

MY EXPERIENCE



DataDisca

Supermarket Sales

1. Basic Forecasting

- Preprocessing and Analysing Trend, Seasonality and Noise
- Tools Used:
ARIMA
FB Prophet (multivariate forecasting)

2. Data Viz and Analysis



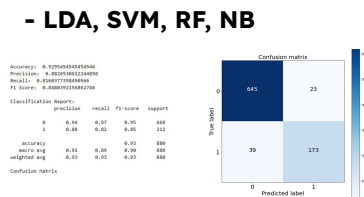
- Created dashboards and interactive charts
- Market Basket Analysis
- Recommender Systems

Packaging Material Plant

1. Predictive Analysis

- Tons of Data
Material
Machine Metrics
QA Metrics
Timestamp Data
Batch Details
Location Data
Weather data
- Data Cleansing and Preprocessing
Rolling window stats

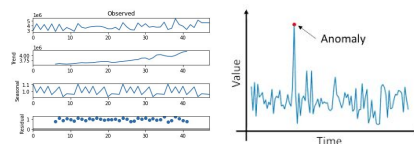
2. Model Training and Evaluation



Industrial Pipelines

1. Pipeline Fluid Flow Analysis

- Clustering
Proximity Based (Euclidean)
Cosine Similarity
Dynamic Time Warping (Out of phase trends)
- 3 Clusters Identified
dynamic time warping
- Time Series Analysis and Anomaly Detection



Computer Vision

1. Image Classification

- ResNet, VGG

2. Object Detection

- YOLO, PyTorch

3. Path Tracing

Curvebeam AI

Data Management Platform

1. DMP

- Building the microservice Architecture for data ingestion and storage
- Tools: AWS Lambda, API Gateway

2. Data Preprocessing

- Preprocessing DICOM images
- Indexing metadata to OpenSearch
- Tools: Numpy, SciPy, SKLearn

3. Data Storage

- S3, DynamoDB, OpenSearch
- ECR

SaaS Infrastructure

1. CI/CD

- CDK
- Automated deployment using Cloudformation and CodePipeline
- Building internal tools for DICOM handling
- Deploying New Services

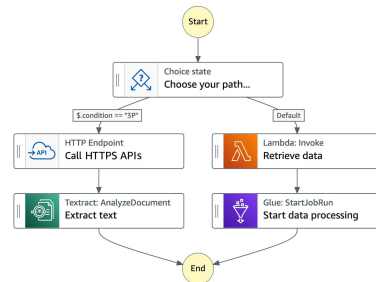
2. Logging and Testing

- Cloudwatch and Datadog
- Test Driven Development
- Unit, Integration and Full Pipeline testing

DevOps and ML

1. ML Orchestration

- AWS Step Functions
- Kubernetes
- ML Pipeline updates and improvements



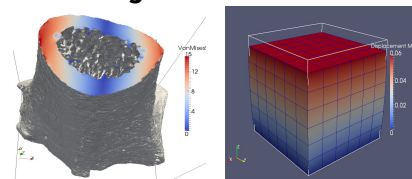
Computer Vis PoC Research

1. FEM

- Research feasibility of analysing stresses in load bearing CT scans

2. DICOM Manipulation

- Manipulating DICOM data
- Converting pixel data to voxel data
- Experimenting with Tetrahedral, Hexahedral and other representations
- Tools: VTK, Gmsh, Paraview, ImageJ

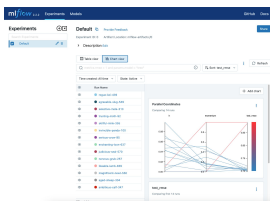


My Projects

MLOps

1. Model Selection

- MLFlow, Hyperopt



2. Model Deployment

- Model Artifact Saving
- Serialization
Pickle, hdf5, gguf, ggml

Machine Learning

1. Feature Engineering

- Keras Feature Spaces
- Feature Crossing
- Imbalanced Data

Credit Card Fraud Data
Banking Data
Diabetes Data

2. Model Training

- Classification and Regression models
- Ensemble models (Libraries and coded)
- XGBoost
- Recommender Systems
Music Recommender
- Mixture Modeling

Deep Learning

1. DL Frameworks

- Anomaly Detection
AutoEncoders
Statistical Analysis
- PyTorch, Keras, Cuda
- Experimentation with parametric models



2. LLMs

- BookGPT
Online Inference
- Testing Lightweight Local Models:
Llama 2, Gemma 2, Phi 2

Statistics and Programming

1. Data Wrangling

- Pandas, Numpy
- Spark
- SQL
- Data manipulation
- Algorithms and Data Structures

2. Tools and Methods

- Built microservices for Curvebeam

- Looking to continue that for my daily tasks

3. Entropy

END

