

## **Project 1 Proposal**

Brian Reppeto

Bellevue University

DSC 680

Frank Neigebauer

December 1, 2024

## **Project 1 Proposal**

Topic: Optimizing of sales forecasting for retail decision making.

This project will predict sales trends for Walmart stores using advanced machine learning techniques, by creating actionable insights for inventory and resource planning.

### **Business Problem:**

Precise sales forecasting is essential for retail businesses to effectively manage inventory, streamline staffing, and enhance profitability. This project will explore how machine learning can boost forecasting accuracy by analyzing historical sales data from Walmart stores to uncover patterns influenced by factors such as holidays and economic conditions.

### **Datasets:**

The primary dataset will be the Walmart sales forecasting dataset from Kaggle. This dataset contains sales data from various Walmart stores across multiple departments, along with features like store size, holiday dates, and markdowns. Additional data, such as economic indicators or weather conditions, may be integrated from public sources like the Bureau of Economic Analysis or NOAA.

### **Methods:**

The project will utilize the following analysis methods:

- Data Cleaning and Exploration: Handle missing values, outliers, and feature engineering.
- Exploratory Data Analysis (EDA): Visualize trends, seasonality, and correlations.

- Machine Learning Models: Use regression techniques (i.e., Random Forest, Gradient Boosting, or XGBoost) to predict sales.
- Model Evaluation: Use metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to assess model performance.

#### Ethical Considerations:

- Bias in Data: The dataset may contain biases that could affect underrepresented store types or regions.
- Data Privacy: Ensure any supplementary data complies with privacy regulations and does not identify individuals.
- Misuse of Predictions: Forecasts should not be used to justify unethical labor or pricing practices.

#### Challenges/Issues:

- Data Quality: Handling incomplete or inconsistent data could impact results.
- Feature Selection: Identifying and engineering relevant features may be complex.
- Model Overfitting: Ensuring the model generalizes well to unseen data.

#### References:

- Kaggle's Walmart Sales Forecasting dataset description and documentation.
- Academic papers on sales forecasting methodologies and retail analytics.
- Online resources for time-series analysis and machine learning techniques.
- Public data sources like the Bureau of Economic Analysis for supplementary economic indicators.

