Guest Lecture

MIE223

Winter 2025

1 Data Science at Walmart

1.1 Introduction

Senior data scientist for Walmart Ecommerce.

1.2 Why Data Science

- passion in applied ML projects in undergrad
- Data Scientist at Citibank
- Worked with Sanner on clustering and data science projects

1.3 What I love about DS at Walmart

- creating value for both customers and Walmart
- Witness real-time impact of your work
- Engaging problem statements and innovative environment
- perfect balance of individual contribution and collaboration

2 Data Science in Retail

2.1 Forecasting

- How integral is demand forecasting in retail?
- Inventory planning
- Financial budgeting
- Workforce management
- Waste management

2.2 Personalization

- ML powered recommedation engine
- Product listing to cater to needs of the customers
- Simple cross-sell items based on category relationships

2.3 Product attribute extraction

- why does AI powered recommendation system work better
- Richer product descriptions
- Better understanding of customer needs

3 Day in the life

3.1 Define the problem statement

- meet with business stakeholders to understand their requirements
- Create a rough timeline with a simple deliverable and follow up meeting
- Use case: demand forecasting for Walmart stores
- Business requirement: workforce management
- Alignment: time series model that forecasts for orders per day at the store level

3.2 Data Pipeline

- lot of time goes into data preparation
- Build SQL query worklflows to bring the data into the data lake
- SQL workflows can be very complex
- Requires joining data fro various tables to arrive at a simple tabular dataset

3.3 Exploratory data analysis

- explore feasibility of the problem statement
- understand the data with simple analysis (missing data, cold-start problem...)

3.4 Modelling and Experimentation

- Most exciting phase of the projects
- Try out innovative modelling ideas
- Run experiments to compare models
- Exploration vs producing an MVP

3.5 Running pilot/test programs

- real-time impact of your work
- anxious waiting period where you want to achieve ”better” results
- A successful pilot program leads to a full rollout
- A/B Testing

3.6 Reporting

- monitor the model performance and provide summarized results to the stakeholders
- business emphasizes on explainability of the models
- Business emphasizes on:
  - Explainability of the models
  - Quick results

3.7 Model enhancements

- univariate versus multivariate time series modelling
- can use features like weather or store level intiatives to model the time series