Muhammad Hatta

Lead Data Scientist / AI developer

mhaaatta@gmail.com

https://hattajr.github.io/

Ulsan, South Korea

Profile

A Full-stack Data scientist working on developing and implementing end-to-end machine learning system. In particular, my work covers exploratory data analysis (e.g., data wrangling, data handling, comprehensive visualization), model development (e.g., algorithm selection, feature extraction, testing/validation, computational performances), and model deployment (e.g., MLOps, CD/CI, monitoring).

Education

2018 Feb – 2020 Feb
Seoul, South Korea

MS
Kookmin University
Department of Mechanical Engineering (advise by Hwataik Han)
Thesis: Predicting Indoor PM2.5 and PM10 Concentrations using Simplified Neural Network

2012 Sep – 2016 Dec
Indonesia

BS
University of Riau
Department of Mechanical Engineering (advise by Awaluddin Martin)
Thesis: Computational simulation and manufacture of a wind turbine for urban area

Professional Experience

2020 Mar – present
Ulsan, South Korea

Lead Data Scientist & ML Engineer
ITS Co., Ltd - Predictive maintenance & IoT-based solution

- Main developer of an Advanced predictive maintenance system (LUDA) which focuses on early detection of faulty and abnormal equipment operation in industries.
  Research domain: Investigate, explore, and apply the latest research trends to build robust machine learning models implemented for fault detection, anomaly detection, RUL(remaining useful life), and other machinery condition monitoring strategy.
  Engineering domain: Designed & developed complex infrastructure to support LUDA ecosystem. Includes big data handling, stream data processing (real-time prediction), product deployment, resource management, and others related to MLOps.
- Understanding business objectives and developing models that help to achieve them, along with metrics to track their progress.
- Designing various efficient algorithms for time-series data handling and processing (data segmentation, data transformation, feature extraction, etc).
- Analyzing the ML algorithms that could be used to solve a given problem and ranking them by their success probability.
- Developed task schedule manager to monitor machine learning workflow using Apache airflow.
- Exploring and visualizing data to gain an understanding of it, then identifying differences in data distribution that could affect performance when deploying the model in the real world.
- Managed cross-functional team ranging from entry-level developer to head of department and collaborated with front-end, back-end developer, and marketing team.
- Have participated in several R&D projects collaboration targeted to AI-based infrastructure as the main developer (*describe in Selected Project Section*).

2018 Feb – 2020 Feb
Seoul, South Korea

MS researcher
Thermal Environmental Engineering Labs

- Researched an efficient and reliable Neural network algorithm to predict Indoor PM's concentration.
- Participated in several local and international conferences as a keynote speaker on computational uncertainty, machine learning explainability, and other machine learning-related topics.

Selected Projects & Collaborations

2021 Aug – present

Korea Aerospace Industries (KAI)
Fault Detection in CNC machine
Build and deploy a Machine learning model to detect several abnormal conditions of the CNC machining in real-time (streaming process). In this project, I also designed the experiment plan for data acquisition and data labeling.

2020 Sep – 2021 Feb

Seoul National University X AI NATION
AI Factory Framework
Collaborated with Seoul National University (Numerical Computing & Image Analysis Lab) and AI NATION to develop several Machine learning models which will be applied as a reference for data-driven maintenance strategy for industries. As the result, we successfully build three (3) main ML models including LSTM, VAE-RNN, and Isolation forest.

Skills

- Two and half years of intensive Python programming experience with a specialization in AI infrastructure and scientific computing, including visualization, big data processing, and machine learning.
- Comprehensive knowledge of major Python packages in data manipulation and machine learning development, including Pandas, Numpy, Tensorflow, Matplotlib, and other related modules.
- Deep understanding of time-series modeling, including data transformation, data segmentation, anomaly detection, data augmentation, and feature extraction.
- Strong understanding of various machine learning architectures and methods, including deep learning architecture (i.e., LSTM, autoencoder), tree-based algorithm (i.e., decision tree, random forest), clustering (i.e., k-mean, dbscan), dimensionality reduction (i.e., PCA, manifold learning) and etc.
- Day-to-day experience with a variety of tools and languages, including bash, Git, SQL, Docker.
- Experience with several Apache ecosystems, including apache airflow and apache spark.

Academic Publications

2021 Jun 29

Predicting indoor PM2.5/PM10 concentrations using simplified neural network models
Journal of Mechanical Science and Technology, volume 35, pages3249–3257 (2021)

2019 Jun

Ventilation Strategy for Acceptable PM2.5 in a Classroom Depending on Building Characteristics
The Society of Air-conditioning and Refrigerating Engineers of Korea (SAREK)
2019.06614 - 617 (4 pages)

2019 Mar

Smart Ventilation for Energy Conservation in Buildings
Journal of Novel Carbon Resource Science, volume 6, issue 1 (March 2019)
Kyushu University, Japan

2019 Jan

Comparison of performance of heat recovery ventilator and air purifier in reducing indoor PM10 concentrations in a classroom
E3S Web of Conferences 111(2):06065

Languages

English
Full professional proficiency

Korean
Elementary proficiency

Bahasa Indonesia
Native proficiency