



BOOK OF ABSTRACTS

LIST OF PROJECTS

SI. NO	PROJECT TITLE (A-Z)
1	A Hybrid Prediction Model for No-Shows and Cancellations of Employee Transport System
2	Analyze Help Desk Tickets/Issues for Operation Optimization and Proactive Action
3	Analyzing Customer Behavior for A Money Transfer Platform Using RFM And Advanced Predictive Models
4	Building A Preventive Detection System for Battery Drain Issues for Bounce
5	FPL Squad Planner for Each Gameweek Using Prescriptive Analytics
6	Leveraging AI in a Skilling Ecosystem
7	Making Energy Intelligent: AI-Based Interventions in The Oil and Gas Sector
8	Manufacturing Analytics for Carpet Industry
9	Operational Analytics for Leads Conversion in Real Estate
10	Operational Risk - Dynamic Claim Allocation at Revenue Cycle Management of an US Healthcare.
11	Optimization of Rewards & Recognitions Contests for A Reputed Insurance Company Using Predictive & Prescriptive Analytics.
12	Optimizing the Discharge Process by Tuning the Discharge Process Time for a Multi-Specialty Hospital.
13	Predicting the Risk of Diabetes, Hypertension and Cardiovascular Diseases in Patients Based on Routine Blood Tests.
14	Prediction of Equipment Breakdown Using Analytics
15	Stock Price and Investment Action Prediction Through Sentiment Analysis of Earnings Call

PROJECT TITLE: A Hybrid Prediction Model for No-Shows and Cancellations of Employee Transport System

ABSTRACT: A no-show occurs when a scheduled employee neither keeps nor cancels the appointment. A cancellation happens when individuals cancels the schedule in the app for their scheduled appointments. Such disruptions not only cause inconvenience to management, they also have a significant impact on the revenue, cost and resource utilization for almost all of the transport systems. We develop a hybrid probabilistic model based on multinomial logistic regression and Bayesian inference to predict accurately the probability of no-show and cancellation in real-time. First, a multinomial logistic regression model is built based on the entire population's general social and demographic information to provide initial estimates of no-show and cancellation probabilities. Next, the estimated probabilities from the logistic model are transformed into a bivariate Dirichlet distribution, which is used as the prior distribution of a Bayesian updating mechanism to personalize the initial estimates for each employee based on his/her attendance record.

In addition, to further improve the estimates, prior to applying the Bayesian updating mechanism, each employee schedule is weighted based on its recency, weekday of occurrence. We also discuss the advantages of the proposed hybrid model and describe possible real-world applications.

PROJECT TITLE: Analyse Help Desk Tickets/Issues for Operation Optimization and Aim Proactive Action for a Huge Real Estate Service Firm

ABSTRACT: Facilities service management collects a lot of operational data. Analytics in Service management leverages the vast amount of information (free flow as well as categorized) captured in the incidents fields to provide universal gains on reduction in operational costs and overall contractual benefits.

This project is for a huge real estate services firm. Its services include sales and leasing, property management, project management, and development. Founded in the 18th Century it is second largest firm of its kind and is listed in Fortune 500. We are working with the Facilities management unit of this firm which deals in providing office leasing space and is responsible for contractual facility management activities like building maintenance, housekeeping, safety, security, cleaning, equipment maintenance etc.

This project entails performing analytics on one year of incidents/ ticket's data, with an aim to reduce the number of incidents by focusing on high volume, high impact tickets. After enriching the given data using external variables like weather, text analytics (Named Entity Recognition) would be performed on text fields to ensure right categorization and routing of tickets. This would be followed by creating a classification model for more accurate and automatic bucketing into proactively solvable and other incident categories. In addition, patterns would be explored to reduce ticket reworks. These proactive actions would allow reduction of incidents and more proactive actions which should reduce the operational costs and also allow the firm to ultimately pen contracts with cost effective SLAs.

PROJECT TITLE: Analyzing Customer Behavior for A Money Transfer Platform Using RFM And Advanced Predictive Models

ABSTRACT: A South Africa based International remittance service provider aspires to better understand their customer base by analysing the untapped trove of historical customer transactions and use the insights to design their marketing strategies.

This project's aim is to analyse all dimensions of their customers and create a model –

1. to predict customer behavior, customer churn (non-returning customers) and
2. targeted campaign designs to reduce predicted churn

This is achieved through right profiling of the customers and identifying their specific needs. Promotions will be tailored to the profiled groups of the customers, offered at the right time to the right customer, in turn improving the customer stickiness to the platform.

Methodology: To enable timely intervention of retaining valuable customers

- An RFM (Recency, Frequency, Monetary) based Markov Chain model is built. This uses customer's existing transactional behaviour and helps capture Customer Lifetime value and Churn.
- Overlaying this analysing with clusters from k-means clustering, to help create optimal clusters for effective personalisation.
- A churn prediction model based on Logistic Regression, will be built to augment and validate the churn predictions from Markov Models.

The expected outcome of this project is to achieve: *Targeted Campaign Designs to Reduce Predicted Churn and Improved Customer Retention.*

PROJECT TITLE: Building A Preventive Detection System for Battery Drain Issues for Bounce

ABSTRACT: Bounce, a pioneering keyless and dockless bike rental company, has a fleet of more than 10,000 bikes in Bengaluru. A seamless customer experience where users book a ride on the app, pick up the bike, enter the OTP, ride to the destination is key to Bounce's success. There are various checks that the Bounce's product team has put in place to ensure that bikes in unsatisfactory condition are flagged immediately to the servicing team and is made 'unavailable for hire' on the Bounce app.

Despite having multiple checks, unpredictability of Battery drain is a sizeable reason for unavailability of bikes for Bounce which affects customer delight and fleet utilization. To address the problem, The Company wanted to build an early warning system that can inform them which scooters are at a high risk of a complete battery discharge. They wanted a system robust enough to outsmart sensors issues and battery swaps, without heavily relying on manual process of referring service logs.

A rudimentary classification technique helped in dimensionality reduction to derive various states of the bikes in the defined period. The model then used multivariate unsupervised classification techniques to cluster bikes based on past behaviour data as derived from multiple sensors along with known lead acid battery research. The flagged bikes were further investigated for sensor anomalies using clustering, finally arriving at the highly probable battery failure cases. Post model deployment an accuracy of about 70% was observed further strengthening the model using semi supervised modelling techniques.

PROJECT TITLE: FPL Squad Planner for Each Gameweek Using Prescriptive Analytics

ABSTRACT: The Premier League is the most-watched sports league in the world, broadcast in 212 territories to 643 million homes and a potential TV audience of 4.7 billion people. With over 6 million players, Fantasy Premier League is the biggest Fantasy Football game in the world. It's FREE to play and you can win great prizes! Around 500 Football Players play in English Premier League in a season.

Mostly users/players of FPL spend time doing research, buying multiple Squads in order to increase their chances of winning. Its tedious to select 15 players of 511 players every week within a budget restriction of £100 million. Every player has a 'Price tag' attached. There are other restrictions/constraints involved while building FPL Squad.

Our objective is to prescribe a team to users to increase chances of winning. Considering each Player performance of last 2 seasons based on points they have scored. Our objective is to maximize the points of any given "game-week" consisting of 38 matches. Along with several constraints as per the rules of the game , we will be applying cost/weights to factors like crucial high scoring matches ,Individual player performance against specific teams and performance when players play home or away game in order to come up with an optimised team recommendation. Multi criteria decision models like Goal programming will be driving the recommendation.

Goal is to test the prescribed team performance with current on-going season competition. Model can be fine-tuned, and accuracy can be determined by testing across several game-weeks of the current season play. This model can be deployed or scheduled to run every week prior the game week begins.

PROJECT TITLE: Leveraging AI in a Skilling Ecosystem

ABSTRACT: This project is aimed at infusing AI into eKaushal, which is the digital skills platform owned and managed by TransNeuron Technologies. eKaushal is designed to bring various stakeholders (Govt, Training providers, Candidates, Assessment agencies, employers) onto a single platform that can help provide visibility and momentum to the Skills Initiatives and schemes launched by the government.

Our project is specifically focussed on developing a recommendation engine for courses that candidates (youth) can enrol through the eKaushal platform – based on historical data and various attributes related to the individual or potential job profile and/ or the region they hail from. AI/ML approaches that are considered include clustering, classification (using a combination of candidate and job-related attributes) and natural language processing (for extracting job-related characteristics).

The recommendations are expected to help the candidates make a more informed and meaningful decision on which courses he/she can enrol onto and enhance the chances of their successful course completion, employability and job placement. This in turn would have an impact on the last-mile outcome realization – given the investment being made by the government towards skills and employability of India’s youth.

PROJECT TITLE: Making Energy Intelligent: AI-based interventions in the oil and gas sector

ABSTRACT: The price volatility in oil and gas is a significant concern for the oil and gas sector across the globe. The supply-side factors influencing the price volatility and the methods to augment productivity in the sector is mostly manual and based on subject matter expertise, especially in today's AI-driven environment. This unlocks the possibility of significant scope for AI-based interventions for monitoring and enhancing the efficiency in the sector to manage costs while controlling emissions to be more environmental-friendly. These AI interventions in the sector could range from drilling and extraction aspects to price determination.

Given this context, this project aims at analysing the oil and gas wells to identify their current status and to facilitate clustering of similar wells based on various characteristics that include location, productivity, and profitability. Employing the hourly data of 48 oil wells over ten years, it attempts to identify and predict anomalies and events at an oil/gas well that could impact production. Furthermore, it attempts to construct a forecasting model for oil and gas production to help make maintenance decisions regarding the oil/gas rigs that are in operation. It also proposes a model for predicting the oil price and examines how the volatility in oil prices can impact the well production cycle. The study would then provide ground for identifying the inter-relationship between the production in one well and all the other wells in the same reservoir so that the production decisions can be optimised. The study expects to provide trained and validated models with actionable outputs with the highest possible accuracy.

PROJECT TITLE: Manufacturing Analytics for Carpet Industry

ABSTRACT: Client asked to study their customer attributes and past purchases and recommend actionable insights for customer through useful and innovative dashboards, which could aid in decision support for the management.

Ordering patterns of existing customers were studied, along with their purchase patterns. Dealers were clustered into three categories using distance-based algorithm. These clusters were further analyzed for specific purchase patterns. Product purchase patterns were analyzed using rule-based mining algorithms, that would help in recommending relevant samples to prospective and existing customers. Similarity between two users were calculated for better recommendations, this was achieved using cosine similarity. Specific patterns were observed for different users and recommendation was generated for sales prospects. Products had various unique attributes which resulted in a large number of unique products as the business is a made-to-order business. To cluster these products, density-based algorithm was used on product attributes. Once the clusters of products were made, the same were analyzed on purchase patterns and revenue distribution.

Manufacturing sector is going through immense stress in current times. This could be due to lack of appropriate analytical techniques specially in the handicraft segment. The techniques employed in the current project could be used by various MSMEs operating in handicrafts domain. These techniques are focused on increasing the revenue by giving an opportunity to upsell products to potential and existing customers.

PROJECT TITLE: Operational Analytics for Leads Conversion in Real Estate

ABSTRACT: Real estate sector in Bengaluru has witnessed tremendous growth over the past decades. In the recent years, reforms like demonetization, RERA, GST disrupted the real estate business and resulted in sales slowdown. With the fierce competition, real estate developers find it increasingly difficult to retain or increase their customer base. Mana Projects Private Limited (MPPL) allocates a major portion of their budget for a project on marketing and sales activities. Digital marketing increases lead quantity but it comes along with challenges related to their quality as well as a low conversion rate of prospect to customer when compared to the spend. In addition, MPPL senior management does not have end to end visibility of prospect to booking life cycle and are therefore looking for ways to optimize operational efficiency.

The project aims to develop a data driven predictive model to identify potential leads that are likely to convert and reduce lead conversion time, optimize marketing spend and lower customer acquisition cost. A three months' timeframe of data for three projects (status of ready to move-in, on-going, recently launched) will be considered. We will start off by consolidating the data sets pertaining to marketing, leads, opportunities and booking data; conducting exploratory data analysis to discover patterns, detect anomalies, test hypothesis and gather insights. Then, logistic regression model will be developed for feature extraction and selection will be applied to determine variables that impact sales conversion. Finally, model will be deployed at MPPL for their on-going projects and further tweaked based on usage.

PROJECT TITLE: Operational Risk - Dynamic Claim Allocation at Revenue Cycle Management of an US Healthcare.

ABSTRACT: The company processes over 300 Million transactions annually to get accounts receivables for providers of US Health care sector, by making manual calls/mails to Insurance provider. In these calls they provide required information and update the status of the claim. On each day new claims are picked up and worked upon. The claims are manually allocated to the employees on daily basis to make calls to the customers and then the outcome of the calls is recorded. According to the SLA employee should work on the claim within 20 days duration. By the end of 20th day the claim needs to be addressed. If the claims are not looked up for the said number of days, the claims fall under non – compliance.

This may be due of either lack of time availability or complexity or by error, the claims would be moved Recall out kick codes (As it was not addressed by end of 20 days). Pushing to Recall, doesn't address the claim, but mostly pushes the claim out of the bucket only to get them back in the queue on a later date (This accounts a part of inflow in a later date).

The project aims to predict the movement of a claim to Recall on a future date, much ahead in the timeline, so that appropriate action can be taken. We are looking at using the Random Forest, PCA, TPM, Graphical Clustering, forecasting techniques, etc to identify the claims that could move to Recall.

By identifying the claims that would move to Recall, we aim to help the company to identify the claims and take appropriate action to address them earlier in the time line and adhere to their client compliance.

PROJECT TITLE: Optimization of Rewards & Recognitions Contests for A Reputed Insurance Company Using Predictive & Prescriptive Analytics.

ABSTRACT: Insurance companies conduct various Rewards & Recognitions (R&R) contests for its employees that are designed to motivate its agents to bring in higher policy sales. These R&R contests consist of various premium slabs & agents who might achieve higher slabs of policy sales will get increasing rewards which can be both cash and non-cash. Measuring the effectiveness and efficiency of these contests is quite challenging. Currently these R&R contests are designed manually based on the judgment of business teams and not data driven, which usually results in sub-optimal lift in sales resulting from R&R contests.

Hence, to tackle these issues, the current project aims to develop analytics driven data based approach to designing the R&R contests that will maximize the lift of policy sales generated for each month. To achieve this, the project would try to predict the policy sales for the next period and obtain the optimal spend on R&R investment month wise. Once this breakdown is obtained, the monthly investment breakdown is further drilled down to obtain the break-up of investment in each reward category. Once this detailed break-down is obtained, the optimal slabs for each contest are derived based on agents' estimated performance.

By optimizing the R&R contest slabs & maximising the returns, significant improvements can be made in both the bottom line and top line of insurance companies.

PROJECT TITLE: Optimizing the Discharge Process by Tuning the Discharge Process Time for a Multi-specialty hospital.

ABSTRACT: An efficient discharge Process in a hospital system is essential to ensure high satisfaction level of customers (patients) and yield high occupancy and revenue. The discharge process is interdependent with other important processes such as admissions, billing etc. Currently, the Discharge Process at a leading hospital system in India is not optimized and it takes around 7 hours on an average for a patient to get discharged leading to lower NPS. The objective of this project is to optimize the discharge process by precisely predicting the discharge turn-around-time (TAT).

Along with bringing transparency in the discharge TAT for both patients and the hospital, through model interpretability we aim to identify most important processes and bottlenecks in the system and recommend operational changes in those processes to make the discharge process faster.

The intended methodology being considered for the prediction is supervised learning algorithm to predict the Discharge TAT using the patient and hospital administrative. Initially starting with models like linear regression we would then move to sophisticated non-linear models for high accuracy that might come up with a cost of lower interpretability. However, with newer methods of model interpretability, the reasons for delay in discharge process can be identified and optimize by 30% - 40%.

The final goal of the project is to improve the discharge process efficiency and see an observable increase in customer satisfaction and achieve high NPS for the discharge process.

PROJECT TITLE: Predicting the Risk of Diabetes, Hypertension and Cardiovascular Diseases in Patients Based on Routine Blood Tests.

ABSTRACT: Risk of Non-Communicable Diseases (NCDs) which includes Diabetes, Hypertension and Cardiovascular Diseases is generally identified through a set of tests which analyses various body fluids and computes a risk score. This process is a very thorough exercise which also includes capturing various data points through a personal questionnaire aimed at gathering information about a person's physical status and living habits. This is dependent upon the selective and presumably correct information disclosed by the patient. Most of the times such detailed personal information is not available. Currently, the SMEs determine risk levels with the help of readily available online models (built using foreign population data). Our client, who is in personalized healthcare domain, wants us to come up with a predictive model which will help them ascertain the risk level for patients based on the blood test results only in absence of other personal information. This will include identification of a select few parameters captured through the blood tests which will have a high impact in determining the risk level. This will help the client unearth new insights which may not be identifiable through SME inputs and equip them with a heuristics-based model for risk estimation. The approach would include developing classification models to determine risks of each type of NCD.

Through this model, the aim is to develop an efficient method to determine the risk level of NCDs with minimal number of parameters. Identification of these parameters in determining the risk will also help cater to a larger clientele where personal information might not be available.

PROJECT ABSTRACT: Prediction of Equipment Breakdown at a Commercial Facility Firm using Analytics

ABSTRACT: Equipment Breakdowns at any facility is one of the major problems in any commercial building. The client currently uses different categories of Mechanical & Electrical equipment (E.g. UPS, Transformers, Chillers, HVAC, Batteries etc) that are deployed in a facility operation.

Currently several maintenance activities are being done manually after a failure occurs and no proactive solution is in place. The information is currently captured by the service personnel in log sheets of each of the equipment installed. These logs captured manually are verified later once there is a failure of an equipment, which is not scalable. Therefore, Project aims to focus on developing a data driven model considering the failure events and data points leading to that failure event using a logistic regression model approach.

The Main approaches we will be considering is to design an advanced early warning system model using the existing data collected over a period of equipment condition. Since the operational life span of production machines are usually several years, historical data reflects the machines' deterioration processes and capture multiple failure events.

Also by automating the data collection process for the entire facility, it aims to help the client to handle such breakdowns and keep the SLA between their clients intact , save replacement cost and increase the efficiency of the machines thereby reducing overburn of the equipment.



PROJECT TITLE: Stock Price and Investment Action Prediction Through Sentiment Analysis of Earnings Call

ABSTRACT: Predicting stock price has always been a major area of study within financial research. The Research Analysts have predominantly used Fundamental approach (based on company’s financial and non-financial data and macro data) or Technical approach (historical stock price data) to predict stock price over different time horizons. Stock prices are difficult to predict given inefficiencies in the market and irrational market behavior. In recent times we have seen emergence of related fields of study such as behavioral economics, text analytics etc. to aid stock research.

With recent developments in NLP and large scope of unstructured data analysis, the analysts are increasingly incorporating textual data available regarding the company (such as earnings transcripts, annual reports, press releases, news articles, social media feed etc.) into their analysis; with a view to uncover new material information which can drive stock prices.

The objective of this project is to develop Management Sentiment Indicator (BUY, SELL or HOLD) from quarterly earnings call transcripts using NLP for 5 US stocks in Financial Sector. We will test if sentiment level or change in sentiment level is predictive of future stock returns. If predictability of future stock returns (1-day, 2-day, 5-day, 10-day) being explained by the management sentiment/change in management sentiment is established, we can reach the conclusion that there exists a window of investment opportunity for an investor to add alpha; basis the management sentiment score (or its change) of an earnings call.
