University of Niagara Falls Canada
Master of Data Analytics
DAMO630 – Advanced Data Analytics
Advanced Research/Industry Data Analytics Project
Due date: Sunday of Week 10, 11:59 PM

Objective and scope

The final group project is designed to provide students with the opportunity to design and execute an end-to-end advanced data analytics pipeline that addresses a real-world research or industry problem. Students will select a problem of interest (with instructor approval), apply at least three course modules, and conduct their analytics using a cloud platform (e.g., Databricks, AWS, Azure, or Google Cloud).

The scope of the project goes beyond technical analysis. Students are expected to:

- Frame a clear business or research problem and connect it to data-driven decision making.
- Prepare, process, and analyze complex datasets.
- Apply advanced methods from multiple course modules to develop meaningful insights.
- Deploy or demonstrate models on the cloud, showcasing scalability and reproducibility.
- Communicate findings effectively through compelling and business-oriented visualizations.
- Provide actionable recommendations that translate analytics into business or policy value.

Project Requirements

1. Problem Definition & Relevance

Select a research or industry challenge with a large-scale datasets belonging to any domain such as healthcare, e-commerce, education, tourism, finance, manufacturing, energy, social media, etc. Clearly state the business or decision-making need your analytics will address.

2. Integration of Course Modules

Use at least three of the following course modules:
- Synthetic Data Generation
- Mining Large-Scale Datasets (Hadoop/MapReduce)
- Mining Data Streams
- Nature Language Processing NLP
- Sentiment Analysis
- Recommendation Systems
- Survival Analysis
- Social-Network Graph Mining

Combine modules in a way that is meaningful for your chosen problem.

3. Cloud-Based Implementation (Mandatory)

All analytics must be conducted on a cloud platform (e.g., Databricks, AWS SageMaker, GCP Vertex AI, or Azure ML). You may use cloud-native datasets, storage (e.g., S3, Blob, BigQuery), and machine learning services. Deployment of models/dashboards in the cloud is encouraged.

4. Business-oriented data visualizations

Apply effective business-oriented data visualizations principles including clarity, audience focus and highlighting key insights for decision-making.

5. Deliverables

- Technical Report: Introduction, problem statement, related work, data preparation and cloud setup, methodology, results, analysis, limitations, business recommendations.

- Business Presentation: Tailored for a non-technical executive audience, focused on insights and business value.

- Code & Documentation: Reproducible Jupyter notebooks or Python scripts that comply with professional coding standards.

Example Project Ideas

- Healthcare: Predict patient survival (survival analysis) + sentiment analysis+ mining large scale datasets + synthetic data generation.

- E-commerce: Build a recommender system + analyze customer churn (survival analysis) + analyze customer reviews.

- Social media: Detect communities (graph mining) + measure sentiment trends + mining data streams.

- Manufacturing: Predict equipment failure (survival analysis) + generate synthetic sensor data + mining large-scale datasets.

- Smart City: Optimize traffic (stream mining) + recommender system for routes + sentiment analysis of citizen feedback.

Project Rubric

	Excellent (90-100)	Very Good (80-89)	Good (70-79)	Satisfactory (60-69)	Failure (F) (<60)
Problem description & business relevance	Clear, insightful problem; strong business relevance	Problem defined; relevance mostly clear	Adequate problem framing; some business connection	Basic problem framing; limited business connection	Unclear or irrelevant problem framing
Integration of course modules	Seamlessly integrates 3+ modules with strong justification	Uses 3 modules with adequate connection	Uses 3 modules with weak justification	Uses only 2 modules or weak integration	Minimal/no integration of course modules
Cloud-based implementation	Full cloud pipeline implemented (data, model, deployment)	Cloud used for most analytics tasks	Cloud used for partial analytics	Partial use of cloud (e.g., storage only)	No or improper cloud usage
Analytics quality & rigor	Thorough, accurate, and innovative analytics	Mostly correct analytics with minor issues	Adequate analytics with some weaknesses	Basic analytics with gaps in rigor	Incorrect or insufficient analytics
Business-oriented visualizations	Clear, compelling visuals; strong storytelling for business decisions	Effective visuals; good storytelling elements	Adequate visuals; some link to business decisions	Basic visuals; limited storytelling or business link	Poor or missing visuals; no clear business connection
Presentation & documentation	Excellent presentation; fully reproducible work; clear, professional documentation	Strong presentation; mostly reproducible; documentation adequate	Adequate presentation; reproducibility with some documentation gaps	Basic presentation; partial reproducibility; minimal documentation	Poor presentation; not reproducible; poor/no documentation