13th Symposium on

Business Analytics and Intelligence

2022 – 23

BOOK OF ABSTRACTS

LIST OF PROJECTS

Title: Accident Risk Analysis using Geospatial Predictive Modelling

Abstract: There were around 32.7 million passenger cars in operation in the United Kingdom in 2020. The increasing number of vehicles on the roads has led to an increase in the frequency of accidents and casualties, resulting in higher insurance claims and payouts for insurance companies.

To proactively plan for potential losses, insurance firms are utilizing accident data to assess the risk across various geographical units. One such measure used is the Accident Risk Score (ARS), which is the ratio of the sum of casualties and the number of accidents. Insurance companies can reduce their risk exposure by deploying different strategies for different postcodes depending on their ARS score. For example, insurance company can quote higher premium for postcodes with higher ARS.

Accidents occur for a variety of reasons, from driver error to environmental factors. Using ML algorithms that leverage historical accident data and geospatial datasets can help reduce the number of factors that contribute to accidents. The project aims to develop an ML model to predict the probability of accidents happening in different places and times based on shortlisted features.

The project aims to achieve the following objectives:

1. Perform feature engineering by leveraging publicly available geospatial datasets like road infrastructure, population density, elevation, weather etc.
2. Leverage machine learning techniques and geospatial data analysis to predict number of accidents/ casualties by postcode.
3. Generate ARS (by postcode) by leveraging the client's data and engineered features.

Title: *Brand Score*: To calculate brand score for a set of keywords basis Twitter, Redditt and Google Trends and YouTube data to enable campaign planning

Abstract: The client, a leading Programmatic Media Company, helps its clients/marketeers around the globe to plan marketing campaigns underpinned by the sharpest data insights. To enable this, the client has a campaign planning & management system which helps client's teams their media and digital campaigns.

This project aims to develop a system that takes data from Twitter, Redditt and Google trends based on a set of keywords for a particular brand provided by user and pulls data from these social media platforms around those keywords. The system will further do analysis on the pulled in data and calculate a score for the brand keyword. This score will fit in the larger campaign planning and management system of MiQ to help the team do digital advertising campaign planning around:

- Top keywords on which to spend.
- Trends around day/time of the week when a keyword is trending.

Further, the project will leverage contextual similarly algorithm by mapping the channel description with the keywords to identify the YouTube channels that the brands may target.

The project aims to enhance the campaign performance and improve ROAS (Return on Advertising Spend) by leveraging the data insights around selected keywords.

Top keywords would help the marketeers narrow the keywords used for a marketing spend and help improve metrices like CTR and ROAS. The trends around day/time of the week, month etc. would help them run the campaign when the target audience is most active and typically searching around those keywords (Because of cyclical nature of business, festivals etc.).

Since live data with free APIs is limited, for purpose of project, the analysis will be limited to predefined 3-4 brands and associated keywords. Also, only parameters and data that can be pulled using free APIs will be leveraged. However, the algorithm will be scalable and can be connected to live APIs in future to get real-time results.

Title: Food for Thought – A Data-Driven Approach to reduce food waste and to optimize Waste Management System

Abstract: Food wastage is a significant issue in the world and it is estimated that one-third of all food produced for is wasted, considering that over 800 million people worldwide suffer from hunger. It results in economic losses for producers and consumers and also has environmental impacts. Furthermore, the United Nations has set a goal to reduce global food waste by 50% by 2030 as part of its Sustainable Development Goal of responsible consumption and production, aiming to promote sustainable use of resources and reduce waste. Also, In Oman, 470K tonnes of food are wasted every year, while the wastage per person is estimated to be 95kg and hotels in Oman throw away food worth 45 million Omani rials (Dh429m) every year.

Food waste and poor waste management system in Oman have several negative effects on the environment and society such as:

- Greenhouse gas emissions
- Wastage and pollution of natural resources
- Economic consequences and increase in expense for food industry and consumers

This project aims to develop a Prototyped Intelligent Waste Management System for food industry using different Innovative solutions with the help of technology and data-driven approaches resulting in:

- Forecast food demand and predict diner usage
- Minimize food wastage
- Optimize supply chain and food inventory

Thus helping improve waste management practices and reduce the overall food waste by leveraging data visualisation tools such as Power-BI and Tableau and using Advanced Data-Science Models such as Regression, Time-Series, Ensemble Models, NLP, Six-Sigma and Supply Chain Management Efficient waste management system will aid the creation of more accurate inventory management and waste reduction strategies thus resulting in

- Reduction in food wastage through different sources viz Kitchen waste, plate Waste and counter waste
- Reduction of costs and maximizing profits by optimizing food production from the menu or uncover any behavioural aspects and factor other hypothesis (Festival season, holidays)
- To correlate if one cuisine has positive or negative impact on other (When served together)
- Predict diner footfalls throughout all meals across all days

Title: Forecasting sales of 500+ unique products through evaluation of multiple models for each product

Abstract: Integrated business planning services adopt innovative digital technologies and applied intelligence platforms to offer several custom-fit supply chain planning solutions. These allow attainment of virtual economies of scale without the upfront investment in people, processes, and equipment.

Forecasting accurate sales is a part of integrated business planning that brings in proficiency in planning, supply and operations.

This project aims to develop machine learning models using Python to generate sales forecast for 500 products, improve forecast accuracy and reduce absolute error. A month on month forecast for a total of 12 months in the future is required for each of the 500 products. Accuracy should be >10% from the baseline accuracy for the last quarter. Monthly Sale quantities in units sold is provided for 5000 unique products for a period of 36 months from October 2019 to September 2022. The products are mainly machine parts.

Exploratory data analysis was conducted to identify a) the patterns of the sales for products in individual categories b) any obvious impact of COVID c) commonality between categories and d) relation of units sold to consistent sales. All products do not have sales in all months and there is a wide distribution of units sold. Some products have sparse data (no sales for months at a stretch) and these were dropped. Data imputation was performed where it was seen as truly representative and appropriate.

Multiple time series forecasting techniques from moving average (MA) to ARIMA and SARIMA are evaluated on each individual product to identify optimal accurate model(s). As an alternate approach to forecasting based on past data, FB Prophet which predicts based on change points in the data set will also be evaluated. A single model for each product might not be a robust strategy as new data might require re-training of the model to arrive at accurate forecasts. Therefore, our deployment strategy would be to have a pipeline of 2-3 models for each product, automate hyperparameter tuning, evaluate the best model based on pre-determined metrics and implement a recursive predictive loop to forecast 12 months into the future. New data can be fed into the pipeline and the best forecast arrived at based on pre-defined decision metric criteria to select the best model.

Title: Forecasting Ticket Sales for a Tour Operator

Abstract: The boat tour industry offers a range of unique boat tours to suit the preferences of individual, group, or corporate customers. These tours are available daily, and tickets can be purchased online or at kiosks. The factors such as weather, holidays, events and promotions can impact tour operations and visitor experience, making it challenging for the tour operators to manage and optimize human resources effectively without understanding crowd patterns.

The project aims to build a forecasting model which assist in the following:

- Footfall analysis to plan special events, promotions, and campaigns for maximum exposure.
- Resource planning to manage employees, contract workers and volunteers.
- Demand planning for food & beverages thereby improving customer experience.

The approach for building the model includes Exploratory Data Analysis (EDA), correlation study, understanding the seasonality, data drift due to covid, univariate and multivariate ML based forecasting models using the historical ticket sales data of the past five years.

This project is intended to build accurate sales forecasting model that will enable the tour operators to optimize their business processes and gain a competitive edge through data-driven decision-making.

Title: Hospital Re-admission Probability Predicition

Abstract: Hospital readmission is when a patient who is discharged from the hospital, gets re-admitted again within a certain period. Hospital readmission rates are now considered an indicator of hospital quality and affect the cost of care adversely. For the hospitals which are currently penalized due to high readmission rates, one solution is to create interventions to provide additional assistance to patients with an increased risk of readmission. The challenge is to identify patients with an increased risk of readmission. Additionally, identify the high readmission risk patients at various patient journey touchpoints.

We can use predictive modelling to calculate the probability of readmission for a patient at various patient journey touchpoints. The journey touchpoints include in-patient, discharged to care-home, and discharge to home. We will be using the following patient's data points:
1. General information: age, gender, race, location, education & insurance details.
2. Health history (last 12 months): allergies, medication, hospital visits, diagnosis, smoking & alcohol habits.
3. Present encounter: Admission type, diagnosis & treatment details, vitals, lab reports, & discharge information.
4. Post-discharge: care-home details, prescribed medications, lifestyle, follow-up visits, caretaker details.

The steps to be followed to build the predictive model are:
1. Hypothesis formulation & data exploration.
2. Feature engineering.
3. Building training, validation & test data samples.
4. Model selection & evaluation

The objective is to develop an explainable predictive model to predict patients' readmission probability so the hospital can take preventive actions to avoid readmission.

Title: Predicting Warranty Claim Decision for Automotive Parts Manufacturer

Abstract: Processing warranty claims is a critical process for any equipment manufacturer as it enables the customer to report and resolve any issues in the equipment, thereby increasing customer satisfaction and brand loyalty. It also provides valuable feedback to the manufacturer about the quality and performance of the equipment. Analysing warranty claims can help the company to identify recurring issues and thereby, making improvements in the product.

However, processing the warranty claims is a tedious and time taking process. It takes ~4-5 months on an average to process a warranty claim submitted by the customer as the product is initially inspected by the engineer at the customer's premise for any visible defects, then shipped to the manufacturing plant for detailed root cause analysis based on multiple functional tests and thus, warranty claim decision is made. i.e., the warranty claim is accepted or rejected by the company.

The project aims to build a predictive model for the automotive parts manufacturer to take warranty decision, i.e., whether the claim will be accepted or rejected, basis customer complaint and basic equipment details at customers' premise only (without shipping the equipment to manufacturer's plant).

For cases where warranty decision cannot be predicted accurately with available details, model aims to suggest preliminary tests to be performed by the engineer at customer's premise, which may help to arrive at the warranty decision. This will also reduce the effort involved for product inspection as the engineer will be required to perform only specified checks for each claim instead of performing all checks for all claims, thereby, optimising the efforts of the manufacturer.

The model will be developed by analysing historical claim database to identify attributes critical for making predictions. Natural Language Processing (NLP) techniques will be used in data pre-processing to extract relevant information from 'customer complaint' details. Thereafter, machine learning algorithms like logistic regression, decision trees, random forest, or a combination will be used to build and evaluate the model to predict warranty decision.

The development of a predictive model for warranty claims will significantly improve the efficiency and lead time of the warranty claim process. It will help the manufacturer to maintain strong customer relationships by providing prompt and accurate resolution of equipment issues, while also reducing costs and streamlining the warranty claim process.

Title: Pricing Analytics for Automobiles and Automotive Components

Abstract: The Indian automobile industry has been growing rapidly in the last few years, with a significant increase in the number of passenger cars on the road. As a result, it is imperative for vehicle component manufacturers to understand the car price trends and macroeconomic factors affecting car prices in India. The aim of this study is to explore the correlation between macroeconomic factors and passenger car prices in India, and to track the component price trends with respect to the car price movement.

The study involves collecting historical data on macro-economic factors such as development indicators, interest rates, metal prices, and passenger car prices in India. In addition, historic data on the manufacturer's prices of vehicle components such as fuel pumps, common rail fuel injectors, electric control units etc. will be collected and analysed to determine the trends in component prices relative to passenger car prices.

The study aims to derive and validate valuable insights from the historical data using data visualisation, statistical inferences, correlation analysis and regression analysis. Also create an interactive dashboard to help visualise the outcomes of this study.

Overall, this study will contribute to a better understanding of the macroeconomic factors that influence passenger car prices in India and provide valuable insights for the manufacturer in determining pricing strategies and enable data-driven decision making to optimise product prices while staying competitive in the market.

Title: Uncovering the Voice of The Customer: Leveraging AI Tools to Analyze, Track and Measure Data.

Abstract: To succeed in a competitive market, businesses must understand customer behavior and drive customer satisfaction. Customer feedback is available through multiple channels, particularly unstructured text data from sources such as social media, surveys, and chat transcripts. Analyzing this data is critical, as it provides insights into buying trends, product flaws, customer interests, and marketing opportunities. In this project, we propose the use of Deep Learning and Natural Language Processing (NLP) technologies to better analyze contextual information and capture customer feedback.

The project aims to provide a significant business advantage by:

1. Minimizing time taken to read the comments / feedback / verbatims
2. Maximize time for business to take data driven decisions

By developing state of the art algorithms for Text Polarization and Topic Categorization thus enhancing the overall customer experience

To achieve these objectives, we will juxtapose the results obtained from traditional models, ML algorithms (supervised and unsupervised), deep learning including transformer models. We explore Ensemble models which combine the predictions of multiple models and transfer learning to produce more accurate and robust results.

Accurate sentiment analysis and categorization of customer feedback allows businesses to identify customer pain points and take appropriate decisions resulting in driving revenue and improve brand reputation.