

Dhiraj Gangaraju

dhiraj.gangaraju@walmartlabs.com | +91-9700281995 | LinkedIn: [Dhiraj Gangaraju](#)

EDUCATION

M.MGT | IISc, BENGALURU | BUSINESS ANALYTICS | 8.1 CGPA | 2016-2018

B.TECH | CBIT, HYDERABAD | PRODUCTION ENGINEERING | 81.5% | 2008-2012

EXPERIENCE

DATA Scientist | WALMART GLOBAL TECH | AUG 2018 – PRESENT

- Incrementality analysis

Calculated the incremental effect of using a Walmart service on the shopping behavior of customers. The shopping behavior is captured through various KPIs like order frequency, average spends, average basket size etc. The lift in the KPIs is computed by creating a test and control setup where the control is the set of customers who do not avail the service in question.

The control set is a lookalike of the test set in terms of the KPIs. And the lookalike control set is selected using the algorithmic matching procedure called covariate matching. This is done for customers falling in different shopping channels like store only, online only, and omni.

The analysis is customized for a multitude of services like Walmart Pay, Money Services, Express Returns, Store Maps etc. The consummation of the analysis for business stakeholders is the incremental metrics served through an interactive tableau dashboard that is refreshed on a monthly cadence.

The entire process from data extraction to growth calculations is automated for every service by building a data pipeline using the Walmart's inhouse machine learning platform called Element. The output tables from the platform are connected to the tableau server for refreshing the dashboard.

A 2% lift was observed on the order frequency of customers because of the use of Walmart pay.

- Next basket prediction

Online grocery is one of the best performing products of Walmart with 3.1M customer base and \$2.5B in sales. An average customer using online grocery spends \$114 in a basket of 26 items taking ~30 min to build the basket.

The business case of the project is to reduce the basket building time of customers by sending them product recommendations. This is enabled by predicting the products the customers would purchase next.

This business case is framed as a machine learning problem (prediction/classification). The goal is to predict if an item will be bought in the next transaction, basis the historical purchase pattern of the customer.

The technique used was Gradient boosted machines (GBM) – which usually fares well in problems such as these. A variety of drivers that may influence this prediction was created. The key genres of drivers were frequency (# of transactions where item was bought), how recently was the item bought by the customer, time between successive transactions where item was bought etc.

The model returned a prediction accuracy of ~50% when predicting 30 items and the accuracy is ~80% when predicting 100 items.

- Long term revenue forecasting

Markov chains are examples of stochastic processes which work on the concept of the Markov property - the future depends only on the immediate past. It is assumed that the transition probabilities between different states becomes constant after a certain point in time.

This concept is used to forecast the revenue generated by customers shopping online in Walmart.com. The entire online customer base is divided into buckets (states) based on the number of online transactions a customer does in the past 52 weeks. Naturally, customers transition from one state to another state across different periods. It is observed that the order distribution (proportion of orders in each state for each period) of customers arrives at a steady state after a few periods. Period here is defined as the frequency for which we are forecasting – monthly/ yearly.

When the steady state occurs, the order distribution at the steady state along with the transition probabilities can be used to forecast the revenue. A -4% MAPE was obtained using this technique when predicting for two years into the future.

- Portfolio analysis

In the digital acceleration ecosystem of Walmart, there is a suite of 12 products which customers interact with. The business had set out to examine the entire DA portfolio to understand how to drive user adoption by leveraging the most used product. This was to measure: which product in the DA portfolio is the first product that customer onboards to? Which products in DA portfolio drive adoption?

It was discovered that a set of core DA products – Store Assistant, Store Maps, Scanner serve as a gateway to the DA ecosystem.

- Markdown Optimization

Excess inventory left at the end of season leads to increase in costs from wastage, disposal, labor cost, shelf space and decrease in overall sales. An optimal Markdown Strategy is essential to plan the markdowns effectively to reduce wastage while minimizing revenue loss.

Markdown Optimization is a solution to recommend the right markdowns at appropriate period at relevant stores, to drive inventory sell-through with maximum revenue. The solution involves two components – Base Demand Forecast and Markdown Elasticity.

Machine Learning Models like XGBoost are used to accurately forecast Base Demand at regular price based on Historical sales patterns by capturing all demand influencing factors like Seasonality, Trend, Holidays, Promotions, Promo Supports, Weather, etc and lag sales features like sales in the recent weeks, sales during same weeks during last year, etc.

Markdown Elasticity of Items are estimated using Log-Log model to capture effect of Price change on Volume based on the Historical Price vs Volume Relationship, Markdown and Sales Data, also accounting for Seasonality, Trend and Product Lifecycle stage. A time-series based clustering methodology (K-shape clustering) is used to group similar stores to account for sparsity of transaction and markdown data at store level.

The Base Demand Forecast and Markdown Elasticity solution in conjunction provide accurate Markdown Demand Forecast for upcoming 16 weeks at item-store level.

The performance of the Markdown recommendations through the weeks is monitored using a Governance framework.

PROJECT INTERN | PHILIPS | JAN 2018 – JUNE 2018

Project title:

To study the impact of demographic factors on the sales of Philips LED products in the US market.

Techniques used:

Regression, Clustering, Decision trees, etc.

SUMMER INTERN | ERICSSON | MAY 2017 – JULY 2017

- IT Usage Analysis

Analyzed the IT spending data and identified cost-reduction opportunities. Identified and suggested licensing and packaging models

- Survey Analysis

Analyzed Ericsson's quarterly survey data through text mining and provided insights which could improve the efficiency of workplace. Identified areas of improvement in Ericsson's IT service area across locations using Decision Trees

- Social Media Analytics

Analyzed Ericsson's Yammer platform to find out the reasons behind employees' satisfaction and dissatisfaction levels. Used Regression Analysis to identify factors which could explain the

variability in satisfaction. Identified the geographies and functional areas where the management must concentrate to improve employee engagement

- Web Scraping

Used the python's Beautiful Soup library to scrape the web content of Ericsson's customers to find insights into their business, so that Ericsson can serve them better

PROGRAMMER ANALYST | COGNIZANT | DEC 2012 – MAY 2016

- Responsible for leading a small team in delivering test results to the customer in the US healthcare sector
- Responsible for status report updates to stakeholders during internal as well as client meetings
- Involved in reviewing and validating the test scripts execution of the team members and sending the status report to the onsite POC
- Closely collaborated with the Dev teams throughout to understand and close the defects
- Involved in Requirement Analysis, Requirement Review, Test Planning, Test case development, test case Walk through, reviews, and defect analysis
- Worked on setting up and running automation test scripts (smoke, regression, and shakeout) in selenium. Worked on RC and Web driver frameworks
- Worked on Requirement Analysis, Smoke testing, System testing, and Regression testing
- Also worked on EDI testing, cross verification with the database, and high-level XML validation
- Resolved over 300 defects which included more than 80 performance bugs
- Also resolved more than 50 ad-hoc business queries with respect to complex domain components
- Worked on a Production-like environment set up project with applications dealing with numerous third-party components

TECH STACK

- Python, R, SQL, Hadoop, Hive, PySpark, MongoDB, Tableau

ACHIEVEMENTS

- Received the SPOT award at Walmart for Q3 2019
- Created the tableau dashboard that was used by the winning team in the Online Grocery hackathon in April 2019
- The analysis on Walmart Pay found its way to the CTO's newsletter in Dec 2018
- The insights from the internship at Ericsson was published in its Q2 bulletin 2017
- Received the quarterly "Best performer" award twice in Cognizant
- Ranked university 8th in Engineering
- Earned a silver medal in engineering for "Excellent Academic Achievement"

EXTRA-CURRICULAR

- Playing national level "Ultimate Frisbee"
- Served as the Placement Coordinator at IISc
- Served for Cognizant Outreach and Cognizant Go Green
- Worked as an "Organizer Trainee" in Great Hyderabad Adventure Club (GHAC)
- Served as the Student Representative for National Social Service (NSS) in CBIT
- Contributed to team building as a chairperson in the American Society of Mechanical Engineers