

## PROJECT ABSTRACTS

### 1) Project Title: Optimizing Digital Marketing Strategy for a Leading Private Sector Bank

**Abstract:** Pricing project requires us to develop a model which will allow the retail products to be dynamically priced based on customer's previous purchases. Model is created for a hypermarket which needs customers to have membership for purchase and that helps us do this pricing at customer-basket level. So, every product in a basket is looked at and sent through the model to check if the price of a product can be discounted and to what extent. This is done in order to provide better savings to customers in turn increasing their allegiance with the store.

### 2) Project Title: Study & assessment of probable impact of catastrophic flood events for managing flood risk:

**Abstract:** One of the most common natural disasters in India is flood. Flood knows no geopolitical boundaries and possess serious threats to human life. An increase in extreme flood events in recent years, has caused massive property damages, economic losses and casualty. This has made managing flood risk one of the toughest challenges, confronting risk & asset managers today. Prolonged heavy rainfall is perceived to be the primary reason for most of the flood events. However, there are many other human introduced factors like deforestation, rapid urbanization, poor city planning that contributes to the escalating flood-risk.

Flood catastrophe models or flood risk assessment methodologies available off the shelves often involve complex Geospatial techniques involving overhead of large, multipurpose geospatial tools and frameworks with varied accuracy. Through this project, we aim to study & assess the flood risk exposure in a city in-case of a catastrophic flood event, using analytical techniques. Our goal is to devise an easy to follow methodology using algorithms developed with open-source platforms & freely available data. We will be analysing Digital Elevation Models (DEM) for understanding the terrain and will include other relevant datasets like weather data, as needed for this study. Focus of our analysis will be centred around Bengaluru but we aim to develop the methodology in a way that can be tuned for other cities or regions easily

### 3) Project Title: Prediction of Customer Churn - An innovative approach using Logistic regression and Markov Model

**Abstract:** As the retail market space is expanding at a rapid pace, attracting customers with better prospect and more options to shop, leading to an extreme competitive market. This leads to a requirement for a retailer to not only attract new customers but also to retain their existing customers to maintain their market share. Customer Churn becomes become major concern as it has direct impact on Customer lifetime value. Churn is an important measure to differentiate customers, which can further help the organization to manage them effectively by strategizing them with better offers for retention.

The objectives of this project are:

1. Building a model to pre-empt customer churn in membership of Sam's club using Markov chain.
2. Construction of states of Markov chain model using recency, frequency and monetary value.
3. Calculate the probability of customer movement from one state to other and condition them on individual characteristics like age, gender, income, education level, etc. using logit model and machine learning techniques.
4. Calculate time taken to reach the churn state.
5. Calculate customer lifetime value using Markov chain.

The classification model would provide us with the probability of customer likely to churn.

Again, for a retail giant like Walmart (SAM'S club), apart from the above classification, a timeline representing customer movement to Churn state serves the greater purpose. The retailer can implement promotional strategies and retain the right set of customers using the above modelling techniques.

#### **4) Project Title: Every Drop Counts: Unleashing the prospective locations for Water Harvesting**

**Abstract:** Water is at the heart of 'Sustainable Development Goals (SDGs)' set by United Nations – with an objective to balance the three dimensions of sustainable development: Environment, Social & Economic -- & is indirectly associated with the success of all the other Goals.

But, with changing climatic patterns, untimely rains, prolonged dry spells, depleting ground water & drought making every drop of water extremely precious, the need of the hour is to gauge & work towards the major aspects of water harvesting -- 'Catchment'. Water Harvesting has to be a key element of any strategy to bring an end to India's perennial swings between drought & flood & to meet the following SDGs for sustained development.

This study presents a structured & meticulous approach, wielding 'Geospatial Analytics' to identify the prospective locations for Water Harvesting in arid & semi-arid parts of the country for sustainable development.

#### **5) Project Title: SCORe: Supply Chain optimizations in Retail with focus on optimizing cross dock operations**

**Abstract:** Dealing with the largest retailing company in the world, it provides us with a valuable example of how supply chain solutions can optimize business. One of the major components to its success in achieving this is cross-docking. The efficiency, cost-savings and logistical advantages of crossdocking have been worldly recognized and is now being implemented by majority of retail chains. The biggest question that still hovers over these retail chains are which items should be categorized as cross dock and which items should be staple stock. Thus, we have been assigned to solve 3 predominantly raised questions 1. Design a workflow drives channel decisions for items. 2. Do a cost based analysis model for classifying both existing and new items as cross dock or staple stock 3. Evaluate existing classification and propose which should be the ideal channel

#### 6) Project Title: Project CRISIS (For Gold loan data) CRISIS: Credit Rating & Increase Sales through In-depth Segmentation

**Abstract:** Financial institutions have introduced innovative lending against gold services at cheaper costs and provide better customer service. The organized gold loan market has witnessed high growth rates in the past decade. The gold loan market has emerged as one of the most reliable and convenient source of credit for low-income individuals and households, compared to other sources of credit available such as loans from Microfinance Institutions, personal loans, loans form self-help groups or community borrowing. Like any other loan, it is imperative for any gold loan lending institute to do a credit rating of the applicants so that defaults are limited. Hence project CRISIS encompasses credit scoring through in-depth segmentation techniques. More so another aspect of this project is to increase revenue by cross selling other loan products to top rated customers.

#### 7) Project Title: Retail Analytics – Price Optimization to increase sales volume

**Abstract:** *“Pricing has long been—and will continue to be—a core capability for retailers”- McKinsey*  
To derive a pricing model for a global retail giant that appreciates the high volume buyers and increase the sales volume of the store. Optimized prices are chosen for retail products adhering to two key principles: EDLP (Every Day Low Pricing) and Minimum profit margin.

Major analytical techniques to be employed will include Customer& Product Segmentation, Demand estimation/Ordinal Regression methods and much more.

Customer & Product segmentation involves selecting the right products and right members for whom the optimized price is applicable.

Pricing Model: Goal of this entire project is to give insights into product demand for different price ranges. This allows the decision makers to price the products in accordance with the demand expectations, which in turn drives the revenue.

#### 8) Project Title: Prediction Algorithm for Machine Failures

**Abstract:** With the growth of sensor technology and telematics, there has been significant use of this technology in both industrial and consumer goods. Industrial machines are far ahead in the curve resulting in huge volume of machine data captured by different industries. This data is typically used to predict the likelihood of machine failure ahead of the actual occurrence in order to provide service support in a timely and pre-emptive manner such that revenue loss and unplanned downtime can be reduced.

This project is primarily focused towards detecting failures in industrial printers. There are two stages of predicting machine failure: First, the identification of what constitutes a failure – partial or complete failure, and, second, the prediction of the actual failure event. The solution would use an anomaly or outlier detection technique to capture the defined failure event. In order to actually predict the occurrence of the failure event, classification techniques are first used – providing the probability of failure. Further study using Survival analysis enables failure prediction at various points in time – that is to say, time left until the occurrence of the failure event.

The project aims to build an algorithm that will enable the identification and prediction of failures for industrial printers with sufficient lead time to mobilize service support.

**9) Project Title: Deploying Machine learning techniques for Feature extraction, Classification and calculating area under cultivation using satellite imagery**

**Abstract:** Agriculture is one of the biggest contributors to the GDP of the nation. It is important that the industry is taken forward along with the technology revolution. For agriculture is crucial to the nation not only in terms of economics but also, to help us gain self-sufficiency in sustaining of our population. Besides, India has been traditionally an agricultural economy and still, much of its population depends on it for their livelihood. With the increase in availability of remote sensing data and Machine Learning capabilities, there is an opportunity to identify, monitor and make intervention at a macro level at much less time horizons and cheaper costs. As part of this, automated pipeline will be created to map farmlands with satellite imagery (Geospatial pre-processing). As other crops can benefit from a similar approach. Following which ML algorithms commonly used like support vector machines, decision trees, ensemble methods, random forest, neural networks, clustering, etc. will be tested and deployed for feature extraction and classification of the crops. To start with wheat and tomato farmlands have been included. Tomato farmlands in particular will be valuable. Based on the results on ML, then the area under cultivation will be calculated.

**10) Project Title: Optimization of Nurse Scheduling at Apollo Hospitals using Prescriptive Analytics**

**Abstract:** Nurse scheduling is an activity of allocating nurses to conduct a set of tasks in a given ward at a hospital for a given period. Some of the challenges in nurse scheduling are the lack of required number of nurses in order to fulfil the needs of the hospital. Currently nurse scheduling is undertaken manually and is time consuming and not data driven, thereby resulting in possible suboptimal usage of the available nurses and/or overburdening the existing staff with irregular schedules. Therefore, this project aims to develop a data driven scheduling model that satisfies all the specific rules set by the management of Apollo Hospital. The developed model considers the recommended policies and hospital objectives by Apollo Hospitals. Hospital objectives include ensuring a continuous service with appropriate nursing skills and staffing size, while avoiding additional costs for overtime. The main approach we will be considering in scheduling the nurse personnel will be optimal schedules of nurses obtained by binary integer programming models for assignment of shifts. Another scheduling model will also be developed to reduce the overtime put in by the nurses due to the shortage of nursing staff. A 2-week planning horizon will be considered for the both the scheduling models.

By automating the scheduling process, the aim is to optimize and increase the efficiency of nurse scheduling while meeting the hospital requirements and reducing the fatigue the nurses are facing due to the irregular schedule. This can help in reducing the excess workload on nurses and eventually help in reducing the nurse attrition rates

**11) Project Title: Life Time Value (LTV) of a New Patient for a Hospital**

**Abstract:** Medical practice is after all a business as well and, they also want to earn profit from their services. This is where the concept of “Life Time Value of a patient” becomes a very critical measure.

Life Time Value of a patient is the amount of revenue he brings to the practice, cumulatively overtime, during his tenure with the practice.  $LTV = Treatment\ Cost * Frequency\ of\ Visits * Referrals * Margins$

The same metric is very important for hospitals as well. However, the derivation becomes very complicated in the case of hospitals because of multiple practices, Inter-departmental referrals, multiple starting points, etc. In our thesis project title “Life Time Value (LTV) of a Patient for a Hospital”, we’ll try to develop an analytical method to derive this measure and help hospitals run a profitable business.

### **12) Project Title: Indian Automotive Industry Sales Forecast: Time Series Analysis and Impact of Macroeconomic Indicators**

**Abstract:** The success of any corporate depends on effective strategies and operation planning. Inaccuracies in planning leads to huge cost burden and loss of opportunities, this is especially true for automobile makers and its suppliers, where the capital investment expenses are large and time taken to develop new product is 2 to 3 years. Bosch is one of the major such suppliers of Auto components in India with a relatively high market share. The service-level-agreements between Bosch and the automobile companies allows the automobile companies to change their purchase orders for components anytime during the year depending on their own demand forecast. Bosch deals with total of 39 such vehicle segments.

The objective of this project is to study the influence of exogenous parameters like GDP, IIP, Fuel Rate, GFD, Forex, CPI and many more on the sales of automotive vehicles in India and build a highly accurate sales forecast model for Bosch to help it in its production planning of components for all such segments.

### **13) Project Title: Persistency Analytics – Developing a Model for Predicting Persistency of Policies**

**Abstract:** For a life insurance company, it is very important to predict policy persistency for the following reasons:

- 1) Retention: It helps the company to focus on risky policy holders that can be prevented from defaulting/ surrendering on a policy.
- 2) Reserve management (regulatory capital requirements): Predicting premium persistency is vital as premium collection has direct impact on reserve accounting and in turn affects the profitability and cash flow of an insurance product.
- 3) Cross-selling: It helps identify potential opportunity to sell other insurance products in future to a paying policy holder.

Overall, predicting premium persistency will reduce the cost of insurance and in turn increase access to insurance. The scope of the project includes:

- 1) Building a predictive model that provides a persistency score for each policy due for payment
- 2) Provide insights into improving persistency
- 3) Provide profiles of policy-holders – persistent and non-persistent

#### 14) Project Title: Text and Sentiment Analysis of BREXIT & FREXIT

**Abstract:** Capture opinions expressed by the voters in social media platforms pre & post BREXIT and categorize them as positive, negative or neutral. Our plan is to leverage social media platforms such as Google trends, Twitter, Facebook and YouTube, to analyze the sentiments. We also plan to explore the geographical and demo-graphical influences of the final decision of Stay Vs Leave and analyze the occurrences of key events/influences that would have helped us predict the final decision earlier. We also aim to explore multiple text mining techniques to understand underlying sentiments and trends in data. We will also attempt to build a predictive model that would predict potential exit of member nations from EU in near future.

#### 15) Project Title: Deep Lens for Air Pollution: Estimation and forecasting of atmospheric pollution from satellite imagery

**Abstract:** In May, World Health Organisation declared New Delhi as the most polluted city in the world! According to India's National Health Profile 2015, there were almost 3.5 million reported cases of acute respiratory infection (ARI) in 2015, a 140,000 increase from the previous year and a 30% increase since 2010! Bangalore has around 26 lakh children and more than half of them suffer from asthma! Primary reason: The rate at which urban air pollution has grown across India is alarming, much higher than that of water pollution and pollution caused by improper disposal of garbage.

Economic Times reports that "India uses machines called high-volume samplers to measure PM 2.5, tiny and dangerous airborne particles that are less than 2.5 microns in size and are fine enough to enter deep into the lungs and the bloodstream. However, most states choose the cheapest manufacturers who make machines using poor quality material and their readings cannot be trusted. The Central Pollution Control Board (CPCB) compared international samplers with Indian ones and found massive inconsistencies in the data." Also, ground sensors are very expensive and hard to maintain. Our goal is to predict the pollution in different parts of Bangalore through satellite images instead of deploying expensive sensors. Idea is to establish correlation between ground sensor and satellite imagery data for the location where we have both the data available and then scale it to other areas where ground sensor doesn't exist.

#### 16) Project Title: Dynamics of Tea Production: Forecasting the Price and Quantity

**Abstract:** Tea industry in India has experienced both increasing and declining trends. India is one of major producer as well as consumer of tea and tea as an industry plays an important role for capacity to earn foreign exchange and provides massive employment opportunities. There are a number of factors including the quality, yield rate of Indian tea gardens, total area under production and other social and macro-economic factors that impact the quantity and price of the tea in market and warrants a good amount of research. Competitor force also plays an important role in the same as the supply market is not very wide spread. Though there have been some studies in terms of general industry analysis from the perspective of a specific company who deals in the business deserves attention. This business case is on one of the India's largest packaged tea brand. Tea costs around 60% of the total cost of the company for the packaged product. So a proper forecasting of price and quantity of tea to better plan procurement through auction or open market has a direct impact in terms of profitability. There are limited time periods for purchase and supply and demand is season specific. Our attempt is to build a forecasting model for procurement which adds a lot of value to the business.

### 17) Project Title: Prediction of change in the phases of Indian equity market

**Abstract:** Equity markets undergo changes in phases. These changes are subtle and the latent forces shaping these changes are complex in nature given the integration of financial markets across geographies and asset classes. These phases can last from a few months to a few years, are characterized by different risk-return profiles and therefore require adoption of different strategies. A more accurate prediction of occurrence of equity market phase changes therefore, is a subject of interest to equity traders, investors and business owners planning to tap the equity market. This study aims to devise a prediction/forecasting model/tool which would predict/ forecast phase changes in the Indian equity market index – NIFTY 50. Statistical and machine learning based techniques would be used to analyse historical NIFTY 50 index data.

### 18) Project Title: Recovery channel and Returns revenue optimization for a U.S retailer

**Abstract:** According to the US National Retail Federation, Americans returned \$260 billion in merchandise to retailers in 2015 which amounts to 8 percent of all purchases. That swells to 10 percent around the holiday season. Less than half of returned goods are re-sold at full price, retailers may end up forfeiting 10 percent of their sales

This project helps in developing a model that gives the optimal mix of recovery channels at a subcategory level to maximize the revenue from returns. The model factors different types of returns like Consumer remorse, product recalls, inventory overstock, and product defects.

### 19) Dynamic airport check-in counter allocation using Linear Programming

**Abstract:** As per IATA1 (International Air Transport Association) the demand of passengers for air travel would double the estimate of 2016 (3.8 billion) by 2035 (7.2 billion). This has fuelled interest in various research communities to look for an efficient way to handle the ever-increasing demand with limited capacity today. Research efforts towards managing the check-in at airport for a given airline has been pursued actively. This involves modelling the passenger arrival process and estimating the optimal counters needed for smooth passenger's check-in thereby reducing congestion near check-in kiosk along with reduced chances of flight delays.

This project work derives a methodology to determine the optimal counter needed to serve multi-fold purposes of reducing the passenger wait time, reduced counter operational cost and lesser congestion. As an outcome, the study would deliver an insight on the required counter operational minutes for a known passenger arrival process, and the expectation on the passenger waiting time prior to serve. The model is developed using linear programming for counter optimization followed with simulation to determine the expected passenger wait time in queue. Unlike, existing work the model outputs dynamic counter allocation in each 30minute time interval (considering counters to remain open for 3-hour duration prior to departure of flight).