Resume

2024/07/19

Toshihiro Yamaguchi

Career Summary

Over the past six years, I have been involved in a variety of projects as a data scientist. In my previous role at a consulting firm, I developed skills in leveraging data science to solve business problems. Currently, I am engaged in the development of machine learning systems using Mynaportal data, where I am responsible for everything from requirements definition to system design. I have honed my skills in system design, including security measures, by utilizing AWS. However, there is a slight gap between my interests and my current tasks, as I have limited opportunities to work on machine learning and data analysis. Moving forward, I am eager to contribute to fields focused on machine learning and data analysis, and I am studying system development utilizing LLMs.

Skills

- Machine Learning (DL/NLP/CV)
- Design, development, and continuous improvement of machine learning systems
- Infrastructure (AWS/GCP/On-premises), Web application development including server-side and front-end

Development Environment / Languages / Libraries

- Cloud: AWS, GCP, Databricks
  - AWS: Sagemaker, ECR, Lambda, Athena, S3
  - GCP: Vertex AI, BigQuery, DataFlow, Cloud Run, Cloud Storage
- Languages: Python, R, Scala, Golang
- Tools: Pyspark, Pandas, scikit-learn, TensorFlow, PyTorch, MLFlow, Terraform

Certifications

- Network Specialist
- Passed Information Security Management Specialist Exam
- IELTS 6.0
- TOEIC 905

Events

- MLOps Study Session Speaker
- Databricks Webinar Speaker

Work Experience

Machine Learning Engineer at HJ Holdings

■Implementation of Personalization Features for VOD Services

- **Period**: 2024/01 - Present
- **Role**: Requirements definition, design, development
- **Environment / Methods**
  - GCP
    - Vertex AI Pipelines
    - Vertex AI Experiments
    - Vertex AI Workbench
    - DataProc
    - BigQuery
  - Natural Language Processing
    - Text embedding using BERT trained with Transfer Learning
- **Team Size**: 8
- **Details**:
- **Point**: Responsible for the entire process from designing to developing and deploying machine learning models.

■Effectiveness Evaluation of Recommendation System

- **Period**: 2024/01 - Present
- **Role**: Requirements definition, analysis, reporting
- **Environment / Methods**
  - Tools
    - BigQuery
    - DataFlow
    - GCS
    - Tableau
  - Analysis Techniques
    - Chi-squared test

        * Bayesian inference
- **Team Size**: 5
- **Details**:
  - **Point**: Designed, analyzed, and reported on AB tests to evaluate the effectiveness of the recommendation system.

Data Innovation Lab at JMDC (2023/06/01 - 2023/12)

■Development of Insurance Company Screening System Using Mynaportal Data

- **Period**: 2023/06 - 2023/12
- **Role**: Requirements definition, design, development
- **Environment / Methods**: AWS
- **Team Size**: 12
- **Details**:
  - **Point**: Designed an insurance company screening system using Mynaportal data.
  - **Reason**: Improved profitability by allowing insurance coverage for individuals who would not typically be approved for screening, using Mynaportal data.
  - **Example**: Designed a secure data infrastructure and API using AWS. Responsible for the machine learning system in a project involving three departments.

■Development of Reporting Function Using SageMaker Inference Pipeline

- **Period**: 2023/07 - 2023/12
- **Role**: Design, development, testing
- **Environment / Methods**: AWS SageMaker
- **Team Size**: 5
- **Details**:
  - **Point**: Developed a reporting function for pharmaceutical companies using the SageMaker inference pipeline.
  - **Reason**: Automated a previously manual process for developing reports on dispensing usage to support efficient business decision-making in pharmaceutical companies.
  - **Example**: Automated the entire process from data collection to analysis and report generation using SageMaker. The system is expected to be expanded for other dispensing operations in the future.

EY Strategy and Consulting (2022/09/01 - 2023/05/31)

■M&A Data Analysis

- **Period**: 2023/01

- **Role**: Front-end development, algorithm development
- **Outcome**: Improved M&A success rates and investment target selection efficiency
- **Team Size**: 6
- **Environment / Methods**: Python, FastAPI, Elasticsearch, Logstash, Kibana for data organization, indexing, and visualization
- **Details**:
  - **Point**: Improved M&A success rates and efficiency in investment target selection.
  - **Reason**: The project was conducted to efficiently evaluate the risk and return of new investment targets.
  - **Example**: Organized, indexed, and visualized data using Elasticsearch, Logstash, and Kibana.

■Estimation of Economic Impact of Used Car Dealers in Emerging Markets

- **Period**: 2022/12 - 2022/01
- **Role**: Data Engineer, Data Scientist
- **Outcome**: Confirmed the positive economic impact of used car distribution on emerging markets
- **Team Size**: 8
- **Environment / Methods**: Data collection using UN Comtrade and World Bank Data, statistical causal inference using instrumental variables
- **Details**
  - **Point**: Confirmed the positive economic impact of used car distribution on the economy of emerging markets.
  - **Reason**: The project was necessary to evaluate the economic impact of used car distribution on emerging markets.
  - **Example**: Analyzed the relationship between economic indicators and used car distribution using instrumental variable methods with countries of similar economic scale as the target countries.

■Financial Model Development

- **Period**: 2022/12 - 2023/01
- **Role**: Front-end development
- **Outcome**: Improved accuracy in company valuation and investment decision-making
- **Team Size**: 4
- **Environment / Methods**: Discounted Cash Flow (DCF) method using Excel
- **Details**
  - **Point**: Improved accuracy in company valuation and investment decision-making.
  - **Reason**: It was necessary to accurately calculate company valuations.
  - **Example**: Calculated company valuations using Excel and the Discounted Cash Flow (DCF) method.

Nihon Keizai Tsushinsha (Contract Period: 3 years, 2019/08/01 - 2022/07/31)

■Design and Construction of Data Analysis Infrastructure

- **Period**: 2019/10 - 2022/07
- **Role**: Data infrastructure design, data infrastructure construction
- **Environment / Methods**: Databricks, AWS, GCP, Python, Scala, Spark, PySpark, Golang
- **Team Size**: 7
- **Summary and Achievements**:
  - Collaborated with Databricks to build a data analysis infrastructure.
  - Enabled efficient data utilization and advanced analysis, improving the personalization features of the Nikkei Wave news app.
- **Details**:
  - Responsible for the design and construction of the infrastructure with a team of 7, considering environmental constraints and goals.
  - Utilized AWS, GCP, Python, Scala, Spark, PySpark, and particularly managed MLOps centrally with Databricks.
  - This project significantly improved the efficiency of data analysis, model building, and deployment.

■Development of Recommendation Algorithm

- **Period**: 2020/02 - 2022/07
- **Role**: Algorithm development
- **Environment / Methods**: Jupyter Notebook, TensorFlow, Pytorch, BERT, collaborative filtering
- **Outcome**:
  - Improved app CTR by approximately 5%
- **Details**:
  - Developed a new recommendation algorithm using BERT, TensorFlow, and Pytorch, as simple collaborative filtering had its limitations.
  - This approach improved CTR by approximately 5%, significantly contributing to user engagement and company value.

■Machine Learning System Development and Operation

- **Period**: 2020/06 - 2022/07
- **Role**: Architecture design, cloud development, back-end development, front-end development
- **Environment / Methods**: AWS, Databricks, GCP, A/B testing, CI/CD, Golang, Flutter
- **Details**:

  - Automated operations using AWS, GCP, and Databricks due to the limitations of manual operations and the need for effectiveness verification.
  - Constructed CI/CD pipelines and conducted A/B testing for quantitative evaluation of effects.
  - Also responsible for front-end and back-end development using Golang and Flutter.
  - This significantly improved system quality and performance.

aiforce solutions (2019/02/01 - 2019/07/31)

- Analysis of Supermarket Purchase Trends
  - Period: 2019/04 - 2019/07
  - Summary: Assigned to a major trading company's DX division, receiving consultations from various departments and subsidiaries regarding AI utilization.
  - Role: Data Scientist
  - Environment / Methods: R, association analysis
  - Team Size: 3
- Development of AI Learning Platform
  - Period: 2019/02 - 2019/07
  - Summary: Developed a platform for AI education.
  - Role: Back-end development, front-end development
  - Environment / Methods: AWS, Bottle, Python, React
  - Team Size: 5

GRID (2017/04/01 - 2019/01/31)

- Implementation of Research Papers for Image Modules in DL Libraries
  - Period: 2018/01 - 2019/01
  - Summary: Responsible for the image module of an in-house DL library.
  - Role: Machine Learning Engineer
  - Environment / Methods: SSG, VAE, VGG
  - Team Size: 5
- Front-end and Back-end Development for Image Recognition System Development
  - Period: 2018/01 - 2019/01
  - Summary: Front-end and back-end development for image recognition system development.
  - Role: Back-end, Front-end
  - Environment / Methods: Vue.js, Flask
  - Team Size: 5
- Reinforcement Learning Platform for Plant Control

  - Period: 2018/11 - 2019/01
  - Summary: Responsible for infrastructure in a reinforcement learning project for plant control.
  - Role: Data Engineer, Data Infrastructure Design
  - Environment / Methods: Python, MQTT
  - Team Size: 3

- Fax Image Classification Model

  - Period: 2017/04 - 2018/12
  - Summary: Developed a classification model for fax images and integrated it into a business system.
  - Role: Data Scientist
  - Environment / Methods: OpenCV, Python, VAE, VGG
  - Team Size: 2

Cuon (2016/01/01 - 2017/03/31)

- Creation of Internal Portal Site

  - Period: 2016/12 - 2017/03
  - Summary: Responsible for designing, building, and managing service sites.
  - Role: Web Engineer
  - Environment / Methods: PHP
  - Team Size: 2

- Development of Websites for Startups

  - Period: 2016/12 - 2017/03
  - Summary: Responsible for requirements definition, design, and construction of service sites.
  - Role: Web Engineer
  - Environment / Methods: Ruby on Rails
  - Team Size: 2

AP Communications (2014/05/01 - 2015/12/31)

- Operation and Maintenance of Internal Infrastructure

  - Period: 2014/03 - 2015/12
  - Summary: Responsible for the operation and maintenance of internal infrastructure.
  - Role: Operation and Maintenance
  - Environment / Methods: Windows Server
  - Team Size: 15

- Creation of Internal Infrastructure Operation Design

  - Period: 2015/06 - 2015/12
  - Summary: Responsible for creating internal infrastructure operation design.
  - Role: Design
  - Environment / Methods: ITIL
  - Team Size: 15

Japan Communications (2013/04/01 - 2014/04/30)

- Design and Construction of In-house Service Site
  - Period: 2013/04 - 2014/04
  - Summary: Designed and built in-house service sites using LAMP stack.
  - Role: Design, Construction
  - Environment / Methods: Apache, Linux, MySQL, PHP
  - Team Size: 2