Microsoft Machine Learning Projects

Project 1: Customer Churn Prediction & Automated Retraining System

Project Overview

Develop a production-grade customer churn prediction system that:

- Ingests streaming customer behavior data
- Builds and serves a churn prediction model
- Implements **automated retraining** based on data drift
- Provides dashboards and alerts for business stakeholders

Milestone 1: Data Collection, Preprocessing, and Exploration

Objectives

• Gather and prepare churn data from various sources (internal CRM, logs, billing)

• Engineer features and understand data quality and imbalance

Tasks

1. **Data Collection**
    a. Collect historical and streaming customer transactional and usage data from SQL databases, logs, or Azure Blob Storage
    b. Ensure comprehensive labels for churn vs non-churn outcomes
2. **Data Preprocessing**
    a. Clean and normalize features (scaling, encoding, null handling)
    b. Address class imbalance (SMOTE or weighting)
3. **Exploratory Data Analysis**
    a. Visualize churn correlates (tenure, usage patterns)
    b. Investigate imbalance and bias issues

Deliverables

- Cleaned and Processed Dataset
- Preprocessing Pipeline Documentation
- EDA Report with insights and issues

Milestone 2: Model Development & Evaluation

Objectives

• Build and evaluate churn prediction models with production focus

Tasks

1. **Model Building**
    a. Implement models such as Gradient Boosting, XGBoost, Random Forest
    b. Perform feature selection
2. **Model Evaluation**
    a. Use confusion matrix, ROC-AUC, precision-recall curves
    b. Validate with cross-validation
3. **Optimization**
    a. Hyperparameter tuning (Grid/Random/Bayesian)
    b. Consider threshold tuning for business impact

Deliverables

- Trained and Optimized Churn Prediction Model
- Model Evaluation Report with business-relevant metrics

Milestone 3: Advanced Techniques & Cloud Deployment

Objectives

• Deploy the churn model as a scalable prediction service on Azure

Tasks

1. **Transfer Learning/Ensembling**
    a. Implement ensemble techniques for improved performance

2. **Azure Deployment**
    a. Deploy via **Azure Machine Learning** as a REST endpoint
    b. Use Docker containers for standardized serving
3. **API Integration**
    a. Build REST API to receive customer data and return churn risk scores

Deliverables

- Deployed Azure ML Churn Predictor
- Integrated API for real-time scoring

Milestone 4: MLOps, Monitoring & Web Dashboard

Objectives

• Implement MLOps best practices for lifecycle management and business integration

Tasks

1. **Experiment Tracking**
    a. Use MLflow or Azure ML tracking for models and metrics
2. **Web Dashboard**
    a. Create a dashboard (Power BI or Streamlit) showing churn predictions & business KPIs
3. **Model Monitoring**
    a. Set drift detection triggers, performance alerts
4. **Automated Retraining Strategy**
    a. Schedule pipelines to retrain the model on fresh data based on drift metrics

Deliverables

- Monitoring & Retraining Pipeline
- MLOps Documentation
- Interactive Dashboard

Milestone 5: Final Documentation & Presentation

Objectives

• Summarize the project and demonstrate full functionality

Tasks

1. **Final Report**
    a. Explain architecture, challenges, business impact
2. **Presentation**
    a. Demo real-time scoring and dashboard insights
3. **Future Roadmap**
    a. Suggest improvements like A/B testing or cross-region deployment

Deliverables

- Final Written Report
- Polished Final Presentation
- Future Enhancements List

Project 2: Real-Time Anomaly Detection for IoT Sensor Streams

Project Overview

Build a production anomaly detection system for streaming IoT sensor data (e.g., manufacturing or smart infrastructure). The system should:

- Detect anomalies in real time
- Scale in production
- Trigger alerts and dashboard updates

Milestone 1: Data Collection, Preprocessing, and Exploration

Objectives

• Collect multi-source time-series streams, clean and prepare

Tasks

1. **Data Collection**
    a. Ingest IoT streams via Azure Event Hubs or Kafka
    b. Store raw data in Azure Blob / Data Lake
2. **Preprocessing**
    a. Windowing, smoothing, feature extraction
    b. Handle missing streams and noise
3. **EDA**
    a. Statistical patterns and correlation analysis
    b. Identify anomaly patterns

Deliverables

- Processed Time-Series Dataset
- Preprocessing Pipeline Doc
- Anomaly Patterns EDA Report

Milestone 2: Model Development & Evaluation

Objectives

• Train models that detect anomalies

Tasks

1. **Modeling**
    a. Evaluate Autoencoders, Isolation Forests, LSTM-based approaches
2. **Evaluation**
    a. Precision, recall on anomaly detection
    b. Real-time detection simulation
3. **Optimization**

    a. Hyperparameter search

Deliverables

- Trained Anomaly Detection Models
- Evaluation Report

Milestone 3: Cloud Deployment & Integration

Objectives

• Deploy pipeline to production

Tasks

1. **Azure Deployment**
    a. Use Azure Stream Analytics + Functions for inference
    b. Containerize model
2. **Scalable API**
    a. Expose REST service for incoming sensor data
3. **Integration**
    a. Alerts via Azure Monitor, Logic Apps

Deliverables

- Azure ML Real-Time Anomaly Service
- API & Alert System

Milestone 4: MLOps & Monitoring Dashboard

Objectives

• Enable ongoing monitoring and automated reactions

Tasks

1. **Tracking & Versioning**

    a. Track experiments and retrains with MLflow
2. **Live Dashboard**
    a. Power BI or Grafana real-time
3. **Retraining Automation**
    a. Trigger retrain based on drift metrics

Deliverables

- Monitoring Dashboard
- MLOps Pipeline Doc

Milestone 5: Final Documentation & Presentation

Objectives

• Complete documentation and demonstration

Deliverables

- Full Final Report
- End-to-End Demo Presentation
- Future Enhancements Proposal

Project 3: Intelligent Support Ticket Classification with RAG

Project Overview

Build a **production-ready support system** that:

- Ingests support tickets
- Uses NLP + vector search + transformer models
- Generates accurate, context-aware automatic responses

This reflects real roles where ML Engineers integrate NLP with production systems and **deliver measurable impact on operations**.

Milestone 1: Data Collection & Preprocessing

Data Collection: Support tickets from databases

Preprocessing: Tokenization, cleaning, embeddings

EDA: Text length, class frequencies

Deliverables: Cleaned corpus, preprocessing doc, EDA report

Milestone 2: Model Development

- Train vector search + transformer models
- Evaluate using metrics like accuracy, recall, BLEU
- Compare traditional vs RAG

Deliverables: Models & Evaluation Report

Milestone 3: Azure Deployment

- Deploy via Azure ML as API
- Integrate vector DB (Azure Cognitive Search)
- Backend service

Deliverables: Deployed API & Integration Doc

Milestone 4: MLOps & Monitoring

- Track experiments (MLflow)
- Dashboard for KPIs
- Retraining triggers

Deliverables: Monitoring Setup, Pipelines Doc

Milestone 5: Final Documentation & Presentation

Deliverables: Final Report, Demo, Future Enhancements

Project 4: Real-Time Anomaly Detection & Alerting System (IoT / Time-Series)

Project Overview

Develop a cloud-based anomaly detection system for streaming IoT sensor data. The system must detect anomalies in real time, trigger alerts, and provide analytics dashboards.

Milestone 1: Data Collection, Preprocessing, and Exploration

Objectives

• Collect real or simulated time-series IoT data

• Preprocess and explore seasonal patterns and noise

Tasks

1. **Data Collection:**
    a. Ingest data from multiple sensors (temperature, pressure, etc.) via simulated streams or public IoT datasets.
    b. Store raw data in Azure Data Lake or Blob Storage.
2. **Data Preprocessing:**
    a. Handle missing values, smoothing, and window segmentation.
    b. Create engineered features (lag features, rolling means).
3. **Exploratory Data Analysis (EDA):**
    a. Analyze trends, seasonal effects, and noise patterns.
    b. Identify baseline normal behavior and potential anomalies.

Deliverables

• Cleaned and Structured Time-Series Dataset

• Data Preprocessing Pipeline Documentation

• EDA Report with Visualizations

Milestone 2: Model Development & Evaluation

Objectives

• Train reliable anomaly detection models suitable for streaming data

Tasks

1. **Model Development:**
    a. Implement models such as LSTM Autoencoders, Isolation Forests, or Prophet-based threshold detection.
2. **Evaluation:**
    a. Evaluate using precision, recall, F1-score for anomaly classes.
    b. Test on held-out test sets and synthetic anomalies.
3. **Optimization:**
    a. Adjust thresholds and retrain pipelines for robustness.

Deliverables

• Trained Anomaly Detection Models

• Model Evaluation Report

Milestone 3: Advanced Techniques & Cloud Integration

Objectives

• Deploy the model in real time on Azure

Tasks

1. **Cloud Deployment:**

    a. Use **Azure Stream Analytics** and Azure Functions for real-time scoring.
    b. Containerize model inference with Docker deployed through Azure Container Instances.
2. **REST API:**
    a. Build a REST API for real-time anomaly scoring.
3. **Integration:**
    a. Integrate with Azure Event Hubs for stream ingestion.

Deliverables

• Deployed Real-Time Anomaly Detection Service

• Real-Time API Endpoints

Milestone 4: MLOps, Monitoring, & Dashboard

Objectives

• Implement production-grade monitoring, alerts, and retraining automation

Tasks

1. **MLOps Tracking:**
    a. Track model versions and experiments with MLflow or Azure ML.
2. **Monitoring:**
    a. Use Azure Monitor and custom metrics for drift and performance monitoring.
3. **Dashboard:**
    a. Build a dashboard (Power BI or Grafana) showing anomaly rates in real time.
4. **Retraining Plan:**
    a. Set up scheduled retraining pipelines triggered by drift detection.

Deliverables

• Monitoring & Alerting Setup

• Dashboard & KPI Visualization

• MLOps Pipeline Documentation

Milestone 5: Final Documentation & Presentation

Objectives

• Summarize and demonstrate end-to-end solution

Tasks

1. **Final Report:**
    a. Document pipeline, deployment, challenges, and business impact.
2. **Presentation:**
    a. Live demo of streaming anomaly detection with alerts and dashboards.
3. **Future Enhancements:**
    a. Propose edge deployment optimization or predictive maintenance features.

Deliverables

• Final Project Report

• Final Presentation Deck

• Future Improvements Document

Project 5: Customer Support RAG-Powered Intelligent Chatbot

Project Overview

Build a production-ready support automation system using **Retrieval-Augmented Generation (RAG)** and vector search. It should serve real queries and integrate with business backend systems.

Milestone 1: Data Collection & Preprocessing

Objectives

• Aggregate support ticket logs, FAQs, documentation, and knowledge base

Tasks

1. **Data Ingestion:**
    a. Collect historical support tickets, product manuals, and FAQ entries.
2. **Preprocessing:**
    a. Clean and unify text data, remove noise, and tokenize.
3. **EDA:**
    a. Analyze common queries, topic distributions, and response effectiveness.

Deliverables

• Processed Text Corpus

• Preprocessing Pipeline Doc

• Support Data EDA Report

Milestone 2: Model Development & Evaluation

Objectives

• Train and evaluate RAG retrieval + generation pipeline

Tasks

1. **Model Building:**
    a. Create vector store (e.g., Azure Cognitive Search or embeddings + vector DB).
    b. Configure RAG model using Hugging Face or similar.
2. **Evaluation:**
    a. Evaluate with metrics like BLEU, ROUGE, and relevance scoring.
3. **Optimization:**
    a. Fine-tune RAG components based on relevance and response quality.

Deliverables

• Trained RAG Pipeline

• Model Evaluation Report

Milestone 3: Advanced Techniques & Azure Deployment

Objectives

• Deploy intelligent chatbot as part of production support

Tasks

1. **Azure Integration:**
    a. Deploy the RAG service via Azure Machine Learning or Azure App Service.
2. **API & Workflow Integration:**
    a. Create REST API and integrate with support portal.
3. **Security & Access:**
    a. Secure endpoints with Azure AD or API keys.

Deliverables

• Deployed Chatbot Service

• Integration & Security Doc

Milestone 4: MLOps & Monitoring

Objectives

• Track performance and automate retraining

Tasks

1. **Experiment Tracking:**
    a. Use MLflow tracking for RAG variations.
2. **Monitoring:**

    a. Monitor accuracy, latency, user satisfaction scores.
3. **Retraining Mechanism:**
    a. Schedule refresh of embeddings and model weights.

Deliverables

• Monitoring Dashboard

• Retraining Pipeline

Milestone 5: Final Documentation & Presentation

Deliverables

• Final Report

• Demo Presentation

• Business KPI Impact Analysis

Project 6: Demand Forecasting & Inventory Optimization Engine

Project Overview

Design a scalable system that predicts product demand using time-series and ML models, and builds inventory recommendations integrated with a dashboard.

Milestone 1: Data Collection & EDA

Deliverables

• Forecast Dataset

• Preprocessing & EDA Reports

Milestone 2: Forecast Model Development

Deliverables

- Trained Forecast Model
- Evaluation Report

Milestone 3: Azure Deployment + API

Deliverables

- Deployed Service
- API Endpoints

Milestone 4: MLOps + Dashboard

Deliverables

- Monitoring & Retraining Automation
- Dashboard

Milestone 5: Final Report & Presentation

Deliverables

- Documentation
- Demo

Project 7: Personalized Recommendation System with Feedback Loop

Project Overview

Build a scalable **personalized recommendation engine** for e-commerce or content platforms that:

- Generates tailored suggestions
- Adapts to user feedback in real time
- Runs on Azure with monitoring and retraining

Real production recommendation systems like those used by Booking.com or similar platforms are central to customer engagement and have measurable business impact.

Milestone 1: Data Collection, Preprocessing & Exploration

Objectives

• Gather user interaction and item metadata

• Prepare unified datasets for training and validation

Tasks

1. **Data Collection**
    a. Aggregate user ratings, purchase history, browsing logs
    b. Include item metadata (categories, tags, content)
2. **Data Preprocessing**
    a. Clean, normalize, and encode features
    b. Create user-item interaction matrices
3. **EDA**
    a. Analyze user behavior patterns
    b. Visualize sparsity and coverage of interactions

Deliverables

• Cleaned Recommendation Dataset

• Preprocessing Pipeline Doc

• EDA Report with insights

Milestone 2: Model Development & Evaluation

Objectives

• Develop recommendation models

Tasks

1. **Model Development**
    a. Content-based filtering, collaborative filtering, matrix factorization
    b. Optionally integrate neural models (e.g., embeddings)
2. **Evaluation**
    a. Use metrics: Precision@K, Recall@K, MAP, NDCG
3. **Optimization**
    a. Hyperparameter tuning and feature selection

Deliverables

• Trained Recommendation Models

• Model Evaluation Report

Milestone 3: Cloud Deployment & Integration

Objectives

• Deploy system as a scalable service

Tasks

1. **Azure Deployment**
    a. Use Azure Machine Learning to host the model
    b. Containerize model with Docker
2. **REST API Service**

    a. Build FastAPI/Flask API for live predictions
3. **Integration**
    a. Connect API with front-end or consumer app

Deliverables

• Deployed Recommendation API

• Integration Documentation

Milestone 4: MLOps & Monitoring

Objectives

• Enable continuous monitoring and retraining

Tasks

1. **Experiment Tracking**
    a. Track models and metrics with MLflow
2. **Monitoring**
    a. Monitor recommendation quality using real-time logs
3. **Retraining Strategy**
    a. Schedule pipeline to retrain based on usage trends

Deliverables

• MLOps Documentation

• Monitoring Setup

• Retraining Pipeline

Milestone 5: Final Documentation & Presentation

Objectives

• Comprehensive report and demo

Tasks

1. **Final Report**
    a. Document methods, results, business impact
2. **Presentation**
    a. Live demo of recommendations and feedback loop
3. **Future Improvements**
    a. Ideas like A/B testing or cross-platform integration

Deliverables

• Final Report

• Presentation Deck

• Future Roadmap

Project 8: Enterprise Feature Store & Model Registry Pipeline

Project Overview

Build an **enterprise-grade feature store + model registry pipeline** to manage features and models consistently across training and serving. Feature stores are a core part of modern production ML workflows.

Milestone 1: Feature Engineering Pipeline

Objectives

• Define reusable features for multiple models

Tasks

1. **Data Collection**
    a. Gather raw data streams and batch sources
2. **Feature Pipelines**
    a. Transform raw data to features using scheduled ETL

3. **EDA**
    a. Validate feature distributions and quality

Deliverables

• Feature Dataset

• Feature Pipeline Docs

• EDA Report

Milestone 2: Model Training Integrating Feature Store

Objectives

• Train models using centrally stored features

Tasks

1. **Model Training**
    a. Use features to train classification/regression models
2. **Evaluation**
    a. Evaluate with appropriate metrics (ROC, MAE)
3. **Optimization**
    a. Hyperparameter tuning

Deliverables

• Trained Models

• Evaluation Report

Milestone 3: Cloud Deployment with Registry

Objectives

• Deploy models with version control

Tasks

1. **Model Registry**
    a. Register models in MLflow or Azure registry
2. **Deployment**
    a. Deploy chosen model version as Azure endpoint
3. **API Integration**
    a. Expose feature + model scoring API

Deliverables

• Registered & Deployed Model

• API Service

Milestone 4: MLOps & Lifecycle Monitoring

Objectives

• Monitor feature quality and model performance

Tasks

1. **Tracking & Logging**
    a. Monitor drift in feature distributions
2. **Alerts**
    a. Set alerts on data and model deviation
3. **Automated Retraining**
    a. Retrain when drift thresholds are met

Deliverables

• Feature & Model Monitoring

• Retraining Pipelines

Milestone 5: Final Documentation & Presentation

Deliverables

• Full Report

• Demo

• Roadmap

Project 9: Automated Explainable AI (XAI) System

Project Overview

Build an ML workflow that includes **explainability dashboards** so business stakeholders can understand model decisions. Explainability is increasingly required in regulated domains.

Milestone 1: Data and Preprocessing

Tasks

- Gather a tabular dataset (e.g., credit scoring)
- Clean and preprocess
- EDA showing feature influence

Deliverables

• Processed Data

• EDA + Feature Analysis Doc

Milestone 2: Model Training & Explainability Integration

Tasks

- Train models (e.g., Gradient Boosting)
- Integrate SHAP or LIME explanations

Deliverables

• Trained Model + XAI Artifacts

• Evaluation & XAI Report

Milestone 3: Cloud Deployment

Tasks

- Deploy model + explanation API via Azure ML
- Build dashboard interface for visual explanations

Deliverables

• Deployed XAI Service

• API & Dashboard

Milestone 4: Monitoring & MLOps

Tasks

- Track model performance and explanation stability
- Monitor for drift

Deliverables

• MLOps Documentation

• Monitoring Setup

Milestone 5: Final Documentation & Presentation

Deliverables

• Report

• Demo

• Future Enhancements

Project 10: Secure & Compliant ML Security Pipeline

Project Overview

Implement a **secure, compliant ML pipeline** for regulated domains (e.g., finance/healthcare), including encryption, audit logs, and secure deployment.

Milestone 1: Data Security & Preprocessing

Tasks

- Collect data respecting privacy compliance
- Encrypt sensitive fields
- EDA on encrypted data

Deliverables

• Secure Dataset

• Security Preprocessing Docs

Milestone 2: Model Development & Risk Analysis

Tasks

- Train models (e.g., fraud detection)
- Risk evaluation with privacy metrics

Deliverables

• Trained Model

• Risk + Compliance Report

Milestone 3: Secure Deployment

Tasks

- Deploy via Azure with Key Vault & secure access
- Harden endpoints

Deliverables

• Securely Deployed Service

• Security Architecture Docs

Milestone 4: MLOps & Audit Monitoring

Tasks

- Track model changes securely
- Create audit logs

Deliverables

• Monitoring + Audit Logs

• Compliance Pipeline

Milestone 5: Final Documentation & Presentation

Deliverables

- Secure ML Report
- Demo
- Enhancement Ideas