# Statistics

## Discrete and continuous distributions

## Conditional probability and Bayes theorem

## Hypothesis testing and confidence interval estimation

## Random variables

## Simple Linear Regression

## Generalized Linear Models

## Non-Parametric Methods

## Causal Modelling

# Foundational:

## Linear Algebra(PCA, SVD, NMF)

## Data Structures and Algorithms

## Stacks,

## heaps,

## queue,

## linked-list,

## hashing,

## tree and graph search

## Distributed Programming

## Object Oriented Programming

## Object-oriented programming (OOP) is **a programming paradigm based on the concept of "objects"**, which can contain data and code: data in the form of fields (often known as attributes or properties), and code, in the form of procedures (often known as methods).

## there are four fundamental concepts of Object-oriented programming – **Inheritance, Encapsulation, Polymorphism, and Data abstraction**.

# Supervised

## Decision Tree (CHAID, C4.5, CART)

## SVM

## Boosting

## Bagging

# Unsupervised

## Clustering

## K-Means,

## hierarchical clustering,

## EM clustering,

## Density-based clustering

## Factor Analysis

# Time Series

## Standard Forecasting(Exponential Smoothing)

## ARIMA

## FbProphet

## Forecast XGB

## VAR (Vector Autoregressive Model)

## Panel Regression

## DTW (Dynamic Time Warping)

Project EDA:

1. Seasonality Shrinkage using Fourier analysis
2. Mean Shift Analysis
3. Neighborhood analysis : Correlation amongst neighboring countries incorporated
4. Velocity Resolution(Vr) is used for People Tracking and Counting" configuration:  L / (2Tf) where L is the wavelength and Tf is the frame period.
   1. Maximum velocity ( V(m) ) = L / (4Tc) where Tc is the duration from one chirp to the other
   2. And Velocity resolution should be: ( V(r) ) = V(m) / numLoop

Ref : <https://e2e.ti.com/support/sensors-group/sensors/f/sensors-forum/978625/iwr6843-maximum-velocity-and-velocity-resolution-calculation-in-mixed-chirp-frames>

1. As market events might have diminishing effects we need to take care of them accordingly
2. Tease out effect of holidays on the data
3. Product Effects:
   1. Change in Prices,
   2. If there are any rebates,
   3. Time from product launch/ Loss of Exclusivity
   4. Market Share
   5. Stocking patterns
   6. Sales Effort(Lagged Effect)
   7. Contracting laws( with payers, distributors)

# Deep Learning

## Linear Perceptron Models, Back Propagation and levenberg marquardt rules

## CNN

## RNN, LSTM, GRU

## GANs

## Attention Mechanism

## Auto Encoder

# NLP

## Tokenization, stemming, stop words, lemmatization

## Information Retrieval - POS tagging, NER, TF\*IDF \*

## Text Similarity Measures (Fuzzy matching, Edit Distance, etc.) \*

## Search applications (indexing, facets, annotations) \*

## Topic detection (such as Latent Dirichlet Allocation (LDA), Latent Semantic Indexing (LSI))

## Text Summarization

## Deep Learning Methods - word2vec, lda2vec, doc2vec

## Speech2Text

## BERT, Transformers, GPT Models

## Language Mode Training

## QnA Chatbots

# Optimisation

## Linear Programming

## Simplex Method and Mixed Integer Programming

## Lagrange Multiples and Duality

## Gradient Descent and SGD

## Advanced optimization methods (Stochastic optimization, Robust optimization, Non-linear Programming)

## Evolutionary (Genetic Algorithms, Simulated Annealing, Tabu search, Ant Colony, ...)

## BFGS, Conjugate gradient descent, quasi-newton methods

# Graph Mode and Simulations

## Graph Clustering( modularity community detection)

## Monte Carlo Simulation

## Probabilistic Graphical Models - Markov models, Hierarchical Bayesian networks, CRF \*

## Page Rank, HITS \*

## Graph based CNN \*

## Probabilistic Circuits \*

## Graph Based Representation Learning \*

# Recommender Systems

## Memory based collaborative filtering (KNN-based) \*

## Model based collaborative filtering (Matrix factorization) \*

## Hybrid collaborative filtering \*

# Semi Supervised Learning

## Pseudo-label generation based \*

## PU learning \*

## Co-training Models \*

# Reinforcement Learning

## Model Based Learning

## Model Free learning

## Planning Network

# Solutions

## Container Based Solutions (Docker, Kubernetes, etc.)

## MLOps Frameworks (MLFlow, Seldon, Azure, etc.) \*

## CUDA programming

## Dash, streamlit

# Technologies

## Cloud Technologies (AWS, Google Cloud Platform, etc) \*

## Tableau

## Web Scraping (beautiful soup, Selenium)

## Dataiku (DSS)

## Altreyx

## HTML, CSS

## Flask

## Django

## Pytorch, FastAI

## Tensor Flow Keras

# External Tools

## Ablation Analysis

## Language Interpretability Tool (LIT) \*

## Automated Feature Engineering Feature Tools

## Automated Model Selection Auto ML

## Explanation Models (SHAP, LIME)

## Project Summaries

|  |  |
| --- | --- |
| **Business Problem** | Enhance existing demand planning team for better forecasting |
| **USP** | Created ML engine integrated with automatic dataflow from/to database, customized algorithms and error metrics  Predict SKU or Segment level demand using certain driver metrics (volatility, medical benefit etc.) and techniques such as ARIMA or AI/ML Optimize the inventory using a multi-tiered objective function with a forward looking projection derived from the SKU demand prediction |
| **Business/Client Impact** | Enhanced current process via   * automation, (automated choice of best features and algorithms) * Max accuracy (through cross validation, ensembles and multivariate model) * reduced turnaround time and * increased efficiency |
| **Data** |  |
| **Skills Learnt** |  |