

## Machine Learning System Design

Mari kita uraikan apa saja materi yang telah dibahas di modul ini.

- **Apa itu Desain Sistem Machine Learning?**

Sebelum memulai proyek machine learning, kita perlu melakukan proses desain sistem machine learning supaya dapat meninjau sebuah sistem secara keseluruhan. Desain sistem machine learning adalah proses menentukan antarmuka, algoritma, data, infrastruktur, dan perangkat keras untuk sistem machine learning guna memenuhi persyaratan (requirements) yang telah ditentukan sebelumnya.

- **Sistem Machine Learning di Dunia Riset/Akademik versus Industri.**

Berdasarkan tujuan dan data yang digunakan, berikut adalah perbedaan sistem machine learning di dunia riset/akademik dan industri.

Di dunia riset/akademik, tujuan utama sistem machine learning kebanyakan adalah untuk mendapatkan akurasi dan performa model yang tinggi, peningkatan kinerja, dan hasil tertinggi atau tahapan terbaru yang biasanya disebut sebagai state-of-the-art. Sedangkan di industri, setiap pemangku kepentingan memiliki tujuan yang berbeda sehingga diperlukan kolaborasi untuk membuat model yang bisa memenuhi semua tujuan tersebut.

Sementara itu, dalam hal penggunaan data, dunia riset/akademik biasanya menggunakan data yang sudah bersih supaya dapat fokus pada pengembangan arsitektur dan proses melatih model. Saat mengembangkan sistem machine learning di industri, biasanya kita masih harus mencari dan mengumpulkan (data collecting) data mentah terlebih dahulu. Sehingga, di industri, diperlukan berbagai proses lebih lanjut seperti data cleansing dan data transformation. Selain itu, data industri biasanya merupakan data tidak terstruktur (unstructured data), tidak seimbang (imbalanced data), dan berubah secara berkala.

- **Persyaratan Machine Learning System Design.**

Sistem machine learning harus mampu menangani perubahan, memudahkan dalam pemeliharaan, dan memudahkan untuk beradaptasi. Oleh karena itu, keterampilan dalam mendesain model machine learning yang dapat memenuhi berbagai persyaratan tersebut, mutlak diperlukan. Apa saja persyaratan yang dibutuhkan oleh sistem machine learning?

- Bersifat andal (reliable).
- Mampu menangani perubahan kapasitas (scalable).
- Mudah beradaptasi (adaptable).
- Mudah dalam pemeliharaan (maintainable).

### ● **Alur Proyek Machine Learning di Industri**

Proyek machine learning di industri merupakan sebuah siklus. Secara umum, terdapat 6 tahap utama dalam siklus ini, yaitu:

- Cakupan proyek
- Manajemen data
- Pengembangan model machine learning
- Deployment
- Pengawasan dan pemeliharaan
- Analisis Bisnis

### ● **Infrastruktur Machine Learning**

Kompleksitas penerapan sistem machine learning di industri berimbas pada kebutuhan infrastruktur yang masif. Kebutuhan infrastruktur dalam sistem machine learning dibagi ke dalam tiga bagian, yaitu Data, Training/Evaluasi, dan Deployment.

### ● **Model Development**

Pengembangan model adalah proses yang iteratif. Anda biasanya memulai dengan proses seleksi model (memilih satu di antara banyak model kandidat). Kemudian, Anda akan mulai mengerucutkan kategori masalah menjadi lebih spesifik, misalnya apakah ini merupakan permasalahan klasifikasi atau regresi? Tahap selanjutnya adalah membuat baseline (model dasar). Terakhir, Anda perlu melakukan

pengaturan parameter untuk mendapatkan performa terbaik.

- **Deployment dan Monitoring**

Terdapat dua kategori pada machine learning model deployment, yaitu batch scoring dan real-time scoring. Batch scoring adalah kategori deployment saat model memproses seluruh dataset yang telah dikumpulkan dari waktu ke waktu untuk menghasilkan prediksi baru. Sedangkan, pada real-time scoring data diproses segera saat diterima atau secara real-time. Sehingga, hasilnya dapat langsung ditampilkan di aplikasi pada saat itu juga.

Pada proses deployment, terdapat beberapa strategi yang menyediakan beberapa versi layanan secara paralel. Strategi deployment ini menawarkan fleksibilitas dalam merilis layanan baru serta memungkinkan kita mengontrol dan mengelola versi mana yang menerima permintaan masuk. Tiga strategi deployment yang umum digunakan antara lain:

- **Recreate deployment.** Ia merupakan strategi penerapan ulang yang bekerja dengan sepenuhnya menurunkan versi aplikasi lama sebelum Anda meningkatkan versi aplikasi baru.
- **Blue-green deployment.** Ini merupakan strategi saat Anda ingin menerapkan versi baru aplikasi sekaligus memastikan bahwa layanan aplikasi tetap tersedia saat penerapan diperbarui.
- **Rolling Update Deployment.** Strategi ini mengalihkan lalu lintas secara bertahap ke versi baru.

Setelah proses deployment, kita harus memastikan model berjalan sebagaimana mestinya. Proses ini disebut monitoring. Apa saja yang metrik yang perlu dimonitor?

- Performa model.
- Model input.
- Performa sistem.

Dengan uraian tersebut, diharapkan Anda dapat memahami bagaimana mendesain sistem machine learning dan hal-hal apa saja yang perlu diketahui

selama prosesnya.

## **Menyusun Proyek Machine Learning**

Mari kita uraikan apa saja materi yang telah dibahas pada modul ini.

Portofolio machine learning merupakan aset penting untuk menunjukkan pemahaman dan keterampilan seseorang di bidang machine learning. Portofolio machine learning tidak harus terdiri dari proyek-proyek yang besar atau rumit. Proyek sederhana yang dapat menunjukkan kompetensi teknis Anda di bidang machine learning dapat dicantumkan sebagai portofolio.

Berikut adalah beberapa ide portofolio machine learning:

- Kursus yang pernah diikuti
- Proses data cleaning
- Eksplorasi dan visualisasi data
- Implementasi algoritma machine learning
- End-to-end machine learning project
- Kontribusi ke proyek open source

Setiap proyek machine learning itu bersifat unik. Tidak ada ketentuan khusus mengenai struktur atau kerangka proyek yang harus diikuti. Tetapi secara umum, ada template yang dapat kita ikuti saat menyusun proyek machine learning. Berikut adalah template yang umumnya digunakan dalam menyusun proyek machine learning:

- Problem statement
- Eksplorasi dan Persiapan data
- Pemodelan
- Evaluasi
- Finalisasi

Setelah menyelesaikan seluruh tahapan dalam proyek machine learning, Anda tentu ingin membuat dokumentasi teknis mengenai hal-hal yang telah dilakukan dalam proyek. Untuk mendukung Anda dalam menulis dan

mengomunikasikan proyek portofolio, Anda perlu menguasai skill **technical writing** (penulisan teknis).

Technical writing merupakan bentuk tulisan yang ditargetkan untuk pembaca yang mencari informasi tentang topik teknis tertentu, misalnya topik tentang komputer, rekayasa perangkat lunak, keuangan, kesehatan, dan lain-lain

Tahap selanjutnya adalah publikasi. Ada beberapa platform yang dapat Anda gunakan untuk menampilkan konten proyek Anda, yaitu blog pribadi dan media sosial (misalnya GitHub dan LinkedIn). Setiap platform memiliki kultur tertentu dan target pembaca yang spesifik. Jadi, pilihlah platform dengan bijaksana dan sesuaikan dengan target pembaca Anda.

Dengan uraian tersebut, diharapkan Anda dapat memahami bagaimana menyusun proyek machine learning.

### **Studi Kasus Pertama : Predictive Analytics**

Modul ini membahas mengenai studi kasus predictive analytics pada bidang marketing. Mari kita uraikan apa saja materi yang telah dibahas di modul ini.

Predictive analytics adalah bidang terapan yang menggunakan berbagai metode kuantitatif untuk membuat prediksi dengan memanfaatkan data. Ia merupakan seni membangun serta menggunakan model prediksi berdasarkan pola data historis. Data diubah dan diekstrak menjadi informasi berharga untuk menghasilkan keputusan analytics. Saat ini, keputusan analytics merupakan senjata strategis di bidang bisnis.

Keberhasilan proyek predictive analytics tidak hanya ditentukan oleh pemilihan algoritma machine learning saja, melainkan juga penerapan metodologi standar dalam mengelola seluruh tahapan atau siklus proyek. Cross-Industry Standard Process for Data Mining atau disingkat menjadi CRISP-DM merupakan salah satu metode standar proses analitik yang paling umum digunakan.

Berikut merupakan fase-fase utama proses analitik dalam metode CRISP-DM:

1. **Business understanding.**

Fase atau tahapan ini mencakup proses klarifikasi masalah, menentukan tujuan atau objective, dan mengidentifikasi sumber daya yang dimiliki.

2. **Data understanding.**

Ini merupakan tahapan untuk memahami informasi dalam data dan menentukan kualitasnya.

3. **Data preparation.**

Pada tahapan ini, Anda melakukan proses transformasi pada data sehingga menjadi bentuk yang cocok untuk proses pemodelan..

4. **Modeling.**

Ini merupakan tahapan saat Anda menggunakan algoritma machine learning untuk membuat model prediksi.

5. **Evaluation.**

Tahapan ini bertujuan untuk memastikan bahwa model akan mampu membuat prediksi yang akurat dan tidak mengalami overfitting atau underfitting.

6. **Deployment.**

Pada tahap ini, model machine learning yang telah Anda buat akan diintegrasikan ke dalam sistem organisasi.

Pada modul ini, Anda telah belajar menerapkan metode CRISP-DM dengan studi kasus prediksi harga diamonds. Pada studi kasus ini, Anda menggunakan tiga jenis algoritma, yaitu:

- **K-Nearest Neighbor.**

KNN bekerja dengan membandingkan jarak satu sampel ke sampel pelatihan lain dengan memilih sejumlah k-tetangga terdekat. Algoritma KNN menggunakan 'kesamaan fitur' untuk memprediksi nilai dari setiap data yang baru. Dengan kata lain, setiap data baru diberi nilai berdasarkan seberapa mirip titik tersebut dalam set pelatihan.

- **Random Forest.**

Algoritma random forest adalah salah satu algoritma supervised learning yang dapat digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Ia termasuk ke dalam kelompok model ensemble (group).

Algoritma ini disusun dari banyak algoritma pohon (decision tree) yang pembagian data dan fiturnya dipilih secara acak.

- **Boosting Algorithm.**

Algoritma boosting bekerja dengan membangun model dari data latih. Kemudian ia membuat model kedua yang bertugas memperbaiki kesalahan dari model pertama. Model ditambahkan sampai data latih terprediksi dengan baik atau telah mencapai jumlah maksimum model untuk ditambahkan.

Dengan uraian tersebut, diharapkan Anda dapat memahami apa itu model predictive analytics dan langkah-langkah dalam membuat modelnya. Jika ada materi yang masih belum dipahami, Anda bisa mengulas kembali materi yang diberikan di modul ini atau menanyakan di forum diskusi. Yuk, sekarang kita lanjut ke materi berikutnya!

### **Studi Kasus Kedua : Sentiment Analysis**

Pada modul ini, Anda telah memahami apa itu NLP, teknik-teknik NLP, dan menerapkannya dalam model machine learning. Anda juga telah berhasil membuat dua proyek sentimen analisis dengan teknik deep learning dan support vector machine. Mari kita uraikan apa saja materi yang telah dibahas di modul ini.

NLP merupakan subbidang dari Artificial Intelligence (AI) untuk menganalisis, memodelkan, dan memahami bahasa manusia. NLP mencakup berbagai algoritma, metode, dan permasalahan yang menggunakan teks sebagai input. Ia menghasilkan beberapa informasi seperti label, representasi semantik, dan sebagainya, sebagai output.

Teknik NLP digunakan di setiap aplikasi cerdas yang melibatkan bahasa alami. Ia merupakan komponen penting dalam berbagai aplikasi perangkat lunak yang kita gunakan dalam kehidupan sehari-hari.

Analisis sentimen merupakan salah satu topik dalam Natural Language Processing (Pemrosesan Bahasa Alami). Ia bertujuan untuk mengetahui pendapat orang mengenai suatu produk, layanan, atau masalah di masyarakat.

Sebuah aplikasi NLP menerapkan berbagai teknik untuk menyelesaikan permasalahan yang berbeda. Berikut adalah beberapa teknik NLP yang sering digunakan:

- **Pembuatan Teks (Text generation)**

Pembuatan teks, disebut juga Natural Language Generation (NLG) adalah proses menghasilkan keluaran teks bahasa alami dari bahasa lain. Teknik ini juga sering disebut sebagai text-to-text generation. Contoh aplikasinya adalah mesin penerjemah.

- **Pencarian Informasi (Information Retrieval)**

Information retrieval (IR) adalah proses menemukan informasi (biasanya berupa dokumen atau teks) yang relevan dengan kebutuhan. Informasi ini diambil dari kumpulan sumber informasi yang tersedia (biasanya di komputer atau internet).

- **Pemodelan Bahasa**

Dalam pemodelan bahasa, kita memprediksi kata berikutnya dalam kalimat berdasarkan histori kata-kata sebelumnya. Tujuannya untuk mempelajari kemungkinan urutan kata yang muncul.

- **Klasifikasi Teks**

Merupakan proses membagi teks ke dalam kumpulan kategori tertentu berdasarkan isinya. Klasifikasi teks merupakan teknik paling populer di NLP dan digunakan dalam berbagai aplikasi, misalnya identifikasi spam email dan analisis sentimen.

Secara umum, komponen utama alur proyek pengembangan sistem NLP adalah sebagai berikut:

- Data Acquisition.
- Text Cleaning.
- Pre-Processing.
- Feature Engineering.
- Modeling.
- Evaluation.
- Deployment.
- Monitoring and Model Update.



Terdapat beberapa cara untuk melakukan akuisisi data, antara lain:

- Menggunakan data publik.
- Mengumpulkan data dari internet (web scraping).
- Data (text) augmentation.

Text cleaning dan pre-processing terdiri dari tiga tahap utama, antara lain:

- Menghilangkan gangguan (noise) pada teks.
- Proses tokenisasi.
- Proses normalisasi.

Teknik rekayasa fitur untuk representasi teks mengubah teks pada data masukan ke dalam bentuk numerik sehingga dapat digunakan untuk pemodelan. Terdapat beberapa metode rekayasa fitur untuk representasi teks. Metode ini diklasifikasikan ke dalam empat kategori, antara lain:

- Basic vectorization approaches (contoh: One-Hot Encoding, Bag of Words, Bag of N-Grams, dan TF-IDF).
- Distributed representations (contoh: Words Embeddings).
- Universal text representation (menggunakan pre-trained embedding).
- Handcrafted features (mengandalkan domain-specific knowledge).

Pada modul ini, Anda telah melakukan proses analisis sentimen menggunakan data IndoNLU dengan teknik Deep Learning dan Support Vector Machine. SVM merupakan algoritma machine learning yang dapat digunakan untuk menyelesaikan permasalahan klasifikasi, regresi, dan pendeteksian outlier. Tujuan algoritma SVM adalah menemukan hyperplane terbaik dalam ruang berdimensi-N (ruang dengan N-jumlah fitur) yang memisahkan titik-titik data input dengan jelas.

Dengan uraian tersebut, diharapkan Anda dapat memahami bagaimana melakukan analisis sentimen terhadap data teks serta mampu membuat modelnya dengan teknik Deep Learning dan algoritma Support Vector Machine (SVM).

## Studi Kasus Ketiga : Computer Vision

Pada modul ini, Anda telah mempelajari apa itu computer vision, penerapan computer vision di industri, dan berbagai teknik dalam bidang computer vision. Anda juga telah berhasil membuat model machine learning untuk klasifikasi rambu lalu lintas (traffic sign) dan deteksi wajah sederhana. Well done!

Mari kita uraikan apa saja materi yang telah dibahas di modul ini.

Computer vision adalah bidang yang meniru kemampuan penglihatan manusia (*human vision capabilities*). Ia bekerja dengan membangun metode untuk pembentukan gambar (meniru sistem sensorik manusia) dan persepsi mesin (meniru sistem kognitif manusia). Sistem sensorik manusia ditiru dengan melibatkan sistem sensor seperti kamera, pengaturan desain, dan penempatannya. Sedangkan, pendekatan modern untuk meniru sistem kognitif manusia terdiri dari metode machine learning yang digunakan untuk mengekstrak informasi dari gambar

Berikut adalah beberapa contoh penerapan computer vision di industri:

- **Bidang Kesehatan**

Di bidang kesehatan, computer vision memungkinkan para tenaga medis menggunakan data citra medis (medical imaging data) untuk membantu memberikan diagnosis, pengobatan, dan prediksi penyakit. Selain itu, computer vision juga digunakan untuk memprediksi dan menganalisis gejala gangguan pada gigi, neurologis dan neuropathy, memantau kehilangan darah untuk mengoptimalkan transfusi, mendeteksi COVID-19, dan lain sebagainya.

- **Autonomous Driving**

Vision (penglihatan) dan perception (persepsi) adalah pilar utama dari self driving cars. Computer vision berperan penting dalam mengembangkan autonomous driving cars.

- **Facial Recognition System**

Sistem pengenalan wajah merupakan salah satu penerapan bidang computer vision yang telah digunakan secara luas. Ia bekerja dengan

memetakan, menganalisis, dan mengonfirmasi identitas wajah dalam berbagai gambar digital (foto atau video).

Dalam klasifikasi gambar, kita mengasumsikan hanya ada satu target utama dalam gambar. Sehingga, model hanya fokus untuk mengidentifikasi kategori target. Pada deteksi objek, kita melakukan proses klasifikasi gambar sekaligus mendapatkan posisi spesifik masing-masing target pada gambar. Pada prinsipnya, deteksi objek melibatkan dua tugas utama:

- Melokalisasi satu atau lebih objek dalam gambar.
- Mengklasifikasikan setiap objek dalam gambar.

Proses ini dilakukan dengan menggambar kotak pembatas di sekitar objek dengan kelas yang diprediksi. Ini berarti, sistem tidak hanya memprediksi kelas gambar, tapi juga koordinat kotak pembatas yang bersesuaian dengan objek. Jadi, dalam tugas deteksi objek, kita ingin menemukan semua objek dalam gambar dan menggambar bounding box (kotak pembatas) sekelilingnya.

Untuk lebih memudahkan dalam memahami berbagai teknik deteksi objek, kita akan membaginya menjadi dua kategori sebagai berikut:

- **Pendekatan Klasik**  
Contoh algoritma: Haar Cascade dan Histogram of Oriented Gradients (HOG).
- **Pendekatan Deep Learning**  
Contoh algoritma: R-CNN, Fast R-CNN, dan YOLO.

Pada modul ini, Anda telah membuat model machine learning untuk klasifikasi rambu lalu lintas menggunakan German Traffic Sign Dataset. Selain itu, Anda juga telah berhasil membuat model deteksi wajah pada gambar dengan library OpenCV.

Dengan uraian tersebut, diharapkan Anda dapat memahami bagaimana computer vision bekerja serta mampu mengimplementasikannya dalam

proyek computer vision.

### **Studi Kasus Keempat : Sistem Rekomendasi**

Kini Anda telah memahami apa itu sistem rekomendasi, tujuan dan fungsi sistem rekomendasi, kategori, serta teknik-teknik dalam sistem rekomendasi. Mari kita uraikan apa saja materi yang telah dipelajari sejauh ini.

Sistem rekomendasi adalah sistem yang memprediksi rating atau preferensi pengguna terhadap item tertentu. Rekomendasi ini dibuat berdasarkan perilaku pengguna di masa lalu atau perilaku pengguna lainnya. Jadi, sistem ini akan merekomendasikan sesuatu terhadap pengguna berdasarkan data perilaku atau preferensi dari waktu ke waktu.

Sistem ini tidak merekomendasikan item secara spesifik. Ia merekomendasikan sejumlah item yang mungkin cocok dengan preferensi pengguna. Oleh karena itu, pada sistem rekomendasi, keluarannya berupa *“top-N” recommendation*.

Secara umum, fungsi dan tujuan sistem rekomendasi adalah sebagai berikut:

- Meningkatkan jumlah item yang terjual.
- Menjual item yang beragam.
- Meningkatkan kepuasan pengguna.
- Pemahaman yang lebih baik tentang preferensi pengguna.

Sistem rekomendasi diklasifikasikan ke dalam dua kategori besar, yaitu non-personalized (tanpa personalisasi) dan personalized (dengan personalisasi).

Pada non-personalized recommender system, setiap orang yang berinteraksi dengan sistem akan menerima daftar rekomendasi yang sama. Sistem ini memberikan rekomendasi pada pengguna tanpa konteks apa pun, baik itu konteks preferensi maupun rating dari pengguna.

Sementara itu, teknik rekomendasi dengan personalisasi dapat diklasifikasikan lagi menjadi beberapa kategori berbeda, antara lain Content-Based Filtering (berbasis konten), Collaborative Filtering (filter kolaboratif), dan Hybrid method.

Content-based filtering mempelajari profil minat pengguna baru berdasarkan data dari objek yang telah dinilai sebelumnya oleh pengguna. Sedangkan, Collaborative filtering bergantung pada pendapat komunitas pengguna. Ia tidak memerlukan atribut untuk setiap itemnya seperti pada sistem berbasis konten. Sementara itu, sistem rekomendasi hibrida menggabungkan dua atau lebih strategi rekomendasi yang berbeda.

Pada kategori collaborative filtering, terdapat beberapa teknik, antara lain:

- Memory based. Contohnya, user based dan item based.
- Model based. Contohnya, matrix factorization dan pendekatan deep learning.

Data untuk sistem rekomendasi terbagi dalam tiga jenis kategori, antara lain: item, pengguna, dan transaksi (hubungan antara pengguna dan item). Item merupakan objek yang direkomendasikan. Objek ini bisa dibedakan dari kompleksitas, nilai, atau kegunaannya. Pada data berupa pengguna, jenis informasi yang dibutuhkan bergantung pada teknik rekomendasi yang digunakan. Sementara itu, data transaksi merupakan interaksi antara pengguna dengan sistem.

Pada dasarnya sistem rekomendasi dibuat berdasarkan tingkat kesamaan atau similarity. Kesamaan paling umum yang bisa diukur dalam sistem rekomendasi antara lain, preferensi dan selera. Selain itu, kesamaan juga dapat ditemukan melalui data dan informasi lain seperti demografi pengguna dan status sosial.

Selain materi-materi di atas, Anda juga telah berhasil membuat model sistem rekomendasi berdasarkan konten (content-based) dan collaborative filtering menggunakan data [restaurant and consumer](#). Saat membuat model sistem rekomendasi, Anda telah belajar memahami data, menggabungkan beberapa

file data, memilih fitur yang dibutuhkan, dan membuat sistem rekomendasi dengan teknik content-based dan collaborative filtering. Well done!

Dengan uraian tersebut, diharapkan Anda dapat memahami bagaimana sistem rekomendasi bekerja serta mampu mengimplementasikannya dalam proyek sistem rekomendasi dengan data yang ada.

Jika ada materi yang masih belum dipahami, Anda bisa mengulas kembali materi yang diberikan di modul tersebut atau menanyakannya di forum diskusi.

## Daftar Referensi

Berikut adalah daftar referensi yang digunakan pada seluruh modul di kelas ini.

[1] Huyen, Chip. "Introduction to Machine Learning System Design". CS 329S: Machine Learning Systems Design. Stanford University. USA. Tersedia: [tautan](#). Diakses pada 18 Mei 2021.

[2] Huyen, Chip. "Machine Learning Interviews". GitHub.com/chiphuyen. Tersedia: [tautan](#). Diakses pada 19 Mei 2021.

[3] Manchepalli, *et al.* "Reliable Machine Learning". Microsoft.com. Tersedia: [tautan](#). Diakses pada 20 Mei 2021.

[4] Saria dan Subbhashwamy. "Tutorial: Safe and Reliable Machine Learning". 2019. ACM Conference on Fairness, Accountability, and Transparency (FAT\* 2019). Tersedia: [tautan](#).

[5] DeepLearning.ai. "A Chat with Andrew on MLOps: From Model-centric to Data-centric AI". (25 Maret 2021). Diakses pada: 22 Mei 2021. [Online Video]. Tersedia: [tautan](#).

- [6] Goodfellow et al. "Explaining and Harnessing Adversarial Examples". 2015. ICLR. Tersedia: [tautan](#).
- [7] Google Cloud. "Introduction to Cloud Monitoring". Tersedia: [tautan](#). Diakses pada 25 Mei 2021.
- [8] D. Sculley, et al. "Machine Learning: The High-Interest Credit Card of Technical Debt". Google Inc. Tersedia: [tautan](#). Diakses pada: 28 Mei 2021.
- [9] Google Cloud. "MLOps: Continuous Delivery and Automation Pipelines in Machine Learning". Tersedia: [tautan](#). Diakses pada: 29 Mei 2021.
- [10] Sculley, et al. "Hidden Technical Debt in Machine Learning Systems". Google Inc. Tersedia: [tautan](#). Diakses pada: 29 Mei 2021.
- [11] Sergey. arayev, et al. (2021). "Full Stack Deep Learning". [Online]. Tersedia: [tautan](#). Diakses pada 29 mei 2021.
- [12] Rogati, Monica. "The AI Hierarchy of Needs". Tersedia: [tautan](#). Diakses pada 29 Mei 2021.
- [13] Gabor Mellis, et al. "On the State of the Art of Evaluation in Neural Language Models". 2018. ICLR. Tersedia: [tautan](#).
- [14] Heidmann. "Introducing MLOps". O'Reilly Media Inc. 2020. Tersedia: [tautan](#).
- [15] Google Cloud. "Application Deployment and Testing Strategies". Tersedia: [tautan](#). Diakses pada 30 Mei 2021.
- [16] Moscow Institute of Physics and Technology. "Technical Writing". [Online]. Tersedia: [tautan](#).
- [17] Laplante, Phillip A. "Technical Writing". O'Reilly Media Inc. 2016.

- [18] Google Developers. "Technical Writing". Tersedia: [tautan](#). Diakses pada: 30 Mei 2021.
- [19] Kelleher, John D, et al. "Machine Learning for Predictive Data Analytics". MIT Press. 2020. Tersedia: [tautan informasi buku](#).
- [20] Delen, Dursun. "Predictive Analytic: Data Mining, Machine Learning and Data Science for Practitioners". Pearson FT. Press. 2020. Tersedia: [O'Reilly media](#).
- [21] IBM Cloud. "Exploratory Data Analysis". Tersedia: [tautan](#). Diakses pada 2 Juni 2021.
- [22] Kang, Hyun. "The Prevention and Handling the Missing Data". Tersedia: [tautan](#). Diakses pada Juni 2021.
- [23] Kuhn, Max dan Johnson Kjell. "Applied Predictive Modeling". Springer. 2013.
- [24] Seltman, Howard J. "Experimental Design and Analysis". 2018. Tersedia: [tautan](#). Diakses pada Juni 2021.
- [25] Fuentes, Alvaro. "Hands-on Predictive Analytics with Python". Packt Publishing. 2018. Tersedia: [O'Reilly Media](#).
- [26] Rhys, Hefin. "Machine Learning with R, the Tidyverse, and MLR". Manning Publications. 2020. Tersedia: [O'Reilly Media](#).
- [27] Scikit-learn Documentation. Tersedia: [tautan](#). Diakses pada: Juni 2021.
- [28] Parker, Mary. "Sampling with Replacement and Sampling without Replacement". Tersedia: [tautan](#). Diakses pada Juni 2021.
- [29] Kemp, Simon. "Digital 2021: Global Overview Report". Published in Datareportal. 2021. Tersedia: [tautan](#). Diakses pada Juni 2021.



- [30] Hagiwara, Masato. "Introduction to Natural Language Processing". Manning Publishing. 2020. Tersedia: [tautan](#). Diakses pada: Juni 2021.
- [31] Vajjala, Sowmya, et al. "Practical Natural Language Processing". O'Reilly Media. 2020. Tersedia: [tautan](#).
- [32] Gu, Xinxing. "Google Translate's instant camera translation gets an upgrade". 2019. Tersedia: [tautan](#). Diakses pada 26 Juni 2021.
- [33] Manning, Christopher D, et al. "Introduction to Information Retrieval". Cambridge University Press 2008. Tersedia: [tautan](#).
- [34] Lovesque, Hector J. "The Winograd Schema Challenge". 13th International Conference on Principles of Knowledge Representation and Reasoning. Tersedia: [tautan](#).
- [35] Kemendikbud. "Badan Bahasa Petakan 652 Bahasa Daerah di Indonesia". Tersedia: [tautan](#). Diakses pada 29 Juni 2021.
- [36] Chaudhary, Amit. "Data Augmentation for NLP". Tersedia: [tautan](#). Diakses pada 30 Juni 2021.
- [37] Wilie, Bryan, et al. "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding". ACL-IJNLP. 2020. Tersedia: [tautan](#).
- [38] Purwarianti, Ayu dan Crisdayanti, Ida Ayu Putri Ari. "Improving Bi-LSTM Performance for Indonesian Sentiment Analysis Using Paragraph Vector". Tersedia: [tautan](#).
- [39] Indobenchmark. "Pre-Trained Models for NLP Tasks Using PyTorch". Tersedia: [tautan](#). Diakses pada Juli 2021.
- [40] Scikit-learn Documentation. "TfidfVectorizer". Tersedia: [tautan](#). Diakses pada 9 Juli 2021.

[41] Lakshmanan, Valliappa, et al. "Practical machine Learning for Computer Vision". O'Reilly Media. 2020. Tersedia: [tautan](#).

[42] Wang, Linda, et al. "COVID-Net: a Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 cases from chest X-ray images". Scientific Reports. 2020. Tersedia: [tautan](#). Diakses pada 23 Jul 2021.

[43] Gupta Abhishek, et al. Deep Learning for Object Detection and Scene Perception in Self-driving Cars: Survey, Challenges, and Open Issue". Tersedia: [tautan](#).

[44] Kortli, Yasssin, et al. "Face Recognition Systems: A Survey". 2020. Tersedia: [tautan](#). Diakses pada 25 Juli 2021.

[45] Russakovsky, et al. "ImageNet, Large Scale Visual Recognition Challenge". 2015. Tersedia: [tautan](#).

[46] Dalal, Navneet dan Triggs Bill. " Histogram of Oriented Gradients for Human Detection". IEEE. 2005. Tersedia: [tautan](#).

[47] Scikit-image Documentation. "Histogram of Oriented Gradient". Tersedia: [tautan](#). Diakses pada 7 Agustus 2021.

[48] Girshick, Ross, et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". 2013. Tersedia: [tautan](#).

[49] Girshick, Ross. "Fast R-CNN". Microsoft Research. 2015. Tersedia: [tautan](#).

[50] Redmon, Joseph. "You Only Look Once: Unified, Real-Time Object Detection". 2015.

[51] Viola, Paul dan Jones, Michael. "Rapid Object Detection using a Boosted Cascade of Simple Features". 2001. Conference on Computer Vision and Pattern Recognition. Tersedia: [tautan](#).

- [52] OpenCV Tutorial. "Casscade Classifier Training". Tersedia: [tautan](#).
- [53] Kane, Frank. "Building Recommender Systems with Machine Learning and AI". 2018. O'Reilly Media. Tersedia: [tautan](#).
- [54] Ricci, Francesco, et al. "Recommender Systems Handbook". Springer Media. 2011. Tersedia: [tautan](#).
- [55] Amatriain, Xavier dan Basilico, Justin. "Netflix Recommendations: Beyond the 5 Stars". 2012. Netflix Technology Blog. Tersedia: [tautan](#).
- [56] Cremonesi, Paolo. "Course: Basic Recommender Systems". [Online Video]. Tersedia: [tautan](#).
- [57] Choi, Sang-Min, et al. "A Recommendation Model Using the Bandwagon Effect for E-Marketing Purposes in IoT". 2015. International Journal of Distributed Sensor Networks. Tersedia: [tautan](#).
- [58] Breese, John S et al. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering". Tersedia: [tautan](#).
- [59] Covington, Paul, et al. "Deep Neural networks for YouTube Recommendation". Google. Tersedia: [tautan](#).
- [60] Kim, Falk. "Practical Recommender Systems". 2019. Manning Publications. Tersedia: [tautan](#).