentingnya AI menjadi nyata pada akhir 1970-an dimana para pemimpin dunia mengetahui potensi AI dan mendukung program-program riset intensif. Jepang juga intensif melakukan penelitian dalam bidang natural language, memahami speech dan *visual scenes*. Demikian juga *Plan* dari pemerintahan British dinamai *Alvey Project*, diikuti Negara-negara eropa lainnya.

Definisi AI:

Proses Berfikir dan Reasoning

Kinerja Manusia	Rasionalitas
<p>Sistem yang berfikir seperti manusia</p> <p>“Upaya baru yang menarik untuk menjadikan komputer berfikir...mesin dengan fikiran, dalam artian sebenarnya dan literal.” (Haugeland, 1985)</p> <p>“[Pengotomasi dari] aktifitas-aktifitas yang dikaitkan dengan fikiran manusia, aktifitas-aktifitas seperti pengambilan keputusan, pembelajaran, ...” (Bellman, 1978)</p>	<p>Sistem yang berfikir secara rasional</p> <p>“Studi tenang kemampuan mental melalui penggunaan model-model komputasional” (Charniak dan McDermott, 1985)</p> <p>“Studi tentang komputasi yang memungkinkan pengenalan, alasan, dan tindakan” (Winston, 1992)</p>
<p>Sistem yang bertindak seperti manusia</p> <p>“Seni membuat mesin yang melakukan fungsi-fungsi yang memerlukan kecerdasan apabila dilaksanakan oleh manusia” (Kurzweil, 1990)</p> <p>Studi tentang bagaimana membuat komputer melakukan hal-hal yang pada masa itu manusia melakukan dengan lebih baik” (Rich dan Knight, 1991)</p>	<p>Sistem yang bertindak secara rasional</p> <p>“Kecerdasan Komputasional adalah studi tentang perancangan agen-agen cerdas” (Poole et al., 1998)</p> <p>“AI ... bersangkutpadu dengan perilaku cerdas dalam artifak (kerja semi)” Nilsson, 1998)</p>

Beberapa definisi kecerdasan buatan, diorganisasikan atas empat kategori

Perilaku

Gambar 1.1 Proses Berfikir dan *Reasoning*

Definisi AI bervariasi dalam dua dimensi:

- ❖ Bagian atas adalah proses berfikir dan reasoning
- ❖ Bagian bawah adalah perilaku (behavior)
- ❖ Bagian kiri adalah mengukur keberhasilan dalam hal kepatuhan atas kinerja manusia
- ❖ Bagian kanan adalah mengukur konsep ideal kecerdasan (rasionalitas). Sebuah system disebut rasional apabila ia melakukan “hal yang benar” berdasarkan yang diketahuinya.

Secara ringkas, AI atau kecerdasan buatan adalah bidang keilmuan yang mempelajari seputar pembuatan “*Intelligent Agent*” dalam komputer. Ilmu ini mencari tahu bagaimana membuat komputer yang dapat berperilaku seperti makhluk dengan

kecerdasan. Hal tersebut erat kaitannya dengan simulasi proses pengambilan keputusan.

Tentunya perkembangan AI masih sangat jauh dari capaian tersebut. Dengan kemampuan komputasi dan algoritma AI kekinian, hal yang tercapai adalah pembuatan "*Rational Agent*", yaitu komputer dengan kemampuan pertimbangan rasional sehingga dapat mengambil keputusan paling optimal dalam satu kasus tertentu. Misal kompor yang dapat menentukan kapan Mie Instan sudah cukup lama diseduh berdasarkan lama waktu dan suhu pemasakan. Perlu digarisbawahi bahwa bidang tersebut sangatlah luas. Mesin apapun yang tidak melakukan sebuah pekerjaan dengan "bodoh" termasuk dalam kategori "*Intelligent Agent*" dan termasuk dalam bidang keilmuan AI.

1.2.2 Machine Learning

ML adalah bidang keilmuan yang mempelajari bagaimana membuat program yang dapat menghasilkan pengetahuan baru dari pengetahuan yang sudah ada (disebut *experience*, atau data) di luar pengetahuan yang "diprogram" secara langsung pada program. Istilah lebih umumnya adalah bagaimana membuat komputer yang dapat belajar dari lingkungan sekitar sehingga memiliki "pengetahuan" yang berkembang. Contoh paling gampang mungkin adalah prediksi kata yang ada di HP kita atau pengenalan wajah pada *facebook*. Hal tersebut mungkin dilakukan karena program di belakang kedua hal tersebut telah menyusun pengetahuan dari data yang ada, biasanya dalam bentuk sebuah model matematis.

Hal tersebut banyak berhubungan dengan algoritma yang dapat mengekstraksi informasi dari berbagai macam data serta mengenali pola dalam data (*pattern recognition*), sehingga bidang tersebut sangat erat kaitannya dengan statistika. Namun secara garis besar, segala sesuatu yang melibatkan proses induksi pengetahuan dari data termasuk dalam bidang keilmuan ML.

ML ini merupakan bidang keilmuan computer science yang paling sering salah dimengerti. Selain konotasi negatif dalam Indonesia dengan istilah "ML", namun ML tersebut juga merupakan

bidang dengan ranah aplikatif yang sangat luas dalam banyak sekali bidang keilmuan. Hampir seluruh bidang yang berhubungan dengan “komputasi cerdas” membutuhkan adanya pengetahuan dalam sebuah program, dan dalam tahap ini ML sangat berperan. Jika AI fokus pada membuat komputer yang cerdas, maka untuk mencapai kecerdasan tersebut digunakan ML. Bahkan sebagian besar ilmu yang dipelajari dalam AI maupun DM adalah ML itu sendiri. Hal tersebut yang banyak menimbulkan ambiguitas bidang antara ketiga hal tersebut.

1.2.3 Data Mining

Data Mining atau disingkat DM adalah sebuah bidang yang banyak berkembang dari ML, namun berbeda secara tujuan. Jika ML fokus pada membuat program yang dapat belajar, DM fokus pada memanfaatkan program untuk membantu manusia belajar dari data. DM dilakukan oleh seseorang, dalam kasus tertentu, pada dataset tertentu, dengan sebuah tujuan tertentu. Jika kedua bidang sebelumnya fokus pada programnya, DM adalah ilmu praktis yang fokus pada manusianya.

Pada sebagian besar kasus, DM digunakan untuk menghasilkan *“insights”* dari data yang ada, sehingga dapat mendatangkan pengetahuan baru. Hal tersebut lebih banyak digunakan dalam kasus praktis seperti dalam proses pengambilan keputusan sebuah perusahaan. Sebagai contoh, dengan DM, sebuah perusahaan dapat mengetahui informasi detil terkait segmen konsumen yang banyak membeli produk mereka. Hal tersebut kemudian akan digunakan dalam menentukan strategi perusahaan kedepannya.

Istilah DM sendiri sebenarnya adalah sebuah istilah yang sedikit rancu. DM sendiri hanyalah satu langkah dalam proses yang lebih besar yang dikenal sebagai *Knowledge Discovery from Database* (KDD) atau sekarang lebih banyak dikenal dengan istilah *Data Science* (DS).

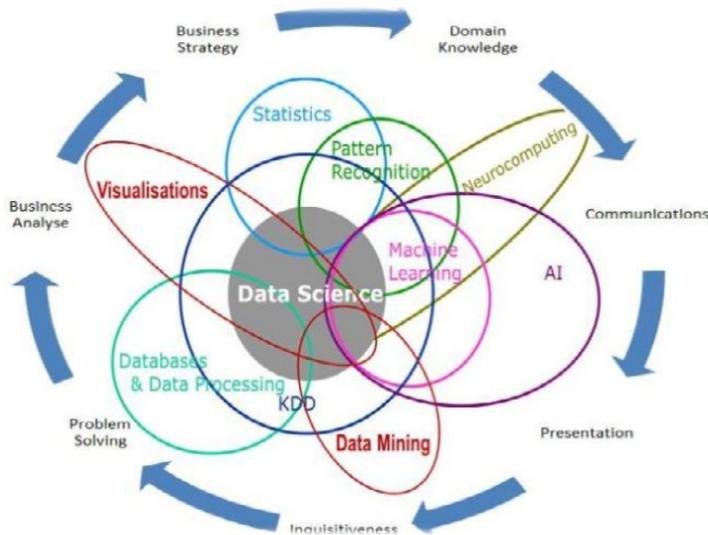
Dengan berkembangnya kemampuan komputasi komputer dan bertambahnya jumlah data di dunia secara eksponensial, sekarang *Data Scientist* menjadi salah satu pekerjaan paling dicari di dunia

karena permintaan yang sangat banyak, namun persediaan tenaga ahli yang sedikit. Dari fenomena tersebut pula muncul dua bidang khusus baru, yaitu Big Data yang berasal dari DM, dan *Deep Learning* yang berasal dari ML.

Secara garis besar ML adalah salah satu ilmu dasar yang banyak digunakan dalam AI maupun DM, yaitu dalam mengolah pengetahuan menggunakan komputer. AI fokus pada membuat komputer dengan kecerdasan, sedangkan DM fokus pada menggunakan komputer untuk membantu manusia memahami data.

1.2.4 Data Science

Data science adalah ilmu yang menggabungkan matematika, statistika dengan ilmu komputer dengan tujuan analisa data (*data analysis*) dari suatu himpunan data baik skala kecil (*sampel*) maupun besar (*populasi*) dengan mengaplikasikan algoritma tertentu untuk tujuan menggali data (*data mining*) dan mendapatkan pola data serta dapat melakukan prediksi data (*prediction*) dengan cukup akurat yang dapat membantu dalam pengambilan keputusan dan dapat digunakan untuk membuat sistem yang cerdas (AI) yang dapat terus belajar dengan sendirinya (*machine learning*).



Gambar 1.2 Ruang Lingkup Data Science

Data Science melibatkan proses berikut :

- ❖ Data Mining adalah proses pengambilan informasi dari pola data dari himpunan data yang sebelumnya tidak diketahui, kadang disebut juga Data Discovery.
- ❖ Data Mining fokus pada mengekstrak pola menggunakan metode statistik untuk dianalisa dan dapat juga melakukan prediksi.
- ❖ Machine learning adalah bidang yang merupakan bagian dari Artificial Intelligence (AI) yang digunakan agar sistem komputer secara otomatis dapat belajar dengan sendirinya tanpa diberi instruksi pemrograman dan dapat meningkatkan prediksi yang akurat dan penggunaannya biasanya sifatnya realtime.

Jadi *Data Mining* dan *Machine Learning* merupakan bagian dari *Data Science*.

NAMA-NAMA dan ISTILAH PENTING

Machine learning

Eksplisit

Task

Artificial Intelligence

Poole & MacWorth

Penggunaan intuisi

Akal sehat (*commonsense*)

Pertimbangan (*judgement*)

Kreativitas (*creativity*)

Pengarahan sasaran (*goal directedness*)

Alasan yang masuk akal (*plausible reasoning*)

Pengetahuan (*knowledge*)

Visual scenes

DAFTAR PUSTAKA

- Gorunescu, F. 2011. Data Mining Concepts, Models and Techniques. Verlah Berlon Heidelberg: Springer.
- Kusuma, P. D. 2020. *Machine Learning Teori, Program, Dan Studi Kasus*. Deepublish.
- Novakovic, Jasmina. 2010. The Impact of Feature Selection on the Accuracy of Naive Bayes Classifier.Vol 2: 1113–16.
- Prasetyo, Eko. 2012. Data Mining Konsep Dan Aplikasi Menggunakan Matlab. Yogyakarta:Andi Offset.
- Putra, J. W. G. 2019. Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. *Tokyo. Jepang*.
- Santoso, B., & Azis, A. I. 2020. *Machine Learning & Reasoning Fuzzy Logic Algoritma, Manual, Matlab, & Rapid Miner*. Deepublish.
- Siringoringo, R. 2018. Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan k-nearest neighbor. Journal Information System Development (ISD),3(1).
- Witten, I.H.; Frank, E.; Hall, M.A. 2011. Data Mining: PracticalMachine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, CA, USA.

--oo0oo--

BIODATA PENULIS



Rozzi Kesuma Dinata, S.T., M.Eng adalah dosen tetap Fakultas Teknik Universitas Malikussaleh. Lahir pada tanggal 29 Agustus 1984. Mengajar di Program Studi Teknik Informatika UNIMAL sejak tahun 2013 sampai sekarang. Sarjana Teknik diperoleh dari FT UNIMAL tahun 2009. Gelar M.Eng diperoleh dari FT UGM (Universitas Gadjah Mada) Program Studi Magister Teknologi Informasi tahun 2012. Aktif sebagai peneliti yang hasilnya dimuat di Jurnal Nasional Terakreditasi.



Novia Hasdyna, S.T., M.Kom.

Lahir pada tanggal 12 November 1992. Lulus S1 di Program Studi Teknik Informatika Universitas Malikussaleh tahun 2014. Gelar M.Kom diperoleh dari FASILKOM- TI USU (Universitas Sumatera Utara) Program Studi Magister Teknik Informatika tahun 2019. Aktif menulis artikel yang dimuat di Jurnal Internasional dan Jurnal Nasional Terakreditasi.

--oo0oo--

MACHINE LEARNING

Machine Learning adalah bidang keilmuan yang mempelajari bagaimana membuat program yang dapat menghasilkan pengetahuan baru dari pengetahuan yang sudah ada (disebut *experience*, atau *data*) di luar pengetahuan yang "diprogram" secara langsung pada program. Istilah lebih umumnya adalah bagaimana membuat komputer yang dapat belajar dari lingkungan sekitar sehingga memiliki "pengetahuan" yang berkembang.

Dengan berkembangnya kemampuan komputasi komputer dan bertambahnya jumlah data di dunia secara eksponensial, sekarang *Data Scientist* menjadi salah satu pekerjaan paling dicari di dunia karena permintaan yang sangat banyak, namun persediaan tenaga ahli yang sedikit. Dari fenomena tersebut pula muncul dua bidang khusus baru, yaitu Big Data yang berasal dari Data Mining, dan *Deep Learning* yang berasal dari *Machine Learning*.

Buku ini dimaksudkan untuk diajarkan baik di tingkat S1 maupun di tingkat S2. Buku ini membahas tentang teori, konsep dan algoritma pembelajaran mesin yang terdiri dari *supervised learning*, *unsupervised learning* dan *reinforcement learning*.

