12th Symposium on Business Analytics and Intelligence

BAI – Batch 12 (2021 – 22)

BOOK OF ABSTRACTS

LIST OF PROJECTS

Sl.No	Project Title	Pg No
1	Alarm Intelligence: Enhancing Alarm Systems Through Machine Learning	2
2	Demand Forecasting for Elective Surgeries at IHH hospitals	3
3	Digital Touchpoint Attribution in Path to Purchase & Spend Optimization	4
4	Emerging and Challenger FMCG Brands in the Indian Market	5
5	Hardness Classification in SAGD Oil Extraction using Soft Sensors	6
6	Kimberly Clark Master Data Analytics	7
7	Reducing Rate of Clinical Indicators at Global Hospitals	8
8	Reducing Laboratory turnaround times for Manipal Hospitals using Prescriptive Analytics	9
9	Ticket Sales Forecasting for Private Tour Operator	10

PROJECT TITLE: Alarm Intelligence: Enhancing Alarm Systems Through Machine Learning

ABSTRACT: Alarm management is necessary in a process plant and is controlled by an operator through a control system such as DCS (Distributed Control System). Alarms are generally triggered when the process value deviates from the given set point or normal operating conditions. The operator manually intervenes under such conditions to bring back the system to normal conditions.

The operator's performance diminishes when the annunciated alarms at any given point of time is more than the operator's handling capacity. An extreme stage of this is observed during "alarm floods" and "alarm chattering", with hundreds of alarms getting triggered at the same time due to myriad factors. They negatively impact the performance of the operator. If the operator is unable to make timely interventions, it can cause disruption in operations and may result in plant downtime.
The project aims to develop an algorithm-based solution for an oil extraction and process plant which can enable the following:

1. Minimize Alarm Chattering
2. Minimize Alarm Floods
3. Alarm Forecasting & Causality

The project aims to enhance plant operator performance through improved alarm management by minimizing alarm chattering and floods through ML based approach.

Minimizing Alarm Chattering/Flooding would directly result in reduction of alarms that get triggered in the plant. This would enable lesser distraction for the Plant Operator and enhance their performance. This will also reduce the threat of disruptions in plant operations thereby ensuring minimal revenue loss for the business.

Since the dataset consists of high-frequency time-series entries, the possible approach would be to start with Logistic Regression and leverage Sequence Classification algorithms such as LSTM-RNN

PROJECT TITLE: Demand Forecasting for Elective Surgeries at IHH hospitals

ABSTRACT: As healthcare systems around the world scrambled to deal with multiple waves of COVID-19, patients were forced to postpone elective surgeries as much as possible, emphasizing that any surgical procedure carries the risk of spreading COVID-19 infection. Elective surgery is a medical procedure that can be planned, if it is highly imperative, but it is not an emergency. These procedures do not mean they are not necessary, as patients may deteriorate if care is delayed for an extended period. Liver transplant, Gallbladder stone removal, angiography or stenting for cardiac patients, and resections for early-stage cancer are a few examples of elective surgery.

IHH Healthcare is one of the world's largest healthcare networks, with 80 hospitals in 10 countries and they have witnessed a decline in elective surgery during the pandemic. IHH expects that the demand for elective surgeries will steeply increase as covid de-intensifies after third wave.

The objective of this project is to forecast demand for the top elective surgeries that would help the management with the decision support for resource planning, where IHH can scientifically come up with mechanism to allocate critical medical resources efficiently and reduce staff redundancy. Insights on the demand will also help them in deriving new package pricing to conform to the expected footfall post covid.

This project aims at building a ML forecasting model for top 10 elective surgeries based on the available historical data of packaged elective surgeries. The forecast would be made for top elective surgeries individually, for given regions.

Our approach would be to experiment with various forecasting ML Models like ARIMA/LSTM for analyzing and predicting the demand for elective surgeries Quarterly.

PROJECT TITLE: Digital Touchpoint Attribution in Path to Purchase

ABSTRACT: Digital touchpoint attribution is a process of assigning credit to marketing channels (Organic search, paid Search, social media, and Email) that influence a customer's buying decision. Example – impact of marketing spends in Google search campaigns in driving online conversions. The insights from digital touchpoint attribution enable organizations to take data-driven decisions on digital marketing strategies and budget allocation across various digital channels.

The first step of any attribution project is to capture & understand customer navigation across digital touchpoints on the web during their buying process. This is achieved via Clickstream Analytics (the process of collecting, analyzing, and reporting aggregate data about which pages a website visitor visits and in what order during their buying journey)

The scope of this project is to understand the buying journey of website visitors for a hair removal brand and develop a regression-based algorithm to estimate the relationships between factors that influence a customer's decision to purchase a specific hair removal brand.

The factors/independent variables include -websites visited, brand & category search terms used in the search journey, brand pages visited, product detail page visited, reviews & deals pages visited, organic pageview, paid pageviews. All these variables need to be feature engineered from Clickstream dataset provided.

Another part of the project is to understand the customer demographics & segment/cohorts that drive conversions. This information is useful for brands to understand the important segments and derive their target audience for future marketing campaigns.

The project aims to have an analytical science of determining which digital marketing strategies are most effective for a brand and how brands should allocate their budgets on paid campaigns (spend more on brand search or category search keywords), SEO optimization to drive organic visibility on search engines, influence more customers to provide feedback/ reviews and other marketing activities that influence purchase thereby driving conversions & revenue to business

PROJECT TITLE: Emerging and Challenger FMCG Brands in the Indian Market

ABSTRACT: In recent years, Indian FMCG market has been seeing acceptance and higher growth of new, emerging and challenging brands (to conventional players) on account of consumer purchase behaviour changing due to external and societal changes. S&D channels are confronted with unique problem of identifying brand potential before they are collaborating with such brands.

In an order to identify brand potential and Go to Market strategy for S&D channels, it is imperative to analyse current market trends along with demographic and technographic data of the sales region. Currently such data research is available at India level with limited information and thus curtail ability to scientific method of choosing successful brands for future growth.

This project aims to bridge this gap by market analysis, finding brand wise common attribute of success and potential of various F&B micro segments for specific sales region of the client. This project will attempt to build a model which can identify brand success potential using available internal data, third party research and socio-economic trends. Which in turn will give market study of specific segments, regions and brand category to client for further GTM strategy preparation.

Various statistical tools such as market basket analysis, regression, brand wise cluster analysis, model development using Python, validation study and overall strategy development are currently deployed for the given project and common attributes of successful brands are being identified.

By adding element of scientific research and analysis, the aim is to optimize and increase the efficiency of brand potential for success in given sales regions, micro segment and GTM which in turn will increase organisation's ability to find, collaborate and work with potential high revenue brands and increase organisation presence in their market portfolio.

PROJECT TITLE: Hardness Classification in SAGD Oil Extraction Using Soft Sensors

ABSTRACT: SAGD is extremely water-intensive and thus proper treatment is essential to eliminate corrosion and scale build-up. The cost of ignoring or guessing about water quality can lead to costly piping failures and downtime. Given this, it is of significant business value to build a soft sensor to estimate whether the water is hard or not.

Analytical measurements such as pH and hardness have traditionally been done in the lab if they were performed at all. However, as this is a variable which is measured by a lab test, operators often prefer to measure them only once or twice a day. Therefore, the dependent variables consist of irregularly sampled data with a significant class imbalance. On the other hand, Xs are a mix of irregularly and regularly sampled data with different sampling frequencies. So, the challenge is to be able to determine Y in the absence of too many known values.

The approach is to develop soft sensors to accurately classify hardness of water. Soft sensors are virtual sensors that employ existing measurements to calculate the unknown values of a quantity of interest. The soft sensor will use time series data obtained through physical sensors and irregularly sampled lab tests. Sensor data such as temperature, pressure, flow of water is sampled at varying frequencies at different stages of the treatment process. Thus, our approach would be to develop a soft sensor using ensemble techniques such as Random Forest, SVM, XG-Boost as well as logistic regression and compare the performance of the model

PROJECT TITLE: Kimberly Clark Master Data Analytics

ABSTRACT: Kimberly Clark has created a data lake to drive data culture in the organization.

The data lake will contain all possible commercial data including internal sales (both primary and secondary), offtake from key customers (both online and offline), external syndicated data including market share information, eCom on-platform execution data (content, rating & reviews, inventory, share of search, pricing) and master data.

Some of the KPIs are present in internal data (e.g. sales, profitability) while others are sourced through external data (e.g. market share, share of shelf). Hence, to look at external and internal KPIs together there is no way unless we do a master data mapping. For e.g. to look at the impact of share of shelf with sales we would need to map external brand to internal brand definition. Similarly, there are multiple use cases for which this need arises across innovation, revenue growth management, financial planning & analysis.

Today this mapping is done manually by someone on an ad-hoc basis leading to gaps (90 out of 507 hierarchies are NOT mapped). This impacts data analysis leading to gut-based decisions thereby impacting sales/profitability for the company.

The task is to automate this mapping using AI/ML so that there are minimal gaps (upto 90% automation expected). We are planning to use below options to do the same:

- RASA open source
- Build our own NLU intent classifier
- Auto correct

PROJECT TITLE: Reducing Laboratory turnaround times for Manipal Hospitals using Prescriptive Analytics

ABSTRACT: Turn Around Time (TAT) is a key KPI to define performance of a laboratory. It is defined as the amount of time to fulfil a request or complete a process. The Medical Industry is very time-sensitive and many diagnoses and courses of actions for ailments are dependent on the timeliness of results of the said tests.

Delays in TAT can critically impact patient satisfaction. Each unit time being consumed in any of the steps of a given request, can also be seen as an opportunity cost for a different request.

Process Classification: A typical laboratory test process includes **Pre-Analytical:** initiation, sample collection, **Analytical:** sample testing, verification, and subsequent **post-Analytical:** report generation

TAT Classification can be done by *test type* (ex: CBC, Blood potassium levels), *degree of criticality* (Urgent or Routine Check-up), *consumer type* (in-patients, out-patient). It is observed that the largest TAT variation occurs in the analytical stage.

The Aim of this Project is to optimize/improve the Blood Testing Process of Manipal Group of Hospitals, by reducing the TAT. This in turn will increase profitability by reducing the cost of testing incurred by the hospital. Identifying tests that utilize maximum resources and are critical TAT variation drivers will be key to our recommendation.

The approach being followed to achieve the same is by critically analyzing each step of the Lab Test Process to identify any redundant steps/tasks that could be eliminated. Using concepts of prescriptive analytics like linear programming, forecasting techniques and six sigma, we aim streamline the process and resource utilization across time periods (looking at patterns within a day and across month).

Additionally, an approach to create a model which can forecast the volume of tests to be carried out over a period of time and enable the Hospitals to better optimize their resources, thereby being in a position to increase number of tests being carried out

PROJECT TITLE: Reducing Rate of Clinical Indicators at Global Hospitals

ABSTRACT: Hospital-acquired infections (HAI) is one of the most common complications occurring in a hospital setting. These are acquired infections that are typically not present or might be incubating at the time of admission. The hospitals have established infection tracking and surveillance systems in place, along with robust prevention strategies to reduce the rate of hospital-acquired infections. Identifying patients with risk factors for hospital-acquired infections is very important in the prevention and minimization of these infections. The risk for hospital-acquired infections is dependent on the infection control practices at the facility, the patient's immune status, and the prevalence of the various pathogens within the community. The risk factors for HAI include older age, length of stay in the hospital, frequent visits to healthcare facilities, mechanical ventilatory support, recent invasive procedures, indwelling devices, and stay in an intensive care unit (ICU). For reference, Central line-associated bloodstream infections (CLABSI) is one of the serious infections that can result in longer hospital stays, increased costs, and increased risk of death. These infections are among the most-deadly types of healthcare-associated infections with a mortality rate of 12% to 25%.

For the last calendar year at global hospitals, infections rate for MRSA, CLABSI, PVAP, CAUTI, SSI indices were measured on regular basis, and it was observed 150 infections against 125,000 opportunities which is around 1,200 defects per million opportunities were reported i.e. between 3 & 4 sigma levels.

Our analytical approach for this project will be to deep dive into common reasons that causes HAI and identify discrepancy in the current quality processes (SOPs). Examine the current RCA details shared by the hospital for the infected cases and draw association with the factors. If possible, observational study of live cases will be done to review the methods and identify any gaps using 5 Why and following DMAIC approach for monitoring and reducing the infection rate. By using the six sigma approach this project intends to add value by recommending improvements in the ongoing processes by improving control on faulty processes that puts patients in danger

PROJECT TITLE: Ticket Sales Forecasting for Private Tour Operator

ABSTRACT: Boat Tour operations usually consists of fleets of cruise boats that provide sightseeing tours as well as a mode of transportation. These tours are offered for covering some major attractions of the city. Operators can build the customer base for each tour through selling ticket either for individuals (or couple/ with infants), or for group activities. Mostly, operators run open tours where customers are sold ticket either through online or at the Kiosks at the location. Multiple factors can impact the boat tour operating business sales and key ones include extreme weather, events in the area, public holidays etc. It is also expected that the Events such as conferences, in the nearby convention centres, School Breaks, and Marketing Offers also impact attendance/ticket sales

The project aims to develop a data-driven forecasting model to enable the following:

- To predict ticket sales on a fortnightly period basis
- Forecast ticket sales by ticket type and location
- Help in inventory management and employ human resources such as boat captains and others, in an optimized way

Accurate Forecasting of attendance is very useful for planning resources (Human and boats, equipment, Food and Beverages).

Through the EDA, three key events have been identified for model building which incorporate identifying impact of conventions and weather. Individual models to be created for each event. Furthermore, to understand the impact of the other variables, secondary data have been sourced for multiple parameters of weather such as rain, fog, snow, temp, precipitation etc., Conventions and Public Holidays.

Approach for building the model includes simple univariate as well as complex multivariate forecasting models that incorporate external data i.e., sales data to identify seasonality, and see correlation impacting the sales with the external parameters.