

職務経歴書

2024/07/19

山口 敏弘

職務要約

過去6年間でデータサイエンティストとして多様なプロジェクトに関与しました。前職のコンサルティング会社では、課題解決にデータサイエンスを活用するスキルを身につけました。現職ではマイナポータルデータを利用した機械学習システムの開発に携わり、要件定義から設計までを担当。AWSを活用し、セキュリティ対策も含めたシステム設計スキルを磨いています。しかし、現職でのタスクには自身の興味と少々ギャップがあり、機械学習やデータ分析に関連する業務に携わる機会が少ないです。今後は機械学習やデータ分析に特化した領域で貢献したいと考えており、LLMを利用したシステム開発の勉強をしています。

スキル

- 機械学習 (DL/NLP/CV)
- 機械学習システムの設計、構築、及び継続的な改善
- インフラ (AWS/GCP/オンプレミス)、サーバーサイド、フロントエンドを含む Web アプリケーションの開発経験

開発環境/言語/ライブラリ

- クラウド: AWS, GCP, Databricks
 - AWS: Sagemaker, ECR, Lambda, Athena, S3
 - GCP: Vertex AI, BigQuery, DataFlow, Cloud Run, Cloud Storage
- 言語: Python, R, Scala, Golang
- ツール: Pyspark, Pandas, scikit-learn, TensorFlow, PyTorch, MLFlow, Terraform

資格

- ネットワークスペシャリスト
- 情報セキュリティ安全確保支援士試験合格
- IELTS 6.0
- TOEIC 905

イベント

- [MLOps 勉強会登壇](#)
- Databricks ウェビナー 登壇

ポートフォリオ

[職務経歴書](#)

- フレームワーク: Next.js(Nextra)
- ホスティング: Vercel

[LangChain・LLM による映画推薦システム \(制作中\)](#)

- パイプライン
 - Vertex AI Pipelines
 - LangChain
- LLM

職務経歴

HJ ホールディングス株式会社機械学習エンジニア

■VOD サービスのパーソナライズ機能の実装

- 期間: 2024/01 - 現在
- 役割: 要件定義・設計・開発
- 環境・手法
- GCP
 - Vertex AI Pipelines
 - Vertex AI Experiments
 - Vertex AI Workbench
 - DataProc
 - BigQuery
- 自然言語処理
 - Transfer Learning を用いて学習した BERT によるテキスト埋め込み
- プロジェクト人数: 8
- 詳細:
- Point: 機械学習モデルの学習から推論まで、設計・開発を 1 人で行っている

■推薦システムの効果検証

- 期間: 2024/01 - 現在
- 役割: 要件定義・分析・レポート
- 環境・手法
 - ツール
 - * BigQuery
 - * DataFlow
 - * GCS
 - * Tableau
 - 分析手法
 - * χ 二乗検定
 - * ベイズ推論
- プロジェクト人数: 5
- 詳細:
 - **Point:** 推薦システムの効果検証のために AB テストを設計・分析・レポートを行っている

株式会社 JMDC (2023/06/01 -) データイノベーションラボ

■マイナポータルデータを用いた保険会社の審査システム構築

- 期間: 2023/06 - 2023/12
- 役割: 要件定義・設計・開発
- 環境・手法: AWS
- プロジェクト人数: 12
- 詳細:
 - **Point:** このプロジェクトでは、マイナポータルデータを用いて保険会社の審査システムの設計をしています。
 - **Reason:** マイナポータルデータを活用し、通常では審査をしないように人に対しても保険加入を認め、収益性を向上させます。
 - **Example:** AWS を用いて、セキュアなデータ基盤・API を設計。3つの部署が関わるプロジェクトで、主に機械学習システムを担当しています。

■SageMaker 推論パイプラインを利用したレポート機能構築

- 期間: 2023/07 - 2023/12
- 役割: 設計・開発・テスト
- 環境・手法: AWS SageMaker
- プロジェクト人数: 5

- 詳細:

- **Point:** AWS SageMaker の推論パイプラインを用いて、製薬会社向けの調剤利用数のレポート機能を開発しました。
- **Reason:** 製薬会社が調剤利用数に基づくビジネス判断を効率よく行うためのレポートについて、手作業で開発していたものを自動化しました。
- **Example:** SageMaker を活用してデータ収集から解析、レポート生成までの一連の流れを自動化しました。今後別の調剤を扱う際にこのシステムを展開することが期待できます。

EY ストラテジーアンドコンサルティング (2022/09/01 - 2023/05/31)

■M&A データの分析

- 期間: 2023/01
- 役割: フロントエンド開発、アルゴリズム開発
- 成果: M&A 成功率の向上と投資先選定の効率化
- プロジェクト人数: 6
- 環境・手法: Python と FastAPI, Elasticsearch, Logstash, Kibana を用いたデータ整理・インデックス化・可視化
- 詳細:
 - **Point:** このプロジェクトで M&A 成功率と投資先選定の効率が向上しました。
 - **Reason:** 新しい投資先のリスクとリターンを効率的に評価するためにこのプロジェクトを行いました。
 - **Example:** Elasticsearch, Logstash, Kibana を用いてデータを整理・インデックス化・可視化しました。

■中古車ディーラーが販売先の中進国に与える経済効果の算出

- 期間: 2022/12 - 2022/01
- 役割: データエンジニア、データサイエンティスト
- 成果: 中進国への中古車流通のポジティブな経済影響の確認
- プロジェクト人数: 8
- 環境・手法: UN Comtrade と World Bank Data を用いたデータ収集, 操作変数 (instrumental variable) を用いた統計的因果推論
- 詳細:
 - **Point:** 中進国への中古車流通が経済にポジティブな影響をもたらすことを確認しました。
 - **Reason:** 中古車流通が中進国の経済に与える影響を評価する必要がありました。
 - **Example:** 対象となる国と経済的規模が似ている国を利用した操作変数法を用いて、経済指標と中古車流通との関係を解析しました。

■財務モデル作成

- 期間: 2022/12 - 2023/01
- 役割: フロントエンド開発
- 成果: 企業価値の精緻な算出と投資判断の向上
- プロジェクト人数: 4
- 環境・手法: Excel を用いたディスカウントキャッシュフロー法 (DCF)
- 詳細
 - **Point:** 企業価値をより精緻に算出し、投資判断が向上しました。
 - **Reason:** 精緻な企業価値の算出が必要でした。
 - **Example:** Excel とディスカウントキャッシュフロー法 (DCF) を用いて企業価値を算出しました。

株式会社日本経済通信社 (嘱託契約期間 3 年 2019/08/01 - 2022/07/31)

■データ分析基盤の設計・構築

- 期間: 2019/10 - 2022/07
- 担当: データ基盤設計・データ基盤構築
- 環境・手法: Databricks, AWS, GCP, Python, Scala, Spark, PySpark, Golang
- プロジェクト人数: 7
- 概要と成果:
 - Databricks と協力してデータ分析基盤を構築。
 - 効率的なデータ活用と高度な分析が可能になり、ニュースアプリ日経 Wave のパーソナライズ機能も高度化。
- 詳細:
 - 環境の制約と目標を考慮して 7 人のチームで基盤の設計・構築を担当。
 - AWS, GCP, Python, Scala, Spark, PySpark などを活用し、特に Databricks で MLOps を一元管理。
 - このプロジェクトにより、データ分析、モデル構築、デプロイの効率が大幅に向上。

■推薦アルゴリズム開発

- 期間: 2020/02 - 2022/07
- 担当: アルゴリズム開発
- 環境・手法: Jupyter Notebook, TensorFlow, Pytorch, BERT, 協調フィルタリング
- 成果:
 - アプリの CTR を約 5% 向上
- 詳細:
 - 単純な協調フィルタリングでは限界があったため、BERT と TensorFlow, Pytorch を用いて新しい推薦アルゴリズムを開発。
 - このアプローチにより、CTR が約 5% 向上し、ユーザーエンゲージメントと企業価値に大きく

貢献。

■機械学習システム開発・運用

- 期間: 2020/06 - 2022/07
- 担当: アーキテクチャ設計, クラウド開発, バックエンド開発, フロントエンド開発
- 環境・手法: AWS, Databricks, GCP, A/B テスト, CI/CD, Golang, Flutter
- 詳細:
 - 手動運用の限界と効果検証の必要性から、AWS, GCP, Databricks を用いて運用を自動化。
 - CI/CD パイプラインの構築と A/B テストで効果を定量的に評価。
 - Golang と Flutter を用いたフロントエンド・バックエンド開発も担当。
 - これにより、システムの品質とパフォーマンスが大幅に向上。

株式会社 aiforce solutions(2019/02/01 - 2019/07/31)

- スーパーマーケットの購買傾向の分析
 - 期間: 2019/04 - 2019/07
 - 概要: 大手商社の DX 事業部に常駐し、AI 活用に関する相談を各事業部、子会社から受ける
 - 担当: データサイエンティスト
 - 環境・手法: R, アソシエーション分析
 - プロジェクト人数: 3
- AI 学習プラットフォームを開発
 - 期間: 2019/02 - 2019/07
 - 概要: AI 教育用のプラットフォーム開発
 - 担当: バックエンド開発, フロントエンド開発
 - 環境・手法: AWS,Bottle,Python,React
 - プロジェクト人数: 5

株式会社グリッド (2017/04/01 - 2019/01/31)

- DL ライブラリの画像モジュールの論文実装
 - 期間: 2018/01 - 2019/01
 - 概要: 自社制作の DL ライブラリの画像モジュールを担当
 - 担当: 機械学習エンジニア
 - 環境・手法: SSG,VAE,VGG
 - プロジェクト人数: 5
- 画像認識システム開発の FE/BE 開発
 - 期間: 2018/01 - 2019/01

- 概要: 画像認識システム開発の FE/BE 開発
 - 担当: BE, FE
 - 環境・手法: Vue.js, Flask
 - プロジェクト人数: 5
- プラント制御の強化学習基盤
 - 期間: 2018/11 - 2019/01
 - 概要: 強化学習によるプラント制御 PJ のインフラを担当
 - 担当: データエンジニア, データ基盤設計
 - 環境・手法: Python, MQTT
 - プロジェクト人数: 3
- Fax 画像の分類モデル
 - 期間: 2017/04 - 2018/12
 - 概要: Fax 画像の分類モデルの作成及び、業務システムへの組み込み
 - 担当: データサイエンティスト
 - 環境・手法: OpenCV, Python, VAE, VGG
 - プロジェクト人数: 2

株式会社 Cuon(2016/01/01 - 2017/03/31)

- 社内ポータルサイト作成
 - 期間: 2016/12 - 2017/03
 - 概要: サービスサイトの設計・構築及び PM を担当
 - 担当: Web エンジニア
 - 環境・手法: PHP
 - プロジェクト人数: 2
- スタートアップ向け作成 Web サイト作成
 - 期間: 2016/12 - 2017/03
 - 概要: 要件定義・サービスサイトの設計・構築
 - 担当: Web エンジニア
 - 環境・手法: Ruby on Rails
 - プロジェクト人数: 2

株式会社 AP コミュニケーションズ (2014/05/01 - 2015/12/31)

- 社内インフラの運用・保守
 - 期間: 2014/03 - 2015/12
 - 概要: 社内インフラの運用・保守

- 担当: 運用・保守
- 環境・手法: Windows Server
- プロジェクト人数: 15
- 社内インフラの運用設計作成
 - 期間: 2015/06 - 2015/12
 - 概要: 社内インフラの運用設計作成
 - 担当: 設計
 - 環境・手法: ITIL
 - プロジェクト人数: 15

日本通信株式会社 (2013/04/01 - 2014/04/30)

- 自社サービスサイトの設計・構築
 - 期間: 2013/04 - 2014/04
 - 概要: 自社サイトのサービスサイトを LAMP 構成で作成した
 - 担当: 設計, 構築
 - 環境・手法: Apache, Linux, MySQL, PHP
 - プロジェクト人数: 2