



Assignment 1

Assignment Introduction and Learning Goals

Course Learning Outcomes Addressed

- **CLO 1** - Identify terms and concepts related to machine learning theory specifically applied to data visualization.
- **CLO 2** - Compare and evaluate different evaluation metrics and methods for assessing the effectiveness of machine learning models in the context of data visualization.
- **CLO 3** - Develop and implement reproducible machine learning workflows that incorporate data visualization techniques.
- **CLO 4** - Apply machine learning tools and concepts to effectively visualize and analyze data.
- **CLO 5** - Differentiate variations in machine learning approaches and their impact on data visualization techniques and interpretations.
- **CLO 6** - Critically analyze and extract relevant information from peer-reviewed manuscripts related to machine learning-based data visualization.
- **CLO 7** - Design and execute a project that investigates a problem or research question, utilizing machine learning-based data visualization to analyze and interpret the results.

Assignment Value

40%

Assignment Instructions

Create a comprehensive, interactive dashboard (or static report) that assists machine learning practitioners and data scientists in visualizing and interpreting various stages of a machine learning workflow. The tool should encapsulate exploratory analysis preprocessing, model evaluation and explainability, ensuring that users have a clear view of how their data and models are behaving and where improvements might be needed.

Example requirements for each module:

Exploratory Data Analysis (EDA) Module: Interactive histograms, scatter plots, box plots, pair plots, and correlation matrices.

Outlier detection using visual aids. Missing data visualization to diagnose patterns of missingness and decide on imputation methods.

If using Time Series data Module: Time series decomposition: Display trends, seasonality, and residuals. Autocorrelation and partial autocorrelation plots. Forecasting results and error metrics over time.

Data Preprocessing Visualization: Visual representation of data before and after normalization or standardization. Encoding methods visualization (for categorical variables). Visualization of feature selection and extraction techniques.

Model Evaluation Module: ROC curves, confusion matrices, precision-recall curves, and other relevant metrics for classification. Residual plots, Q-Q plots, and prediction vs. actual plots for regression. Interactive feature importance plots.

Model Interpretability Module: SHAP and LIME interpretations for model predictions. Partial dependence plots. Feature interaction visualizations.

Interactive User Features: Data filtering and zooming. Ability to change visualization types dynamically. Option to upload custom datasets and evaluate custom models.

Implementation: Backend: Python (with libraries such as Pandas, NumPy, Sklearn, PyTorch, SHAP, LIME, TensorBoardX, torchviz, etc.) Frontend: Dash by Plotly or Streamlit for interactive web-based dashboard creation. Challenges: Making sure that the dashboard handles large datasets without much latency. Ensuring compatibility with a wide range of machine learning models. Simplifying complex visualizations for non-technical stakeholders while retaining depth for experts.

Potential Impact: Such a tool can significantly speed up the machine learning model development and troubleshooting process by giving practitioners a bird's eye view of the entire process, highlighting areas of concern, and providing insights into data and model behaviors. Furthermore, by fostering better interpretability, this tool can also promote trust and understanding in machine learning models among a wider audience, including non-technical stakeholders.

Due Date

Refer to the **Course Schedule and Due Dates** located under Course Information in the Content.

How to Submit your Assignment

Present capstone projects orally (including code)

Grading Criteria

View the **Capstone Project Rubric** ([/d2l/common/dialogs/quickLink/quickLink.d2l?ou=1686122&type=rubric&rCode=FANSHAWE-7282964](https://www.fanshaweonline.ca/d2l/common/dialogs/quickLink/quickLink.d2l?ou=1686122&type=rubric&rCode=FANSHAWE-7282964)) to review how this assignment will be assessed.