



11th Symposium on
Business Analytics and Intelligence 2021

BOOK OF ABSTRACTS

SI.NO	PROJECT TITLE
1	A Quick and Accurate Fall Detection In a IOT Enabled Emergency Care Device For Elders
2	Application of Deep Learning for Information Extraction of Digital Resume
3	Customer Sentiment Analytics for Passenger Car Segment
4	Electricity Load and Price Forecasting for Short Term Period
5	Engineering, Procurement, and Construction (EPC) Project Optimization using Evolutionary Algorithms
6	Identification of Root Cause of Higher Customer Backlogs
7	Combining MOOCs Clickstream Data with Analytical Thinking to Optimize Learning and Benchmark the Learning Environment
8	Improve Evaporator Performance Applying AI/ML to Define and Forecast Key Operating Metrics.
9	Logistic Planning Last Mile Performance Management
10	NPS Driver Analysis Using Effect-Based Kano Model
11	Predicting Medication Errors at Multi-Specialty Hospitals
12	Prediction Of M&A Activities Post Covid-19, Across Industries in the USA
13	Profiling the People Persona and Developing Intelligence on People Behavior Using App-graph Data
14	Smart Meter Disaggregation
15	To Predict Likelihood of Prospective SPONSORs to Enable SPS for a Multinational Financial Service Company (MNFSC).
16	To Reduce Overall Cost by Improving Seat Utilization and Vehicle Route Optimization

A Quick and Accurate Fall Detection in a IOT Enabled Emergency Care Device for Elders

Abstract: Fall and associated injuries with fall is a major health hazard for the elderly people. In such cases immediate medical assistance becomes very crucial. If elders are not monitored, then detection of fall and alerting of medical assistance may not happen in time causing serious risk for the life and well-being of the person. To overcome the problem of alerting medical facilities about the fall, an IOT enabled device is designed. The elderly person should wear the device on his wrist and the device is paired with a special application of the smartphone. The device is fitted with Accelerometer (ADXL345 and MMA8451Q) and Gyroscope (ITG3200) which measure the acceleration and angular velocity every 1/5th of a second. The readings from the device are fed into the smartphone application which monitors the data and generates an alarm when it detects a fall in the readings. The alarm will be received by the medical assistance facilities and prompt care will be provided to the elderly person.

Different models for fall detection like RNN, LSTM will be evaluated. The goal is to select model that must be fast and that must rely on minimum computing resources available on the smartphone. The model must achieve high sensitivity and high specificity. False positive has a high-cost impact and True negative has serious impact on life.

Application of Deep Learning for Information Extraction of Digital Resumes

Abstract: Recruitment process is one of the most critical activities performed in any organization. The process begins with understanding the Job requirement, followed by screening candidates through "Resumes" posted on Job portals, In-house Company portal or Referrals and finally

calling the right candidates to the interview process. Resume Screening is not only tedious but time-consuming; As each candidate tries to differentiate from the rest, resumes have become more unstructured and unique not to mention digitized, making it difficult to read all the resumes manually. Usually resulting in errors such as filtering out a right candidate or selecting a candidate that doesn't match the requirement. Thus, there is a need for automation to eliminate human errors. Automatic resume screening and data extraction is still an on-going process. Applicant Tracking System (ATS) is one of the popular tools in recent times for recruitment and talent acquisition process or HR department. Most ATS works through keyword matching and questionnaire pattern. However, the major drawback is its false positive or false negatives while selecting the resume for next level based on the information extracted. Our work here addresses this issue and takes a step forward as to rank the candidates matching to the job requirement.

We propose two approaches here. Initially, a base model using simple pattern matching and rule based. The second using applications of Deep learning techniques such as NLP. Finally comparing the accuracy. Our main objective is to provide a model that is capable to extract relevant information from any resume formats and provide the screener with the insights to select the right candidate.

Customer Sentiment Analysis for Passenger Car Segment

Abstract: Automobile industry in India was highly protected in favour of domestic car manufacturers till late 1980s. World leaders in passenger cars such as Toyota, Honda, General Motors, Ford, and Hyundai set up manufacturing hubs in India, cashing on the liberalized Foreign Direct Investment Policy of the Government of India. They captured the hearts and minds of customers, with their choicest of car models with high technological and innovative product offerings. This transformed the automobile scene from a seller's market to buyer's market. Car customers had started developing their own personal preferences and purchasing patterns, which were hitherto unknown in the Indian automobile segment. In this context, it becomes important not just to know how the overall economy is driving the PC segment, but also to specifically capture and analyze the other factors like geographical, demographics, technology, social, regulations and psychographic factors which may influence the customer sentiment in enabling growth of this segment. Use of data analytical techniques such as text scraping, text mining, NLP, classification and forecasting techniques aids us in deriving a Customer sentiment index for the a) Segment level, b) Sub segment level, c) OEM level, d) Model level end user preferences which would then be translated into studying the drivers for the preferred models. A recommendation matrix would then help in archotyping the product portfolio and augment the customer strategies to penetrate further into new business opportunities. The results of the research would help in capturing customer behavior influencing buying patterns in passenger car segment in automobile industry.

Electricity Load and Price Forecasting for Short Term Period

Abstract: The energy trading industry has developed and made progress with uttermost rapidity in the last few years. Electricity is now traded under market rules using spot and derivative contracts in many countries across the world. Alberta currently operates a wholesale power market that sets a price for electricity every hour for the whole year that determines the revenue for generators as well as the cost for consumers.

However, electricity is a very special commodity. It is economically non-storable, and power system stability requires a constant balance between production and supply. The main factors of the energy industry are demand & supply which depends on weather parameters and the intensity of business and everyday activities. These features cause some important characteristics of the electricity prices: high volatility, sharp price spikes, mean-reverting process, seasonality in different frequencies etc. Because of all these idiosyncratic features, forecasting the electricity prices accurately becomes a very challenging task.

This project aims to make high-frequency short-range price forecast to take electricity trading decisions at asset as well as macro-level for the deregulated Alberta Canada market. The model will be able to help the users to make immediate operational and financial decisions based on deep learning models. This will enable electronic trading decisions.

Forecasting energy prices is a function of load prediction. We will be considering historical load, climate data, and calendar information for forecasting load at an hourly interval of time and in succession use load and other variables to forecast price for the day-ahead market in Alberta Canada.

Using an accurate price forecast, energy supply companies would be able to optimize their procurement strategies. Companies can also schedule their operations according to the low-price zones. In a nutshell, it will enable users to make informed energy trading decisions.

Engineering, Procurement, And Construction (EPC) Project Optimization using EvolutionaryAlgorithm

Abstract: Planning and managing an EPC project requires consideration of a lot of factors (time, budget, quality, the occurrence of unfavorable events, etc.) that can impact the project. The project scheduling within a defined budget and time with optimal resource utilization is very crucial. Currently, most of the managers do it manually which is time-consuming and is usually opinion-based resulting in a suboptimal plan. Others use Project Evaluation and Review Techniques (PERT) and Critical Path Method (CPM) which solve time scheduling but fail to optimize budget, quality, and resource level utilization.

The core objective of the proposed model is to support data-driven decision making for horizontal and vertical project planners by optimizing resource utilization to minimize project duration and its cost while maximizing quality. The framework uses a combination of the Critical Path Method (CPM) and multi-objective Genetic Algorithm (GA) to generate optimal solutions. It also predicts the possibility of occurrence of risks based on historical data and its impact on the project. The model will also assess real-time data and propose the next best feasible solution that can be viewed in an interactive Gantt chart.

The model enables the project managers to focus on planning and scheduling activities in such a way that (1) Increases the resource efficiency; (2) Minimizes construction total time and cost (3) Measures and improves construction quality; (4) Provides the ability to actively monitor risks and act on the proposed solution.

Identification of Root Cause of Higher Customer Backlogs

Abstract: Customer Service Levels (CSL) is an important KPI that provides the firm an insight in understanding overall efficiency in customer engagement. The CSL is measured at 2 levels, CSL1 is for matching an order line to a physical part & issuing a part post billing /invoice, CSL2 is on delivering the part to the customer. The project's focus area is CSL1. The firm would like to achieve higher fulfilment of CSL1. The scope of analysis spans across all major distribution centers which serve customers.

The project aims to develop categories of failures using the data driven root cause model considering the thresholds on the service levels requirements set by the firm's logistics team. In addition to considering some recommended factors for analysis by the firm, the developed model will try and identify the dominant factor in failure chain. Logistics team objectives also include optimally enhancing the current service levels.

By automating the analysis process, the aim is to identify, prioritize and address any trends that induce the increase in customer backlogs and thus impacting the customer service levels while meeting the firm's requirements to increase operating efficiency.

Combining MOOCs Clickstream Data with Analytical Thinking to Optimize Learning and Benchmark the Learning Environment

Abstract: MOOCs offered by IIMB covers a wide range of subjects in core management that includes Marketing, Strategy, Accounting, Information Systems, Human Relations, Finance, Operation Management and Statistics. It also includes contemporary management topics such as Analytics, Banking and Financial Markets, Information Technology Management, International Business Management and Sustainability. Pressing demands to optimize the learning, benchmark the learning environments, adjust the quality of the hosted courses, as well as controlling the high dropout ratio and the lack of interaction are the motivation for MOOC data analysis. Given this context, this project aims at analyzing MOOCs data to provide good experience to learners and interesting insights to instructors, pedagogical researchers, and management to improve the content, design, and delivery of the courses. Thereby intend to equip stakeholders to achieve high level of learner engagement, interactive experience, and completion rates.

Data Source: 1. A dump of clickstream data from student browser and platform tracking logs. 2. Course calendar details: Start date, End date, Release date, Assignment deadlines. 3. Course communication details: Emails, marketing, updates and announcements. 4. A dump of video stream data: Start, Pause, Speed Change, Stalled etc. 5. Assessment Data: Number of attempts, Submission records, First attempt score, last attempt score, In-video quizzes etc. 6. All discussion forum posts, edits, comments, and replies data related to the courses.

Analytics in Action: 1. Evaluate the reported data against hypotheses - to improve the quality of hosted courses. 2. Understand the context of the course run: the environmental factors and choices that make each run unique - to drive learner engagement and enhance learning experience 3. Decide whether action is called for - to Predict user-dropout and drive higher completion rate 4. Which action to take and when? - to Optimize learning and benchmark the learning environment. 5. Develop MOOC-Bots and Integrate into e-learning platforms to moderate discussion forum.

Improve Evaporator Performance Applying AI/ML to Define and Forecast Key Operating Metrics

Abstract: The oil and gas sector currently face challenges due to depressed oil prices. The sector needs AI based intervention for improving efficiency, reducing cost and emissions there by making it greener.

This project offers data that has been collected past few years on an entire facility, including multiple well pad and central processing facility which includes thousands of sensor data. The main objective is to optimize boiler and evaporation process. The optimization will be done for both evaporator performance as well as capacity.

Evaporator Operation Performance is Optimized by minimizing the energy consumption. Evaporator Capacity is taken care by optimizing the performance of system to maximise production of distillate, by minimizing the evaporator blowdown rate and by Optimizing the washcycle frequency to minimize energy and chemical cost while maintaining target distillate rate.

Therefore, this project aims to differentiate various evaporator assets and their behaviour, identify optimal set of inputs to provide the best \$/m³ of distillate and suggest the best distillate. Finally provide an implementable model with increased accuracy along with insightful recommendations.

Data visualization is done to understand the evaporator behaviour and to identify features for modelling. Predictive Models such as Regression and Time series forecasting have been explored to identify optimal set of inputs and Non-linear optimization for model development.

Logistic Planning Last Mile Performance Management

Abstract: As manufacturing organizations expand their customer base and geographic presence year-after-year, the distribution network grows in size and complexity. Administering On-Time Delivery (OTD) of shipments, a Key Performance Indicators (KPIs) for customer satisfaction, demands organizations to adopt a strong and efficient transportation management framework. This framework must be able to detect and eradicate any inefficiencies that are identified within the network.

This project, pertaining to a leading Industrial Technology supplier, aims to solve similar operational challenges by diagnosing anomalies within their current last mile transportation operations and providing meaningful business rules to ensure optimal OTD. The proposed solution leverages internal (i.e., transportation management system data) and external data sets to identify non-performing OD pairs (i.e., Origination-Destination) by unearthing hidden patterns of inefficiencies. The main approaches we will be considering to maximize OTD levels will be tree-based models as well as statistical models.

NPS Driver Analysis using effect-based Kano Model

Abstract: The advent of e-commerce has resulted in a smaller, but more competitive battleground for retailers. Therefore, determining how effective your business is at making customers happy is essential. This is why, the Net Promoter Score is so valuable. Net Promoter Score (NPS) is used to measure customer loyalty and how likely they are to refer your products and services to others. NPS helps identify who among your customers are: promoters, passives, and detractors. An NPS customer feedback survey aims to answer two questions:

- i. How likely are your customers to recommend your product or service?
- ii. What are the key factors influencing your customers' likelihood to recommend your product or service?

Computing the Net Promoter Score indicates how likely your customers are to recommend your product or service but doesn't explain the 'why' behind their response. The underlying goal of driver analysis is to determine the key attributes of your product or service that determine your Net Promoter Score. These attributes are referred to as 'drivers'.

This information can further be used to influence how to tailor products and where to focus your efforts. To identify key drivers to a differentiated product or service, it is imperative to extract 'meaningful information' from the factors influencing customers' likelihood to recommend. Suggested approach to accomplish this objective is to categorize every comment into a sentiment brackets, followed by using topic modelling techniques to extract features/topics that influence NPS. This information can be further used to identify a key driver/attribute across geographies and product categories using KANO model and Factor analysis.

Predicting Medication Errors at Multi-Specialty Hospitals

Abstract: Hospitals are an integral part of modern-day society, and healthcare analytics is a burgeoning field of interest, to ensure adherence to quality. One process immensely vital to quality is ensuring that the patients get the right medications. Medication errors can happen at the hands of the doctors, during creation or updation of prescriptions; by nurses / typists whilst transcribing these prescriptions; or by pharmacists during dispensing. Any of these could lead to a wrong dosage and mode of delivery, an inappropriate timing or even an entirely incorrect drug being administered to the patient. This could have adverse effects, and at times could even prove fatal.

This project aims to effect performance improvement by developing an analytics model that effectively predicts medication errors, driven by the data from India's leading multi-specialty hospital chain. Even a moderately sized Tier-II city's hospital issues nearly 1.5 million drugs per year to its patients. The scope of this project is to identify key predictors that cause and influence errors, so as to reduce, if not completely avoid their future occurrence. Both statistical and machine learning techniques are employed to achieve this, making the model serve as a decision support system. Along with such a system, the project also suggests continual process improvement methodologies like lean six sigma to ensure that the model can be effectively utilized post deployment.

Prediction Of M&A Activities Post Covid-19, Across Industries in the USA

Abstract: Due to COVID-19 pandemic, the world is going through a transformation across different industries and making a long-lasting impact. Studies done during the subprime mortgage crisis, especially from late 2007 to early 2009, suggest that situations of this nature often trigger consolidation within sectors that bear the brunt. Almost 17% of public companies don't survive a global downturn. The organizations that made significant mergers and acquisitions aligning to their long-term strategy outperformed those that did not. In many markets globally, it is expected to see multiple M&A activities over the next couple of years. The forward-thinking leaders need to rebalance their risk, liquidity and strategy while assessing significant opportunities for resilience and growth coming out of this downturn.

Looking deeper into M&A history, the testing times like recession and pandemic/epidemic situations leave a trail on how the M&A would play out. So, in this situation of global crisis, forming a view on who will drive consolidation, who will get consolidated, and how one should respond is critical to making better decisions.

With this background, we look forward to predicting possible M&A scenarios across industries in the USA.

This project analyses the M&A trends during the recession, pandemic/epidemic situations across different industries using data available from several M&A and financial databases specific to US market like Bloomberg professional services, Factiva, Refinitiv - SDC Mergers & Acquisition, Frost& Sullivan, Venture Intelligence, World bank, etc.

Based on the past data, we intend to build a model that continuously learn and forecasts future mergers and acquisitions that are likely to happen by categorizing companies into different buckets based on various parameters around financial health, market conditions, etc. The project aims to achieve this with the support of a custom-developed scraping tool that can help the model predict possible M&A opportunities for companies across industries in the USA.

Profiling the people persona and developing the intelligence on people behavior using the App-Graph data

Abstract: Marketing and ad-tech companies leverage the sources of intelligence on people and places for enrichment and marketing, in a privacy-led environment. Due to an increase in privacy restrictions, many digital marketing companies are noticing a reduction in the user profiles to create the target audience. It is important to generate accurate user profiles for the business. The recent privacy laws do not allow the companies to create the profiles deterministically which is leading to develop the probabilistic models by using the plethora of machine learning techniques.

This project conducts a study to profile a user using the information about the apps s/he uses. It helps in deriving further insights, to create targeted audience, device, location and highly detailed campaigns.

Today, a person can be profiled using the information about the apps s/he uses. An app graph gives the details of the different apps installed in smartphones. Obtaining the app graph information will enrich the user profile and thus help in deriving further insights.

This project aims to develop a robust prototype of a machine learning model to create the user profiles which can be implemented by Near - the sponsoring company. This approach can identify high value profiles and understand people's patterns and preferences in using the functions of the mobile app. The approach allows developers to gain insights into customer behavior and function usage preferences by analyzing big data.

Further, this study could be combined with the unstructured (ex: social media, visitation data etc.) data to understand customer behavior. Such an analysis could be used to shape product improvement and customization decision-making. Future studies could also examine how to use time-series data to track individual customers and analyze their behavior.

Smart Meter Disaggregation

Abstract: Smart Meter enables us to collect the electricity consumption readings at granular level of data for a household. This project aims to break down total energy consumption at household level into the contributions of single electrical appliances.

The breakdown of the energy consumption has multi fold usage, First, informing the household's occupants of how much energy each appliance consumes, helps consumers to take steps towards reducing their energy consumption. Second, personalized feedback on appliance specific advice helps empower customer take financial decision on replacing low-efficient appliances into high-efficient ones and reducing energy waste or maintain the existing ones. Third, disaggregation of energy when combined with appliance usage time can recommend, deferring appliance usage to a time of day when electricity is cheaper.

The objective of this project is to develop a deployable model to monitor non-intrusive load monitoring system for smart homes with the goal of identification, prescribe and optimize the overall energy consumption and provide a real time feedback to the customer. The aim is also to evaluate the feasibility of the model on multiple electric devices.

To Predict likelihood of prospective SPONSORs to enable SPS for a Multinational FinancialService Company (MNFSC)

Abstract: Stock Plan Services is a special service by which clients have the opportunity to participate in their respective company stock plan. SPS is a plan that caters to corporate employees to manage their stocks and allows employers to customize & manage their stock offerings. The SPONSOR can be a Fortune 500 company or a start-up and there are various parameters that determine whether a Sponsor would buy the MNFSC's SPS. They are financial performance of the SPONSOR such as Market Cap, Revenue, Total Assets, EPS; Macro-economic factors such as GDP, Inflation rate, unemployment rate, taxation; Sponsor's performance as Number of employees, Age of the company etc. and finally the MNFSC's performance with respect to SPS.

The MNFSC has an existing model for this problem for Fortune 500 companies but not for Start-up. For Fortune 500 companies, the parameters they have chosen are very different from the parameters or secondary data we have selected.

Thus, the project aims at the following: Based on secondary data, find features that are common across SPONSORS that would offer SPS- Stock Plan Services to its employees via MNFSC using classification techniques such as Logistic regression/Random Forest/XGBOOST and ultimately provide propensity of a prospective client to enable Stock Plan Services.

The scope of this project is that the MNFSC will have a list of new Fortune 500 SPONOSOR's and a new perspective of Start-up SPONSOR's whom they can target in the future.

To Reduce Overall Cost by Improving Seat Utilization and Vehicle Route Optimization

Abstract: Vehicle Routing Problem (VRP) is an important element of many logistic systems which involve routing and scheduling of vehicles from a depot to a set of customers node. This is a hard combinatorial optimization problem with the objective to find an optimal set of routes used by a fleet of vehicles to serve the demands of customers. It is a classical problem which encompasses multiple facets starting from already defined time windows, fleet and driver scheduling, pick-up and delivery in the planning horizon to optimize Route planning, better seat occupancy (existing state 85% to target state of 95%), minimizing single/double occupancy routes, cab allocation basis zones/routes, Reduction in No Show by employees, minimize escort trips (female employee last drop and first pick up). Some of the challenges include siloed operations for different offices and hubs, sub optimal fleet occupancy, no work-around for ad-hoc cancellations and no-shows, similar approach across all times and seasons, and no method of penalizing the users.

The objective is to minimize the overall costs of all routes over the planning horizon. We model the problem as a Linear mixed Goal program and a combination of heuristics and exact methods for solving the model. This project aims to develop a stochastic model that satisfies all the specific rules with further additional constraints which will be accounted as part of the objective function. Solution approaches shall include and not limited only to Data Analysis and data interpretation for noise/ patterns, Regression analysis to understand factors influencing the overall cost/ occupancy, Demand forecasting vehicles for a specific hub and office location basis time, Mathematical goal programming models for assignment of cabs/ hubs and outline factors to optimize No show and escort trips as opportunities for the client. Automation of the scheduling process will help in reducing the cost as well as better employee service with forecasted demand and better routing for the cabs on all routes.