Applied Data Science Final Portfolio

Student Name: Rithvik Rangaraj
SUID Number: 220806328
Email: rrangara@syr.edu
Faculty Advisor: Professor Ying Lin (ylin65@syr.edu)

Table of Contents

Overview

This portfolio highlights my mastery of core competencies in Applied Data Science through three major projects. Each project demonstrates the application of data collection, processing, analysis, visualization, and predictive modeling, while addressing real-world problems using machine learning, big data processing, and visual storytelling.

The selected projects showcase my expertise in:
- Applied Machine Learning (Cybersecurity threat detection)
- Visual Analytics (Dashboard design and data storytelling)
- Big Data Analytics (Optimizing supply chains through predictive models)

Learning Goals and Project Deliverables

Throughout the program, I have achieved the following learning goals:
- Apply machine learning models to real-world datasets.
- Develop interactive visual analytics dashboards.
- Efficiently process and analyze large-scale datasets.
- Use predictive analytics for business and operational decision support.
- Communicate technical insights effectively to stakeholders.

Project Deliverables:

- CyberPulse – Predictive Cybersecurity Threat Detection (Applied Machine Learning)
- Rethink the Roads – Vehicle Collision Analysis Dashboard (Visual Analytics)
- Efficient Supply Chain Management for HIV Health Commodities (Big Data Analytics)

Applied Machine Learning: CyberPulse - Predictive Threat Detection

Course: IST 707 - Applied Machine Learning

Learning Goals Addressed:

- Apply machine learning models to cybersecurity datasets.
- Develop predictive models for threat detection.
- Communicate data-driven insights to technical teams.

Project Summary:

CyberPulse leverages machine learning to strengthen cybersecurity defenses. The project analyzed over 40,000 cybersecurity incident reports using XGBoost classification models, achieving an 87% prediction accuracy and an AUC-ROC score of 0.92. Feature engineering

focused on timestamp correlations, anomaly detection, and intrusion signature analysis.

Key Achievements:

- Designed a binary classification model for proactive attack prediction.
- Applied feature selection, hyperparameter tuning, and ensemble methods.
- Analyzed network traffic patterns to detect malicious behavior.
- Developed feature extraction pipelines for enhanced model performance.

Skills Demonstrated:

- Supervised Learning (XGBoost, Random Forest)
- Feature Engineering and Data Preprocessing
- Model Evaluation and Optimization
- Cybersecurity Analytics

Visual Analytic Dashboards: Rethink the Roads - Vehicle Collision Analysis

Course: IST 737 - Visual Analytic Dashboards

Learning Goals Addressed:

- Develop interactive dashboards using Tableau.
- Apply geospatial and predictive analytics for public safety insights.
- Implement dynamic data storytelling for decision support.

Project Summary:

This project explored motor vehicle collision data across New York State from 2018 to 2022. Using Tableau, I designed interactive dashboards that visualized accident severity, crash-prone locations, and influencing factors such as weather, road conditions, and time of day. Predictive modeling through Poisson regression and association rule mining was integrated to enhance decision-making insights.

Key Achievements:

- Designed a four-panel dashboard with user-driven filters and geospatial visualizations.
- Applied calculated fields and parameters for dynamic analysis.
- Conducted spatial, temporal, and predictive analysis of accident patterns.
- Integrated association rule mining to uncover high-risk conditions.

Skills Demonstrated:

- Tableau Dashboard Development and Interaction Design
- Geospatial Analysis and Visualization Best Practices
- Predictive Analytics and Statistical Modeling

Big Data Analytics: Efficient Supply Chain Management for HIV Health Commodities

Course: IST 718 - Big Data Analytics

Learning Goals Addressed:

- Process large-scale datasets efficiently.
- Apply predictive analytics for logistics and supply chain optimization.
- Leverage machine learning models to enhance operational performance.

Project Summary:

Using USAID's Supply Chain Shipment Pricing dataset, this project focused on predicting shipment delays for HIV health commodities. Shipments were classified into on-time, moderate delay, and severe delay categories using Random Forest models. Linear regression was applied for continuous delay prediction. The project proposed actionable logistics improvements based on predictive insights.

Key Achievements:

- Developed a 99.58% accurate model for shipment delay classification.
- Identified key delay drivers including shipment mode, vendor terms, and freight cost.
- Conducted ANOVA and Chi-Squared statistical tests for feature impact analysis.
- Proposed data-driven recommendations for vendor selection and route optimization.

Skills Demonstrated:

- Big Data Processing and Feature Engineering (Python)
- Predictive Analytics and Machine Learning Modeling
- Statistical Testing and Hypothesis Validation

Reflection and Future Learning

Key Takeaways:

- Strengthened skills in machine learning model development and evaluation.
- Mastered interactive visual analytics and data-driven storytelling.
- Applied big data analytics for real-world logistics and operational improvements.
- Gained practical experience integrating predictive modeling with business decision-making.

Future Learning Goals:

- Expand expertise in Deep Learning and Natural Language Processing (NLP).
- Enhance cloud-based analytics skills (AWS, Spark, Databricks).
- Explore real-time data processing and dynamic decision-making systems.
- Deepen knowledge in MLOps and AI deployment pipelines.

This portfolio reflects my journey through the Applied Data Science program, highlighting my preparedness for roles in data science, analytics, and AI-driven decision-making.

GitHub Repository Link for Projects:
https://github.com/rithvikrangaraj/IST782PortfolioRithvik