9th Symposium on
Business Analytics and Intelligence- 2019

BOOK OF ABSTRACTS

9th Symposium on Business Analytics and Intelligence - 2019	
Sl. No	**List of Abstracts**
1	Credit rating model for rating of unrated companies
2	Named entity recognition using machine learning for a multinational insurance giant
3	Transport demand modelling using cell transaction records.
4	In-season markdown optimization for fast fashion products with short lifecycles.
5	Forecasting sales of aftermarket products
6	Personalized ancillary product recommendation for airlines
7	Study of Air Quality in India
8	Demand forecasting and inventory optimization of blood units for the blood bank at VHS
9	Analytical customer relationship management for bank
10	HR analytics at multi-specialty tertiary care hospital
11	Advanced data analytics for modelling and prediction of real-world two-wheeler emissions
12	Forecasting warranty claims for different variants of a two-wheeler company
13	Neural network-based OCR model to recognize handwritten characters
14	Identification and estimation of remarketing campaign parameters for an online real-estate platform
15	Automatic blade defect detection
16	Shelf assortment model for independent stores
17	Optimization of logistics cost at an automotive supplier company using prescriptive analytics
18	Comparative study between generalized linear modelling and gradient boosted machines, in claim frequency models
19	Resume ranking using text analytics
20	Agricultural yield prediction and interactive dashboard for visual analytics.
21	Machine learning model for document labeling.

1) PROJECT TITLE: Credit rating model for rating unrated companies

ABSTRACT: A credit rating is an evaluation of the credit risk of a prospective debtor (an individual, a business, company or a government), predicting their ability to pay back the debt, and an implicit forecast of the likelihood of the debtor defaulting. The credit rating represents an evaluation of a credit rating agency of the qualitative and quantitative information for the prospective debtor, including information provided by the prospective debtor and other non-public information obtained by the credit rating agency's analysts. In India, there are a large number of companies that sit outside of the Credit rating agencies purview & thereby are not rated. This lack of rating forces lenders to do independent evaluation with not enough data points or resources. The project aims to build a credit rating model based on companies with existing credit rating. Quantitative data such as historical annual financial statement & qualitative data such as compliance matrix & key management personnel credentials will be used for building the model. A multitude of classification techniques (including logistic regression, random forest & neural net) will be utilised to accurately classify the company to a particular credit rating. This model should help prospective lenders in gauging credit worthiness & lending accordingly.

2) PROJECT TITLE: Named entity recognition using machine learning for a multinational insurance giant

ABSTRACT: In the insurance sector, organizations need to process hundreds of documents and claims daily. This includes manual perusal of each claim and supporting documents, classifying into relevant categories and subsequent processing. Additionally, periodic analysis of data must be conducted to identify patterns and derive insights from it. Given that a large chunk of the business is still paper – based and forms filled, and claims applied physically, digitalizing all this information can be an extremely daunting task which is often plagued by human errors.

Given this context, it is of paramount importance that a reliable system be developed to translate important information from the physical to an intelligible digital format. There are three steps to do this, namely, optical character recognition (OCR) to read scanned physical copies to convert to their digital counterparts, topic classification to ensure same topic documents are clubbed together and finally named entity recognition (NER) which is used to extract relevant context specific text entities from these digital documents.

This project aims to tackle the NER portion of the problem. The idea is to develop a language agnostic NER model which can be used to extract common entity types such as name, gender, age etc. along with context specific entity types such as injury, severity etc from the text data. The primary approach would be to employ Deep Learning workflows using pre – existing libraries such as SpaCy, NLTK etc. to train a classifier which can be then used in varied business contexts via transfer learning.

3) PROJECT TITLE: Transport demand modelling using cellphone transaction records (CTRs)

ABSTRACT: In a city where rapid urbanization has led to overbearing stress on transportation system, the need to analyse the existing system and design sustainable mobility solutions that will further aid economic growth, cannot be understated. Realistic Travel Demand Estimation is an integral part of analysis and design of sustainable mobility systems. The travel-demand model is a mathematical model between travel-demand flows on one hand, given activity and transportation supply systems on the other. The greatest challenge in estimating travel demand across various locations in a city is the limited understanding of where, when, why, and how people travel. Traditionally, this behavioral understanding is achieved through sampling surveys. However, in this project, anonymized cell phone transaction records (CTRs) are used to infer social behaviour that feeds into the demand model. Due to the ubiquity of cell phone usage data from CTRs have better spatial coverage & temporal component when compared to static traditional surveys. Using the data extracted from CTRs, this project attempts to estimate mobility metrics such as, static travel demand, dynamic travel demand, route choice, link flows and travel times. Visualization of the metrics is also attempted in order to give a better idea of travel activity / behaviour which in turn helps in decision making with regard to identifying transportation bottlenecks or identify areas for investment in infrastructure.

4) PROJECT TITLE: In season markdown optimization for fast fashion goods at top fashion retail company in India

ABSTRACT: Markdown optimization for Fast Fashion goods is to provide the right discount at the right time to the customers so that inventory is cleared, and profits are increased. Current scenario is that the company provides discounts on every item defined at style code level only during End of Season Sales due to which they end up giving huge discounts to clear the inventory. Also, it is not data driven and hence would result in possible sub optimal discounts given to clear the inventory. Therefore, the project aims to provide a data driven solution by satisfying all the rules set by the management of the company. Their objectives include clearing the inventory within 8 weeks of launch of a product. The proposed solution helps them to order the quantity of style codes based on the demand and also use the Optimal markdown to clear the inventory. Demand estimation will be done using Delphi Technique. Dynamic programming model is used to estimate the Optimal markdown for each week. It is a nested model which would propose the Optimal Markdown in each week based on the available inventory, by estimating profit in the current week and also by the end of the 8-week period. The scope of this project is limited to couple of style codes and important thing to consider while choosing the style codes is cannibalization of style codes as purchases behaviour of one style code can influence the purchase of other style codes. For example, in a particular store if a white shirt with a certain design/pattern is selling good at Full price without any discount and another white shirt with a slight variation in colour/design is not selling good due to which if discount is given to this product then this affects the selling behaviour of other white shirt selling at Full price.

By providing this solution, the aim is to optimize the markdown process while meeting the management requirements. This can help in gaining more profits and also optimization of inventory.

5) Project Title: Forecasting sales of aftermarket products

Abstract: This project involves analyzing the historical data (over 11,000 products) for the aftermarket division of a global automotive company and forecast their sales with good accuracy.

The Automotive **Aftermarket Parts** industry is a secondary industry to automobiles and deals with spares and service. Vehicles require replacement consumables, servicing items and replacements for failed or damaged components throughout their life. The aftermarket brings profitable advantages to manufacturing companies by providing value adding services for the costumers. Accurate forecasting of the demand of these products is essential to achieve a smooth material flow of spare parts.

Uncertainty in demand for spare part derives from huge number of stock keeping units (SKUs) and this can be mitigated through high inventory levels or responsive transportation solutions. Such approaches for dealing with uncertain demand are associated with high costs. Naturally, improving forecasts will have a significant impact in securing spare parts at the right place, in the right time and for the right cost. In effect it is possible to decreasing inventory costs without impacting the availability negatively. The availability of spare part is increasingly important for the automotive industry because it is moving towards providing solutions and services to its end-customer, rather than just selling a product. This point of view brings multiple benefits such as increased profitability and customer loyalty.

The current adopted approaches like market driven sales numbers, last quarter averages etc have yielded low accuracy, and also thrown challenges with forecasting for each and every product given the sheer number/scale of parts that were being manufactured, inventory stocking by clients.

This project will try to cluster large number of products (SKUs) that are logically connected into groups of similar series and the series in each cluster are then pooled together and their forecasts are obtained. Some of the methodologies for classification of products are business relevance, product categories, product lifecycle, and demand characteristics to create groups for large scale analysis across 11,000 different products.

6) PROJECT TITLE: Personalized ancillary product recommendation for airlines

ABSTRACT: Airline fares were all-inclusive until a few years ago with little to no "ancillaries". Over the last decade, the emergence of no-frills airlines has led to airlines unbundling their product to compete effectively. For example, in several countries, only the right to fly is the basic purchase. Adding a bag, ability to cancel or reschedule, adding insurance, more legroom, food & beverages are all ancillaries available at additional cost during or after purchase.

Users rarely login when shopping for airline tickets. Therefore, Airlines are often left only with unlinked session information (eg. Origin, destination, Departure and Arrival Dates, date of shop, number of travelers, cabin and type of trip (one-way, round-trip, multicity) and what was purchased (if a purchase happened). The session information along with ancillary purchase history will be provided.

Building a recommendation engine for a shopper to improve ancillary purchase conversions during a shopping session.

Recommending personalized ancillaries to clients based on their previous transaction behavior and also looking at 'LOOK A LIKE' clients is the key here. We would start with exploring the data to capture travelers' behavior and build a model for each ancillary product.

Step 1: Clustering clients based on purchase behavior and booking activity (through KNN or K Means)

Step 2: Personalized recommendations through: -

- Market Basket Analysis – Association rule (98% Clients purchased C along with A and B)
- Next Best Product (Probability of purchase)
- Collaborative Filtering (Recommendation engine – Cosine similarity)
- Multivariate Regression (Propensity for each Ancillary product)

7) PROJECT TITLE: Study of Air Quality in India

ABSTRACT: Poor air quality is one of the most serious environmental problems. Recent studies, with respect to adverse health impacts caused by exposure to polluted air, call for immediate actions to curb pollution.

India appears to have a very high proportion poor air quality by international comparison - host to 14 of the 20 most polluted cities in the world.

It has become imperative to come up with a meticulous study of the air quality index/ profile.

The project aims to study pattern of the critical components, which form a part of Air Quality. This data shall be obtained from various sources such as Pollution Control Boards, Internet Data Sets, Satellites, etc. We propose to study the behavior of these components to arrive at a relationship model, utilizing learning from correlations and regression techniques, imputation, etc.

A key challenge we envisage is the availability of reliable, complete and relevant (robust) data for the Indian states.

The intent is to feed the learning, to enable various policy related decisions, once the source & drivers of poor Air Quality (harmful) are understood in the coming phases.

Techniques of Imputation, Clustering, Correlation, Regression, Forecasting would be extensively used.

8) PROJECT TITLE: Demand forecasting and inventory optimization of blood units for the blood bank at VHS.

ABSTRACT: VHS hospital blood bank is the focus unit of a blood supply chain; and it has among its main assignments, the following activities: blood collection from donors, blood processing and blood component distribution to hospitals and healthcare institutions. It is surely the nervous center of the chain where planning, monitoring and controlling of these activities must be efficiently performed in order to prevent the stockout or outdate of blood components. All efficiency performance provided by a blood center will depend on the quality of its planning process, which starts with a good accuracy of forecasting required for blood supply and blood components. In this paper, a computational environment, which is oriented for forecasting of blood components, is presented. The idea is to improve planning of the inventory balance process of the blood supply chain.

Currently blood component distribution is undertaken manually and is time consuming and not data driven, thereby resulting in possible suboptimal usage of blood units. Therefore, this project aims to develop a data driven forecasting model that satisfies all the specific rules set by the management of VHS Hospital The main approaches we will be considering is forecasting blood demand and developing inventory optimization models. A 2-4-week planning horizon will be considered for the model.

By automating the demand forecasting process, the aim is to optimize and increase the efficiency of blood processing and distribution while meeting the hospital requirements such as minimizing waste and maximizing commercial effectiveness.

9) PROJECT TITLE: Analytical customer relationship management for bank

ABSTRACT: Syndicate Bank intends to implement future ready and agile Analytical CRM in order to provide efficient customer experience and aid the bank in increasing revenue. It is an initiative in the right direction towards becoming and analytically driven Organization thereby improving the operational efficiency by having overall view of business and facilitating rapid performance assessments and decisions to drive business faster.

The goal of the project is to enable Bank in becoming Customer Centric by focusing on Customer Satisfaction and Customer Lifetime Loyalty by uncovering the hidden information in raw data by converting the same to actionable intelligence. The model is expected to provide Bank with a better understanding of customer lifetime value, profitability to drive business growth, analytically derived Next Best Actions (NBAs) and Next Best Offers (NBOs) at Customer household/segment level, generate detailed information on the channel usage such as preferred channel, most value added/profitable channels, frequency of use by channel, channel costs and profitability, etc. This sort of analytics will help Bank to leverage the channel analytics information while planning changes in the existing channel structure, introducing new channels and migration of customers to low cost channels etc.

Based on the Customer analytics, segmentation analysis and LTV analysis Bank will be able to design campaigns for Cross-sell/Up-sell, churn prevention etc., that will provide the next best offer which the customers are most likely to take up.Integration of the Analytics Platform with Marketing Performance Tracking as well as Feedback Analysis will help the Bank to assess Marketing ROI thereby optimizing the Marketing process as well as to reduce the marketing expenses and effort.

In the long run the ACRM solution will help the Bank in generating a single view of the customer across all systems, touch points and channels viz. customer's activity, Average Balance, expenditure patterns, product penetration, Channel preference, etc. thereby enabling Bank to run more focussed campaigns and to achieve improved product penetration as well as improved product mix in the portfolio.

10) PROJECT TITLE: HR Analytics at multi-specialty tertiary care hospital

ABSTRACT: The nurses form an important cog in the hospital's machinery. Multi-Specialty Tertiary Care Hospital has a pressing need to optimize the number of nurses across its wards and specialty units. The hospital primarily faces issues with unplanned absences, attrition and not having the right ratio of nurses to patients, which has led to an increase in the workload of nurses. The hospital has identified some of the areas of attrition but would like to ratify those via the usage of analytics as well as check if there are other factors which have not been inherently apparent. Another key decision at the Hospital needs to take will be to decide on how fast to replace nurses who have attrited, depending on the predicted occupancy rates. Our primary goal will be to optimize the number of nurses required at the Hospital and also plan for a buffer; the project will also cover ancillary tasks of confirming the ratio of the number of nurses required per ward by using a hypothesis method to check its validity. We would be using predictive methodologies to classify nurses who have a higher probability of quitting. This information will feed into the optimization algorithms and help provided a holistic set of parameters for decision making.

The end objective will be to provide Multi-Specialty Tertiary Care Hospital with an algorithm which will enable decision making around speed of hiring and the optimal number of nurses required to ensure that Hospital has the right set of nurses with extensive experience. This will reduce the workload of nurses which will lead to better service of the patients and thereby meets the hospital's end objectives.

11) PROJECT TITLE: Advanced data analytics for modelling and prediction of real-world two-wheeler emissions

ABSTRACT: India is a country with 10 most populated cities of the world, and this is one distinction we should not be proud of. Vehicular emission is a major contributor to the worsening air quality of Indian cities and taking a toll on people's health arising the need for a stricter norm that could reduce the emissions considerably and put India on track to meet the BS VI standard.

Bharat Stage VI (BS VI) is an emission standard that will bring much needed changes in the Indian automobile industry in terms of pollutant emissions. BS VI norms include a wide list of technology modifications under the hood, the most significant being making On-Board diagnostics mandatory for all vehicles to make sure that the emission control component work at its optimum efficiency. As Indian regulatory bodies set new emission-regulation standards (the BS VI standards will go into effect for vehicles in these categories manufactured on or after April 1, 2020.), there is a need for Indian sub-continent to relook the way automobiles are produced and consumed and design solution to keep a check on pollutant levels.

Our objective is to build Machine Learning models considering the patented Bosch Physics based models to predict the emission amount in different vehicle riding scenarios so that the prediction solution can be used to prescribe solutions for reducing the current automobile emission levels, improving performance of vehicle & fuel consumption and uncovering the hidden insights in the data which can further enhance/corroborate business opportunities.

12) PROJECT TITLE: Forecasting warranty claims for different variants of a two-wheeler company

ABSTRACT: Guarantee, warranty, upgrade or repair service forms a critical part of after-sales support in automobile business. Simply tracking and monitoring warranty claims will only make the company backward looking in an era where the data is available to predict or forecast the future. Currently forecasting warranty claims activities is based on qualitative techniques and not data driven, thereby resulting in long service times for customers due to stock out of parts coming under warranty claims. Therefore, this project aims to develop a data driven forecasting model based on the historical time series data that is available and enable the organization to have an accurate view of the warranty claims thereby using this as an input to their inventory management activities to ensure availability of stock and reduce service time for its customers. The organization aims to improve its customer service and reduce its inventory carrying costs. Carrying higher inventories not an efficient option for improving the customer service. Reducing the error rate in forecasting the demand arising out of warranty claims is necessary for achieving optimized and efficient supply chain and ensure customer satisfaction. The main approaches we will be considering in forecasting the warranty claims will be time series forecasting, machine learning and deep learning models.

By deploying the forecasting model, it will provide inputs to the demand and supply planning teams and the production planning team for ensuring availability of spares and motorcycle models at the right place at the right time.

13) PROJECT TITLE: Neural network-based OCR model to recognize handwritten characters

ABSTRACT: In the insurance industry, there is a high demand for extracting information from the data available in printed or handwritten documents or scanned images to later re-utilize this information to take data driven decisions to achieve higher share market and customer conversions. Optical character recognition is an active research area that attempts to develop a computer system with the ability to extract and process text from images automatically. The objective of OCR is to achieve modification or conversion of any form of text or text-containing documents such as handwritten text, printed or scanned text images, into an editable digital format. Some of the font characteristics of the characters in paper documents and quality of images are only some of the recent challenges. Project aim is to extract handwritten or printed characters from scanned documents with high accuracy. The main approach we are using to handle the challenges and achieve high accuracy is to divide problem in image pre-processing, segmentation, normalization, feature extraction and classification. We are considering using open source library such as openCV and kersas to target poor image quality of scanned documents**. Pytesseract open source library has been used to identify coordinates of boxes and extracts characters from the box. We have found limitations of using Pytesseract therefore exploring neural network techniques to achieve better extraction of handwritten text.** Convolution Neural Network architecture will be used to extract features and classify them.

Goal is to achieve higher accuracy in recognizing characters in insurance documents to further process the data for entity recognition and predictive models.

14) PROJECT TITLE Identification and Estimation of Remarketing Campaigns Parameters for an Online Real-estate Platform

ABSTRACT: Remarketing is a form of online advertising that enable sites to show targeted ads to users who have already visited their site. Our client, India's leading online real estate information exchange platform, 99acres.com is a gateway to one of the fastest growing property markets of the world, an information 'exchange' for buying, renting and selling of all types of residential and commercial properties anywhere in India, uses this Marketing technique to target a section of its visitors who leave their website without leaving an enquiry (such visitors account for 80% of the total count).

There is a cost associated with advertising and generating enquires through this channel which is termed as Cost per Qualified Lead. CPQL is a primary KPI for the client, hence its minimization is of great importance. As of now, the client follows a majorly subjective and manual approach towards this objective which may lead to sub-optimal outcomes. Therefore, this project aims to develop a data driven analytics model that would help the client meet its objective.

Project's primary objectives would include -

1) Identifying major parameters/features that contribute towards variation in CPQL

2) Using these parameters and Campaign settings to bring the lowest possible CPQL.

After performing EDA, we will run clustering algorithms to identify customers which are more likely to respond to remarketing efforts which would ultimately help minimizing CPQL. Predictive analytics techniques that we will be used for CPQL minimization will be regression/decision trees to identify relevant parameters that impact the cost.

The final leg of this project would be to make these models flexible and reusable so that client can continue leveraging them for their future campaigns and ever shifting parameters.

15) PROJECT TITLE: Automatic blade defect detection

ABSTRACT: As part of Quality Control Process, each wind turbine blade in manufacturing (final product, sub-product and raw materials) passes through multistage manual inspections. The inspection time can vary depending on the blade length & product complexity. This approach is labor intensive & outcome may differ due to skill/performance variation between Inspectors. In few cases defect can escape between processes.

In today's world, Defects can be identified using computer vision (through images of the blade), thereby, reduce human intervention and manual errors/variation among inspectors.

The project would aim to develop a feasible solution on few high occurring defects that would be piloted in Test Lab (through a simulated shop floor environment). The deliverables would include defect detection methodology using camera, classification algorithm able to handle noise factors like product types, image quality, and dust during inspection, non-product background on images etc. The company would decide its implementation at a later stage after project closure.

16) PROJECT TITLE: Shelf assortment model for independent stores.

ABSTRACT: Optimizing the shelf space and product selection within the space provided in the stores has always been a critical requirement for producers and retailers since the correct assortment is critical to demand generation and increasing the sales. The shortage of shelf space, and the intensity of competition have magnified the importance of retail assortment and shelf-space planning. Increasing number of products, not just from competitors, even the different brands from same producer / different SKUs of same brand also conflicts with the limited shelf space. So planning the right assortment model makes it necessary to specify the type of brands to be placed at each shelf and determine the space and facings for each brand.

Currently, there is no shelf assortment model present for independent stores for a top beer manufacturer. So, an optimal recommendation for shelf capacity for each brand / store is the problem in hand. The assortment model should help in recommending optimal shelf capacity for each brand / store which helps in increasing the sales.

17) PROJECT TITLE: Optimization of logistics cost at an automotive supplier company using prescriptive analytics

ABSTRACT: Optimize the logistics cost by using the descriptive analytics to identify the key optimization opportunities and to specifically use prescriptive analytics for optimizing the outbound transportation network. Outbound transportation network planning and execution is a process of managing the shipping of several different products from several origins (factories) to **several** destinations (customer locations) through various transportation modes, vehicle types and facilities (warehouses, consolidation, cross docking, deconsolidation) to meet the service level goal of the organization at a minimum cost. This is a complex network problem considering the variety of products, origin-destination pairs and existing facilities.

Currently, they use 4flow for outbound transportation planning and execution. The transportation legs from factory to warehouse and from warehouse to customer location for a product have not changed in the past few years. The product variety, product volume and customer base has grown over these years and will grow & change in the future. This project aims to develop an outbound transportation optimization model to account for these changes and remain optimal on an ongoing basis.

As part of this optimization model, we will be evaluating the current network design and propose a new network in terms of flow of products through the same existing factories, warehouses and customer locations. We would be evaluating a greenfield approach of possible introduction of new facilities to identify if they will provide better optimization than the optimization of AS-IS network. We plan to use simulation to verify if the proposed network flow is better than the AS-IS network flow. The data collection for the optimization and simulation would involve historical data of shipment volumes (cubic) of all products through the current transportation legs, the associated freight costs, service level, cubic utilization of resources, lead times, on time delivery performance, vehicle types available, maximum capacity of vehicles and other critical data elements. Optimizing the outbound transportation would result into reduction in freight costs and better utilization of resources.

18) PROJECT TITLE: Comparative study between generalized linear modelling and gradient boosted machines, in claim frequency models.

ABSTRACT: This project is in alliance with one of the leading service providers in insurance and asset management domains across the world serving almost 88 million customers across 70 countries. Their main LOBs include Property & Casualty Insurance, Health & Life Insurance, Business Insurance and Asset Management. The insurance domain involves the provider to assess risk and accordingly underwrite policies for leads. Leads with higher risk of making a claim are typically charged a higher premium.

In order to scale and adopt a data-driven approach to the process of risk assessment, predictive modelling is used. Predictive models can be broadly divided into two genres – Statistical and Machine Learning models. Statistical models are traditional and form the premise on which machine learning techniques are built. The latter offers a higher degree of accuracy but is not as explainable as the former. This poses a unique challenge – Stakeholders from the business feel more confident about leveraging the output of a model that they can understand, while Data Scientists who build these risk assessment models recommend the usage of predictions from the sophisticated machine learning models which are more accurate.

This project using the data and mentorship this leading insurance company aims at bridging this gap by comparing the GLM and GBM models in terms of Fairness, Reproducibility and Explainability. Visualizing data using dashboards made on Apache Superset helps in explaining how different variables influence model predictions (GLM and GBM) which can be consumed easily by the stakeholders. Dynamic dashboarding and incorporating derived features using existing ones, help increase the ease of consumption. The last leg of the project will involve model building which will aim at improving the accuracy of existing models.

19) PROJECT TITLE: Resume ranking using text analytics

ABSTRACT: Recruiting and assigning right candidate for right job consumes significant time and efforts in an organization. Currently candidate selection process involves manual intervention in form of resume screening for keywords from the Job Description and matching the overall requirement while ranking them simultaneously. This process is time consuming and may result in errors while rating the candidates which may lead to higher costs.

This project aims at developing a data-driven resume ranking model that ranks a set of resumes irrespective of the format for a job description based on the criteria set by Management team of the Organization.

The approach will be to extract information on Skills, Professional experience and Academic Background etc. from candidate's resume and create a Term-Document Matrix which will be further used for weighted similarity scoring of resumes. Then, using Machine Learning techniques that involve syntactical and semantical analysis, a similarity check will be performed on the resumes with respect to the job description. Based on these measures of similarity a score for each resume will be determined. These scores and other extracted attributes will be pushed to algorithms like Vector based algorithm like Support Vector Machines and learn to Rank Based Point Wise Algorithm like ORPF and CRR etc.

By automating the process of screening and ranking the resumes, this project aims at increasing the efficiency of Recruitment Team while meeting the requirement of a particular Job Opening. It will also help in reducing errors and bias involved while ranking candidates for interview process.

20) PROJECT TITLE: Agricultural yield prediction and interactive dashboard for visual analytics.

Agriculture is one of the main sectors to be impacted by different sources like climatic changes, soil attributes, seasonal changes, diseases etc. The traditional insurance strategies that were used for the financial risk management are mostly impractical mainly in the developing countries because of the high transaction cost, adverse selection, information asymmetry, poor distribution and many other challenges which hinder the entire process of protection from financial losses. Therefore, estimating the yield of the crop is very critical for the farmers and can reduce the enormous toil faced by farmers including crop selection, watering and harvesting. In the past, this was being performed by considering the farmer's experience on the field and the crop, however, estimating the crop yield by means of statistical techniques has gained popularity in the recent past as it is more accurate and reliable. The project aims at predicting the yield of the crops for sericulture farmers using predictive modelling techniques like linear regression and providing actionable insights to Jayalaxmi Agrotech. This will help the sericulture farmers in selection of right variety of crop and managing agricultural expenses.

In addition to this, we also intend to build various interactive dashboards and extract critical business insights that can help the Jayalaxmi Agrotech in data-driven decision making and robust strategic planning.

21) PROJECT TITLE: Machine learning model for document labeling.

ABSTRACT: Document labelling is a problem of document classification. Classifying a document means identifying a category/label which it belongs to. Every organization deals with documents of some kind. Most of the organizations still label documents manually. It becomes impractical and inefficient when number of documents increase say to millions. Adoption to digital world made it more complex and costlier.

We are doing this project for a large multinational insurance company that receives around 8.1 million communications per year. All the documents are classified manually. There are 30 full time employees dedicated to this job. It costs the company more than Euro 1.5 M every year.

This project tries to address the problem the company is facing today. Objective is to use Natural Language Processing techniques to come up with features and use them to build an analytic model. Plan is to start with the simplest model and move to complex ones until a desired accuracy is met.

The model would take text from all kinds of documents as an input and suggest a label for each document. Manually labelled documents would be used as a corpus to train and test the model first. Once the model is developed, it would be used for the labels where accuracy of the model reaches 95% or above (success criteria).