Final Project: Proposal

Team: Deep Learners

Ria Gupta, Sachin Kumar, Abhishek PV, Ishan Sahay

1. List 3 questions that you intend to answer (1 point)

1.) What and how the various features like holiday, temperature have on individual sales prediction (in time series) at a particular aggregation level for a retailer?
2.) How to provide a forecast at various aggregation levels in a hierarchical time series for merchandise planning?
3.) How to make this forecast automatic, consistent and optimized in nature?

2. List all the datasets you intend to use (1 point)

The dataset of Canadian fashion retailers will be provided by MYFIND.AI (3-4 years of anonymized transaction data). We have not seen the data yet.
We might add data related to retail stores that contain some features not mentioned in provided dataset but can provide some hidden insights (variables like temperature, holiday days).
We expect to obtain the dataset soon.

3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)

- Data Collection: Where/how to get data
  Data (a csv file) will be provided by MYFIND.AI.

- Data Exploration: Do you need to conduct EDA in order to understand the data?
  Yes.
  a) Missing Values: to see if all values of data are provided for the duration of 3-4 years. If there are missing values present, what percentage they account for and how to best impute them.
  b) Time Series Components: We expect to see time series components like trend, cyclicity, seasonality and non-stationarity.
  c) Feature Analysis: Perform some feature engineering, manual feature selection and analyze correlation of various features.
  d) Output Variable Analysis : What is the distribution of sales wrt to various relevant features, is it linear or non- linear?

- Data Cleaning: Do you need to clean data? How to clean them?
    a) We have not seen the data yet, but we assume not to perform huge cleaning of the provided data. However, after integration with other data, we might have to perform cleaning.
    b) This also involves removing unwanted and duplicate observations as it is a transaction based data. As we would be aggregating by product, temporal and spatial levels, we need to make sure that there are no missing or duplicate values.
    c) Transformations on data like applying log transformation, order differencing to make time series stationary. Making the residuals or white noise normalized.

- Data Integration: Do you need to integrate data from multiple sources?
  Yes we might have to fetch temperature data from an API (e.g. OpenWeatherMap API - Free Weather Data (for Developers) and factors like holiday data from API (e.g. https://calendarific.com/api-documentation).

- Data Analysis: Do you need to analyze data? What analysis do you intend to do? (e.g., SQL, Statistics, Deep Learning) How to evaluate your analysis results? (e.g., evaluation metrics, confidence intervals, benchmark)
  Yes.
  Potential Approaches for analysis:
    a) (Hierarchical) Time Series Analysis via top down, bottom up and middle approach. Application of ARIMA, SARIMA models at various levels. Selecting best parameters(p,q,d) for these models via hypertuning or grid search.
    b) Statistical analysis like ADF test, rolling statistics to analyze stationarity.
    c) Analysis and modelling via HTS prophet method in python.
    d) Potential use of deep learning techniques like LSTM.

  Potential Metrics: MSE, MAE, MAPE
  Potential Analysis:
  What effect do various feature columns like holiday, temperature have on predictions?
  Which have better sales- online or offline transactions? (Assuming we have that information)
  Which regions, products perform best and worst and at what time of year?

- Data Product: What data product do you want to build? (e.g., visualization, an interactive web app, a report, a model)
  Model(s) that forecasts hierarchical time series. Report of our findings, which models/parameters are optimal for a given aggregation level.

4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)

Provide an optimized idea of future demand of fashion products. Help retailers plan when, how and how much quantity of a product to stock to optimize their sales.