

MACHINE LEARNING BOOTCAMP (4 BULAN)

Mengenai Bootcamp ini: Kita sedang memasuki era baru dari Artificial Intelligence (AI) dikarenakan sebuah teknologi yang disebut Deep Learning. Deep learning digunakan untuk mengenali wajah seseorang, membuat sistem pengenalan suara, menganalisa artikel, dan banyak hal lainnya. Walau demikian, kita perlu mempelajari ilmu di balik Deep Learning sebelum kita terjun ke bidang tersebut. Hal tersebut adalah Machine Learning yang merupakan bidang yang lebih luas daripada Deep Learning.

Di bootcamp ini, kita akan mempelajari apa itu Machine Learning dan berbagai tipe Machine Learning seperti *supervised machine learning* dan *unsupervised machine learning*. Kita akan mulai dari nol, yaitu mempelajari tools-tools yang dibutuhkan seperti **Python, Numpy, Pandas, dan Matplotlib**. Setelah mempelajari tools dasar tersebut barulah kita terjun ke dunia Machine Learning menggunakan **Scikit-learn** dan **J.Cop ML**, dan akhirnya kita akan mempelajari Deep Learning menggunakan **PyTorch**.

Di setiap materi akan disiapkan sebuah studi kasus yang dikerjakan bersama instruktur agar dapat memahami materinya lebih baik, dan tentunya agar kita terbiasa untuk mengaplikasikan ilmu yang dipelajari ke kasus-kasus lainnya.

Untuk semua kalangan: Bootcamp ini tidak hanya bermanfaat untuk yang baru ingin mempelajari Machine Learning, tetapi untuk praktisi dan akademisi juga. Hal ini dikarenakan kita menggunakan teknologi terbaru (misalnya workflow scikit-learn terbaru) dan tools yang dapat membuat pembuatan machine learning menjadi lebih mudah dan efisien. Selain itu, materi kami juga dilengkapi dengan tools visualisasi algoritma yang mempermudah siapapun untuk mengajarkan machine learning.

Prasyarat :

Matematika level SMA — Walau kita mengajarkan semuanya dari nol, ada baiknya peserta bootcamp ini paling tidak mengetahui statistik sederhana seperti apa itu rata-rata, median, modus, dan apa itu histogram. Tentunya pengalaman lebih di bidang pemrograman atau matematik merupakan bonus untuk mengikuti bootcamp ini.

Laptop — Sistem operasi 64-bit dengan ram setidaknya 4GB, dan memiliki sisa storage setidaknya 2GB untuk instalasi dan materi.

Frequently Asked Question (FAQ):

Bahasa apa yang digunakan dalam bootcamp ini?

Semua materi dan pembawaan akan menggunakan bahasa Indonesia, tetapi beberapa istilah yang umum di machine learning akan tetap dalam bahasa inggris.

Saya sudah familiar dengan python dan scikit-learn. Apakah saya masih perlu mengikuti bootcamp ini?

Kamu sudah ada bekal yang bagus! Bootcamp ini dapat membantumu agar lebih rapi dan efisien karena kami menggunakan workflow scikit-learn terbaru dan J.Cop ML. Hal tersebut membuat pengolahan data campuran menjadi mudah dan lebih efisien. Kami juga menggunakan LuWiji untuk membuat algoritma machine learning dapat dipahami secara visual dan interaktif. Selain itu, kami juga memberikan banyak tips yang practical, dan tentunya kesalahan umum yang perlu dihindari.

KURIKULUM

Course 1 Pengenalan Machine Learning dan Tools Data Science	
Pengantar Machine Learning	→ Jenis-jenis machine learning → Machine Learning dan aplikasinya
Pengenalan Miniconda	→ Kenapa kita membutuhkan Miniconda → Instalasi Miniconda → PyPI vs Conda → Anaconda vs Miniconda
Pengenalan Jupyter	→ Instalasi Jupyter → Belajar menggunakan jupyter notebook
Pengenalan Python	→ Belajar menggunakan python
Membuat environment	→ Apa itu environment → Kenapa kita butuh environment → Membuat environment untuk bootcamp ini
Tools Data science	→ Belajar menggunakan numpy untuk operasi matematik → Belajar menggunakan pandas untuk pengolahan data → Belajar menggunakan matplotlib dan seaborn untuk visualisasi
Course 2 Supervised Learning: Regresi dan Klasifikasi	
K Nearest Neighbor	→ Apa itu K Nearest Neighbor (KNN) dan cara kerjanya → Memahami parameter KNN
Alur Dasar Machine Learning Part 1	→ Dataset Splitting → K-fold Cross Validation → Meningkatkan performa melalui data (Feature Engineering) → Meningkatkan performa melalui model → Feature Engineering: Feature Scaling → Parameter Tuning → Menghindari Data Leakage → Pipeline → Evaluasi Model → Inferensi / Prediksi → Menyimpan Model Case: Survivability Prediction
Alur Dasar Machine Learning Part 2	→ Analisis Eksplorasi Data → Feature Engineering: Feature creation → Feature Engineering: Encoding dan Binning → Column Transformer → Kenapa mulai dengan model yang sederhana

	Case: Survivability Prediction
Baseline dan Benchmark	<ul style="list-style-type: none"> → Apa itu baseline dan benchmark → Kompetisi Kaggle
Algorithm: Support Vector Machine	<ul style="list-style-type: none"> → Apa itu Support Vector Machine (SVM) dan cara kerjanya → Kernelized SVM → SVM untuk nilai yang kecil → Memahami parameter SVM <p>→ Exercise: Regresi menggunakan SVM Case: Car Price Prediction</p> <ul style="list-style-type: none"> → Exercise: Klasifikasi menggunakan SVM → Imbalance dataset dan class weight <p>Case: Credit card fraud detection</p>
Algorithm: Decision Tree dan Random Forest	<ul style="list-style-type: none"> → Apa itu Decision Tree dan cara kerjanya → Memahami parameter Decision Tree → Apa itu Random Forest dan cara kerjanya → Memahami parameter Random Forest <p>→ Exercise: Regresi menggunakan Random Forest</p> <p>→ Feature Engineering: Feature crossing</p> <p>→ Feature Engineering: Feature importance (MLD dan MSD) Case: House price prediction</p> <p>→ Exercise: Klasifikasi menggunakan Random Forest</p> <ul style="list-style-type: none"> → Pengenalan workflow 4 fasa → Randomized Search <p>Case: Mobile phone classification</p>
Algorithm: Boosting	<ul style="list-style-type: none"> → Apa itu Ensemble Learning → 2 Ensemble Learning paling umum → Weak, good, dan strong learner → Apa itu Adaptive Boosting (AdaBoost) dan cara kerjanya → Memahami parameter AdaBoost → Apa itu Gradient Boosting (GradBoost) dan cara kerjanya → Memahami parameter GradBoost <p>→ Exercise: Regresi menggunakan XGBoost</p> <ul style="list-style-type: none"> → Latihan workflow 4 fasa → Bayesian Search <p>Case: Life Expectancy Estimation</p>
Course 3 In-Depth Supervised Learning: Evaluasi Regresi dan Klasifikasi	
Regresi	<ul style="list-style-type: none"> → Metric untuk Regresi → Apa itu Linear Regression dan cara kerjanya → Pengenalan Loss, dan optimizer → Gradient dan non-Gradient based Optimizer

	<ul style="list-style-type: none"> → Apa itu gradient descent → Learning rate → Polynomial Regression → Prinsip Bias-Variance Tradeoff <p>→ Exercise: Analisis Bias-Variance Tradeoff Case: Parameter KNN, SVM, dan Random Forest</p> <ul style="list-style-type: none"> → 3 cara mengevaluasi regresi → Exercise: Regresi menggunakan Polynomial Regression → Feature Engineering: Time Feature Extraction → Manfaatkan Regression Evaluation untuk dapat insight <p>Case: Taxi fare prediction</p> <ul style="list-style-type: none"> → Feature Engineering: MinMax, Standard, dan Robust Scaling → Feature Engineering: Power Transform (Box-Cox & Yeo-Johnson) <ul style="list-style-type: none"> → Regularization → Ridge, Lasso, dan ElasticNet <ul style="list-style-type: none"> → Exercise: Regresi menggunakan ElasticNet → Manfaatkan Residual Plot untuk analisa → Tips Regresi: Matrix Korelasi <p>Case: Taxi fare prediction</p>
Klasifikasi	<ul style="list-style-type: none"> → Metric untuk Klasifikasi → Kelemahan Gradient Descent → Apa itu Logistic Regression dan cara kerjanya → Cross Entropy Loss → Sigmoid dan Probabilitas → Klasifikasi Multiclass dan Multilabel <p>→ Exercise: Klasifikasi Multiclass menggunakan Logistic Regression Case: Flower Classification</p> <ul style="list-style-type: none"> → 3 cara mengevaluasi klasifikasi → Precision-Recall Tradeoff → kurva PR dan kurva ROC → Efek imbalance data pada ROC dan PR <ul style="list-style-type: none"> → Exercise: Klasifikasi menggunakan Logistic Regression → Tips Klasifikasi: Matriks Asosiasi <p>Case: Poisonous mushroom identification</p>
Course 4 Menangani Unstructured Data	
Unstructured Data	<ul style="list-style-type: none"> → Apa itu Structured dan Unstructured Data
Data berupa gambar	<ul style="list-style-type: none"> → Menampilkan gambar menggunakan Pillow dan matplotlib → Komponen gambar digital: channel dan pixel → Strukturisasi data berupa gambar

	<ul style="list-style-type: none"> → Exercise: Image Recognition Case: Handwritten digits recognition → Mempersiapkan metadata dengan os → Exercise: Persiapan metadata Case: Handwritten digits recognition
Data berupa tulisan	<ul style="list-style-type: none"> → Dasar preprocessing data text: tokenisasi, normalisasi, dll. → Strukturisasi data berupa text: Bag of Words, TFIDF → Exercise: Spam classification Case: Indonesian SMS spam classification
Course 5 Transisi ke Deep Learning	
Pengantar Deep Learning dan Neural Network	<ul style="list-style-type: none"> → Deep Learning vs Machine Learning → Kombinasi Linear, Representasi Node, dan Fungsi Aktivasi → Apa itu Neural network dan cara kerjanya → Istilah-istilah di neural network → Neural network playground → Neural network menggunakan scikit-learn → Exercise: Neural Network Playground Case: Spiral Dataset
Neural Network menggunakan PyTorch Sequential API	<ul style="list-style-type: none"> → PyTorch introduction → Feedforward, dan Backpropagation → Momentum, dan Adaptive Optimizer → Neural Network menggunakan PyTorch Sequential API → Apa itu Dropout → Save dan load model di PyTorch → Exercise: Klasifikasi menggunakan neural network Case: Diabetes Classification
Neural Network menggunakan PyTorch Module	<ul style="list-style-type: none"> → Pengenalan Class dan Object Oriented Programming (OOP) → Neural Network menggunakan PyTorch Module → PyTorch Workflow, Dataloader, & Callbacks → Torchvision → Exercise: Klasifikasi gambar menggunakan Neural Network Case: Fashion Image Recognition
Course 6 Deep Learning menggunakan PyTorch	
Convolutional Neural Network (CNN)	<ul style="list-style-type: none"> → Blur Filter dan konvolusi → Filter dan edge detection → Stride, padding dan pooling

	<ul style="list-style-type: none"> → Apa itu Convolutional Neural network dan cara kerjanya → Kelebihan CNN → Augmentasi data → Exercise: Klasifikasi gambar menggunakan CNN <p>Case: Pest in Plants</p>
Transfer Learning	<ul style="list-style-type: none"> → Apa itu Transfer Learning dan cara kerjanya → Apa itu kompetisi ILSVRC → Arsitektur CNN yang populer: VGG, Inception, Resnet, dkk → Exercise: Klasifikasi gambar menggunakan Transfer Learning CNN <p>Case: Natural Scene Multilabel Classification</p>
Recurrent Neural Network (RNN)	<ul style="list-style-type: none"> → Apa itu Recurrent Neural Network dan cara kerjanya → RNN menggunakan PyTorch → Memahami parameter di RNN → Melakukan Forecasting Time series → Exercise: Regresi menggunakan RNN <p>Case: Daily minimum temperature Time Series</p>
Varian dari RNN	<ul style="list-style-type: none"> → Apa itu Long-Short Term Memory (LSTM) dan cara kerjanya → Apa itu Gated Recurrent Unit (GRU) dan cara kerjanya → Apa itu Bidirectional pada varian RNN → Exercise: Regresi menggunakan LSTM, GRU, dan BiLSTM <p>Case: Daily minimum temperature Time Series</p> <ul style="list-style-type: none"> → Exercise: Latihan tuning parameter pada varian RNN <p>Case: Damped Sine & Stock price</p> <ul style="list-style-type: none"> → Pendapat saya mengenai relevansi algoritma machine learning untuk prediksi harga saham, maupun time series forecasting secara umum → Menggunakan input yang multivariat pada varian RNN → Exercise: Memanfaatkan informasi quarter sebagai fitur <p>Case: Daily minimum temperature Time Series</p> <ul style="list-style-type: none"> → Aplikasi RNN lainnya
Course 7 Unsupervised Learning	
Clustering	<ul style="list-style-type: none"> → K-Means, DBSCAN, dan Agglomerative Hierarchical Clustering → Clustering menggunakan scikit-learn → Menentukan jumlah cluster dengan Silhouette score and Elbow Analysis → Agglomerative Hierarchical Clustering menggunakan scipy

	<p>→ Exercise: Image Quantization Case: Indonesian Endangered Animals</p>
Dimensionality Reduction	<p>→ Principal Component Analysis (PCA) dan cara kerjanya → Kompresi data menggunakan PCA</p> <p>→ Exercise: PCA untuk membantu supervised learning → Menentukan jumlah komponen PCA Case: Breast Cancer Dataset</p> <p>→ PCA dan NMF untuk ekstraksi fitur → Exercise: Ekstraksi fitur secara unsupervised Case: Labeled Faces in Wild feature extraction</p> <p>→ LSA dan LDA untuk Topic Modeling → Exercise: Topic modeling Case: Kompas News Dataset</p> <p>→ t-SNE untuk visualisasi dimensi orde tinggi → Exercise: Visualisasi dimensi orde tinggi secara unsupervised Case: Fashion Image data visualization</p>
Course 8 Deep Unsupervised Learning	
Autoencoder	<p>→ Apa itu Autoencoder dan cara kerjanya → Encoder, ruang laten, dan decoder</p> <p>→ Exercise: Mengkompres dan rekonstruksi gambar → Interpolasi pada ruang laten Case: Handwritten digit dataset</p> <p>→ Apa itu Denoising Autoencoder dan cara kerjanya → Exercise: menghilangkan noise pada gambar Case: Handwritten digit dataset</p> <p>→ Variasi dan aplikasi autoencoder lainnya</p>
Generative Adversarial Network (GAN)	<p>→ Apa itu GAN dan cara kerjanya</p> <p>→ Exercise: GAN untuk gambar grayscale Case: Handwritten digit dataset</p> <p>→ Apa itu Deep Convolutional GAN (DCGAN) dan cara kerjanya</p> <p>→ Exercise: DCGAN untuk gambar berwarna Case: CelebA dataset</p>
Course 9 Sistem Rekomendasi dan REST API	

Sistem Rekomendasi	<ul style="list-style-type: none">→ Demographic Filtering→ Content-based Filtering→ Exercise: membuat sistem rekomendasi Case: TMDB movie recommendation→ Collaborative Filtering→ Exercise: membuat sistem rekomendasi hingga REST API nya Case: Movielens dataset→ Deploy model menggunakan REST API menggunakan Flask→ Exercise: membuat REST API untuk rekomendasi film
---------------------------	---