Emily Bridges and Akshay Rangasai Devalla

13 April 2022

15.077 Term Project Proposal

Background:

Walmart operates ~10,500 stores in 24 countries, managing inventory across a wide variety of climates and cultures. Its "Every Day Low Price" motto is the cornerstone of the brand's strategy, and underscores its dedication to helping people save money and live better by bringing value and convenience to consumers around the world. Extreme weather events such as hurricanes, blizzards, and floods, can have a huge impact on the demand Walmart sees at the store and product level, and accurately predicting this demand is crucial in order for Walmart to continue delivering upon its value proposition for consumers.

Objective:

As evidenced throughout the COVID-19 pandemic, consumer behavior can be dramatically influenced by extreme events - remember trying to buy toilet paper or Lysol wipes in spring 2020? Our goal is to accurately predict the sales of various potentially weather sensitive products around the time of major weather events. Although our analysis will focus on forecasting demand surrounding extreme weather-related disruptions, in the future, the research could potentially be further expanded upon to encompass the impact of non-weather related events, such as holidays and anticipated lockdowns.

Data:

In our model, we will look at sales data for 111 different products across 45 different Walmart locations and weather data from 20 different weather stations covering the 45 Walmart locations in question between January 1, 2012 and January 1, 2015. We will define an extreme weather event as a day in which more than an inch of rain or two inches of snow was observed, and will predict the units of each item sold over a window of three days before and after the storm. The data we will use comes from Kaggle (https://www.kaggle.com/competitions/walmart-recruiting-sales-in-stormy-weather/data).

The store dataset contains three fields: the store id, the item id, and the date. The weather dataset contains 18 fields including the date, the weather station id, the min and max temperature, the amount of snowfall, the total amount of precipitation, and the wind speed. Finally, we have a key matching each of the stores to a nearby weather station.

We will use various machine learning methods such as linear regression, CART trees, random forests, and boosted regression to predict the demand for each product over time.