# ECE 537 Data Mining, Winter 2024

## Project progress report

**Project title:** Analysis and Prediction of Superstore Sales Using Machine Learning

**Students in the project group:**

Vidarshana Govilesh – Tableau Dashboard preparation and preprocessing and 2 ML models

Aswin Gunasekaran – Tableau Dashboard preparation and feature engineering and 2 ML models

**CONTRIBUTION:** Vidarshana Govilesh: Preprocessing

Aswin Gunasekaran: feature engineering

**RESPONSIBILITIES AND REMAINING TASKS**

**Yet to complete**: Vidarshana Govilesh –ML Algorithms and comparison, Tableau charts

Aswin Gunasekaran – ML Algorithms and Evaluation, Tableau charts **ESTIMATED TIME FOR COMPLETION**– 30 days.

**Project Description**:

In this project what we aim to do is to develop data mining algorithm that will help in forecasting sales data.Firstly we plan to begin by utilizing the python libraries like numpy,pandas,matplotlib,seaborn we will analyze and visualize data.Similarly we will also create dashboard in Tableau for visualizing the data and to get better insights using that tool as well.But our main goal here is to create and develop a data mining algorithm that can forecast sales at Superstores. Our objective is to create a predictive model that can accurately forecast Superstore Sales using machine learning approaches, taking into account past data and features. We want to determine the most accurate and dependable algorithm for this prediction by conducting experiments with several models. In order to assess our model, we will use feature engineering approaches and metrics like mean absolute error, root mean square error, and R-squared. Lastly, we'll use Tableau to create a dashboard that shows the data's patterns and trends.

**Data Description:** For this project we will be using the Sales data of Superstore which has attributes related to the sales of their products, order details, product information, and other shipment details. The dataset is sourced from Kaggle. Our dataset has 9800 no of entries and 18 attributes. LINK: <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting/>

**Proposed/Modified Method**: We would predominantly be using Regression ML models like, decision tree, Random Forest, SVM and other models. We would be using Tableau to create a dashboard for visualization and finding patterns, trends in the data and getting insightful information. The Software used for this includesGoogle Colab for Python Machine Learning and Tableau.

**METHODOLOGY:** 