



CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

DATA GEEK

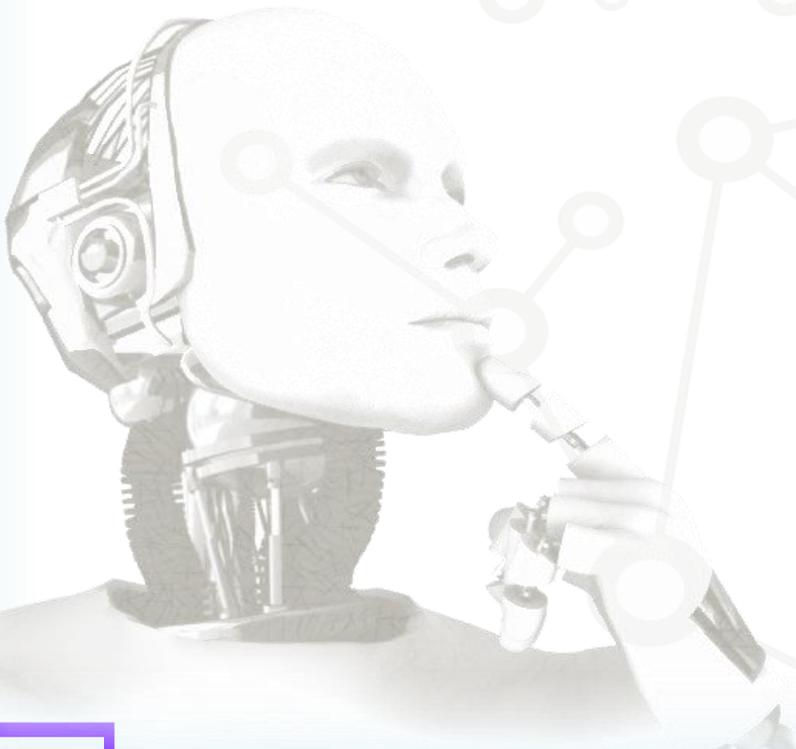
MBA (BUSINESS ANALYTICS)

Volume 1 | Issue 3 | August 2019



CHRIST Institute of Management

INDUSTRY INTERNSHIP INSIGHTS



CONTENTS

1	Introduction	1
2	Development of dashboard for Amazon sponsored products at Selvitate Technologies Pvt Ltd	2
3	Data visualization and analytics using Qlik sense at Greencube Global	4
4	Analysis of online food ordering apps and their growth opportunities in India	6
5	Binary regression model on online order cancellation using machine learning algorithm at Voylla Fashions Pvt Ltd	8
6	Last mile operations visibility at Grofers	11
7	Exploratory Data Analysis and Visualization of Client Data at CIEL HR Services.	13
8	Fraud detection in e-commerce industry at Kilobyte Tech Solutions	15
9	Analysis of customer sentiment towards digital marketing and modelling of topics often discussed	17
10	Implementation of machine learning algorithm in optimization of service request allocation in service desk query optimization at Unisys India Pvt Ltd	19
11	Revenue estimation at Wipro Technologies	22
12	Material requirement planning using SAP at Roots India Pvt Ltd	24
13	Corporate Connect	26
14	Crossword	30

INTRODUCTION

Industry Internships done in the summer months after the completion of the first year is a key requirement to complete the MBA programme. Internships are expected to provide students with an opportunity to apply their classroom learning to a real-life business situation. During the process, students get an exposure to the real time working environment in an organization. Internships provide an opportunity to be in an on-the-job work setting where there exists well defined targets and timelines. Students get a feel of the challenges associated with specific roles which they might have to face in an organization. Internships are expected to improve the critical problem solving and analytical skills prior to taking up jobs in the Industry. Students also develop interpersonal skills which are very much needed to excel in the competitive corporate world.

Students identify and get in touch with reputed organizations in their area of Specialization with clear focus on the learning process. Business Analytics students get to work on specific business problems and find possible solutions by applying various tools & techniques of machine learning. It is essential for a management student to identify problem, apply appropriate methodology and find a solution to the problem with analytical mindset. Business understanding or domain knowledge plays a major role in the problem identification step. Identifying and understanding the variables which have an impact on the problem in-hand involves the next step. Designing the methodology, applying appropriate algorithms to build a model and validating the same is the crucial step which will take the model close to solving the identified problem. Students are evaluated based on their final report and the capability to speak about their project through a viva process. Overall, Internships are a valuable intervention for MBA students while they are in the process of making a successful career path towards the Industry.

Business Analytics Specialization proudly presents the brief of selected Summer Internship Projects taken up by the current batch of students. Needless to say, all of them had a great learning experience.

DEVELOPMENT OF DASHBOARD FOR AMAZON SPONSORED PRODUCTS AT SELVITATE TECHNOLOGIES PVT LTD

Selvitate is a start-up which was founded by Charith G Kashyap and Ruthvic R, and was incorporated on 11th July 2017. It is an e-commerce company which aims to help the micro, small, medium enterprises as well as the offline retailers to sell their products online. The company has expertise in catalog creation, inventory management, product photography, promotions and advertisements, selling the products in multiple marketplaces as well as managing customer feedbacks and thereby provide services to accelerate the business by identifying the right marketplaces to boost the online sales. The company is good at maximizing brand reach through product research and market analysis, boosting e-commerce visibility by creating customer centric content, providing ease of business by complete online e-commerce management.

Selvitate works as an integrated platform working on both a Software as a Service (SaaS) model as well as an operational model to improve sales which differentiates them from their competitors who only provide software to improve sales. The company's target group is cottage industries and small industries, that don't have any digital knowledge. They are also targeting big companies who have problems juggling different platforms for sales. In the long run, Selvitate plans to help take micro and small enterprises to international markets and to bring international brands to India. The company wants to help India gain the largest market share in international exports and earn their reward in the process.

The problem identified in the ecommerce industry is the absence of a platform that provides a structured data that integrates the data from various business reports as well as the time-consuming nature of manual analysis report generation. Again, the limited information available to Amazon Seller Central is yet another problem that limits the scope of analysis within the company.

The project aims to integrate various business reports to develop a standardized data and utilize it to provide business insights through automation and integration of business intelligence tools like Microsoft Power BI, Google Data Studio and Microsoft Excel in a well standardized template format that optimally utilizes the available data and analyzes its performance over a period of time, which in return help the clients to:

- Accelerate the growth of newer or low-exposure ASINs
- Increase discoverability for top Buy Box offers
- Act as an incremental revenue driver

Process Flow



The entire project can be analyzed in five stages as mentioned above wherein each stage of the objectives is fulfilled.

The final outcome of the project is a report generation template that generates analysis reports for multiple clients in a short span as well as a classification algorithm that categorizes the products as critical, important and additional on the basis of performance over a period of three months. The instruction manuals containing the operationalization of the report templates is yet another outcome of the project.

LEARNINGS

1. An understanding of the various processes and the flow of business in the e-commerce industry and techniques applied to power up the data analytics in Selvitate.
2. Learnt how to record and execute macros and utilize it for the standardization of data, work with analytical tools namely Google Data Studio and Microsoft Power BI.



PRATHISH SHOBI M

1827723

DATA VISUALIZATION AND ANALYTICS USING QLIK SENSE AT GREENCUBE GLOBAL

The origin of Business Intelligence goes back to 1958, when the potential of BI was recognized in an article written by IBM Computer Scientist, Hans Peter Luhn, titled “A Business Intelligence System” which described about the automatic system developed to disseminate information to the various sections of any industrial, scientific or government organization

The evolution of Data Warehousing, Online Analytical Processing (OLAP) became the facets of Business Intelligence. This phase of development became business intelligence 1.0. BI 2.0 involved a host of different technologies such as real-time processing, which incorporated information from events as they happened into data warehouses, allowing companies to make decisions based on the most recent information available. While BI 3.0 mainly focused on Business centric where users could leverage the interaction with dashboards on clicks and access from any device.

As we are on the verge of Industrial Revolution 4.0 with emerging trends like Artificial Intelligence and Machine Learning, it’s not enough to provide static output, business needs a means to interact with and explore the results of advanced analytic models. These needs led to the BI.4.0 movement.

Greencube Global Pte is a Management consulting company that provides personalized & customized service and implement solutions for organizations in the field of enterprise management and business intelligence.

This study is made to understand the concept of visualizations using Qlik Sense, a leading data Visualization tool and carry out research for the analytical integration with R to give better insights on prediction and forecasting in visualizations.

The objectives of the project are

- To identify valuable business opportunities from the data to drive profitable decisions
- Use Predictive Analytics to enable putting the right data at the fingertips of the people with the potential to generate meaningful insights

The study has used Cross Industry Standard Process for Data-Mining – CRISP-DM approach for Data Mining. Once the data was extracted, for the visualizations involving advanced analytics, the data is sent to Advance analytics engine. The advanced engine runs calculations and sends

the resulting data back to the Qlik Engine. Lastly, the resulting data is combined with the existing in-memory data and immediately visualized for the user.

For the study, a Retail Apparel store data is used to build a Qlik Sense Visualization application that was used to build Dashboard, analyze Region wise Sales, Product wise Sales, Customers analysis, Time Series Analysis and forecasting. In the application, Clustering and Time Series Analysis and Forecasting is done using R integration. Kmeans algorithm is used for Clustering and Holt-Winters Algorithm is used for Time Series. From the time Series forecast, the business will be able to take better investment and production decisions depending upon the product category which could fetch more sales. From the, it was inferred that there would be a rise in the sales of Children's wear and fall in the sales of Sportswear in the future.

Another Visualization application is built on a Health care data set. The objective was to build an application to analyze the parameters that would lead to heart attack and predict if a patient is likely to get heart attack. The model is having 77.55% accurate and is having 60.31% precision in predicting if a person would have a heart attack. A prediction template was built in a Qlik Sense application for a real time output.

This work has presented an advanced approach in Business Intelligence where not only pre computed questions are visualized, but also the extensive machine learning algorithms are built to provide predictive insights for the Business, which would help in better data driven decisions

LEARNINGS

- The various functionalities of Qlik Sense such as data loading, App design, visualization creation, reuse of visualizations (dimensions and measures) and data storytelling.
- The real time applications of Business Intelligence using Qlik Sense and Advance Analytics Integration in Qlik Sense using R programming.



**NAMRATA
MANGALGI
1827648**

ANALYSIS OF ONLINE FOOD ORDERING APPS AND THEIR GROWTH OPPORTUNITIES IN INDIA

My Internship at Kilobyte Tech Solutions, Chennai started in April, 2019 for two months. The company is a start-up that mainly develops Web and Mobile based applications and planning to expand its services to Industrial Automation, Artificial Intelligence and Analytics. The company planning to launch an online food ordering and delivery aggregator, and the project includes analysing the data of restaurants present in different cities of India.

The main objective was to give insights on the appropriate geographical locations to launch the app, the preferred cuisines in that particular geographical **location**, services to be included to gain customer satisfaction and the factors that affect the restaurant ratings given by customer. Even though online food industry has seen a humongous growth in the last 5 years, aggregators like Swiggy, Zomato and Food panda are not actually making profits. It is required for the later entrants in this industry to track the trends in order to sustain in the business by overcoming the competitor's strategies.

The initial days of Internship started with industry analysis involving online food ordering architecture, trends, competitor strategies and growth factors in India. Later the data of various restaurants containing the details like geographical location, services offered, cuisines served, average price charged, aggregate rating, customer feedback etc. were given. This was followed by analysis to draw useful insights working on quantitative tools like R, Python and MS Excel. Models were built to predict the rating that will be given to the restaurant.

One of the key outcomes of the analysis is, most of the metro cities are occupied with the existing players and it is advisable to launch this service in tier-2 and tier-3 cities in India.

Another interesting outcome is, even though India is known to be a price sensitive market, low prices in restaurants didn't show any positive effect on the ratings. This shows the complete paradigm shift in customers' perception as they are not being very cost sensitive but are only expecting a value for the money from the restaurant. Top 10 popular cuisines preferred by customers in India were also listed.

The internship has been a great learning experience. I was exposed to the working culture of a start-up. As an Intern in Business analytics, I witnessed that data is everywhere and analytics plays a key role in taking business decisions.

The internship has definitely given me a better understanding of my skill sets and the industry expectations. This internship has provided a practical exposure to the theoretical concepts learnt from classroom courses.

LEARNINGS

1. In depth understanding about food industry and importance of analytics in decision making
2. Understanding about classification models, various algorithms and how to choose accurate model



MONICA CHOWDARY

1827042

BINARY REGRESSION MODEL ON ONLINE ORDER CANCELLATION USING MACHINE LEARNING ALGORITHM AT VOYLLA FASHIONS PVT LTD

Industry

The e-commerce industry has transformed the way business is done in India. The Indian e-commerce market is expected to grow to US\$ 200 billion by 2026 from US\$ 38.5 billion as of 2017. Much growth of the industry has been triggered by increasing internet and Smartphone penetration. The ongoing digital transformation in the country is expected to increase India's total internet user base to 829 million by 2021 from 560.01 million as of September 2018. India's internet economy is expected to double from US\$125 billion as of April 2017 to US\$ 250 billion by 2020, majorly backed by ecommerce. India's e-commerce revenue is expected to jump from US\$ 39 billion in 2017 to US\$ 120 billion in 2020, growing at an annual rate of 51 per cent, the highest in the world.

Problem Statement

It is an e-commerce company which delivers jewellery and other accessories for both men and women across the country, but from the past 2 years the company has been facing the issue of large number of orders getting cancelled by the customers before being delivered. The product return rate was 25%. So, the company had to face a huge loss due to the cost involved in packaging the product, dispatching it and then again bringing it back.

So, in order to reduce the Return Rate, a binary logistic Regression model was built to predict whether the given product will be delivered or returned based on certain features, and also predicts the probability of the product being delivered or not.

Objectives

1. To determine what is the probability of getting the product returned by every customer ordering the product, and also whether the product will be delivered.
2. To analyse the various important features that are responsible for returning of the product and also give suitable measures to reduce it.

Project Design and Methodology

Data Understanding

A semi-structured data of more than 1 lakh customers with 47 features was provided. The data of customers ordering product online was provided by the company.

Data Preparation

• Data preprocessing

- **Removing NAs:** The model cannot be Run if there are NA's in the dataset. Therefore, all the NAs in character variables were replaced by creating a new category of NAs, and NAs in numeric variables were replaced by the mean. Deleting the NAs leads to a lot of data loss. So, NAs were replaced.
- **Converting the class of variable:** As per the requirement, the character variables were converted into categorical variables using as.factor.
- **Dividing the data into Train & Test:** For a machine learning model the data has to be divided into training and testing. Training data helps the model in learning and understanding the pattern and Test data is used to check the accuracy of the model.

Modeling

The tool used to build the model is R and algorithm used is **Extreme Gradient Boosting (XGBoost)**. To check the Cut-off Value & prediction ability ROC curve and AUC curve is used.

The model is built using the following steps:

After preprocessing the data:

1. BORUTA algorithm is used to select important variables out of 47 variables.
2. Created Sparse Model Matrix used in XGBoost Model
3. Defined desired parameters Manually
4. Built the XGBoost model after getting Best Iterations
5. Checked accuracy using Prediction Matrix
6. Built Feature importance graph
7. Made ROC curve to check which cut off value is giving higher accuracy
8. Made Area Under Curve which gives us the discriminative power of a model
9. Built a machine learning model on XGBoost which will automatically set the parameters which will give Higher accuracy

Findings

- The company is facing huge loss due to many products getting returned from customers.
- There was a lot of cost involved in packaging, dispatching and then returning it back.
- In case of RTO, the product is usually in good condition and is sent back to inventory.
- In the first case, when the product is received back at retailer's warehouse, a quality check is done and if good, it is sold again. If the product cannot be sent back to vendor for any reason, then it goes for liquidation, where the retailer tries to liquidate it to someone who can buy it generally at a lower cost
- The percentage of RTO mainly depends on the following features Amount Total, Previous RTO percentage, Device Group, State & Tier, Address Length, Gender, Source, TOD,
- ROC curve gives us the Cut-off value at which the accuracy is Max. In this model the cut-off is 4.8 and max accuracy is 72%.
- Area under curve tells us how much our model is able to predict the values, here the AUC is 0.79 that means out of 100 values our model is able to predict 79 values.

LEARNINGS

1. The internship provided an opportunity to learn and develop machine learning models, and to learn how such models could be used to decide for the best suited parameters for optimizing the business processes.
2. It helped in understanding the correlation between various factors and the significance of certain features. Learning of the nature of such factors helps extensively in future decision making.



SACHIN S KATIYAR

1827862

LAST MILE OPERATIONS VISIBILITY AT GROFERS

Grofers is an e-commerce company which initially started off as a hyperlocal delivery in December 2013. It later on changed its business model to become an online grocery provider. It currently delivers in 18 cities around the country and has tech offices in Gurgaon and Bangalore.

The summer internship project was done in last mile logistics. Last mile logistics is the movement of goods from the delivery points to the customer. It is the last leg of order fulfilment. The last mile logistics team provides technical solutions and support for smooth operations. The company is looking at optimizing its last mile logistics operations by reducing the time and cost of order fulfilment, and is hence undertaking several projects to achieve the same.

The different levels of management do not have good visibility on all the activities performed during last mile operations. The company operates in 18 cities, and in order to optimize the activities, it is essential to have a clear view of the current operations.

Operations data is being captured in their various application databases. Each team captures data in different systems. The primary task during the internship was to extract the data from application database and perform data preprocessing on the different data sets, and was finally presented on tableau dashboard. Different levels of management require different levels of visibility. The lower level management will require a more detailed report of all the activities being performed in their stations, and the top-level management requires an overview of operations in the different cities. Different reports need to be generated to provide the views.

SQL queries in Redash application was used to extract the data from the application database and to perform the data pre-processing. Reliability data for reschedules, cancellations, on-time delivery and delivery failed have been arrived at. This data will be used to track reliability of the last mile operations. Further analysis of this data can aid in-time and cost optimization through identification of pain points.

Automatic mailer has been setup to send the reports to different stakeholders providing visibility on the number of orders successfully delivered by field executives. This enables managers

to track the performance of field executives. Another automatic mailer was setup to send a report that gives information on the items settled by GSP's so that the station manager can deduct the pendency from the GSP's salary.

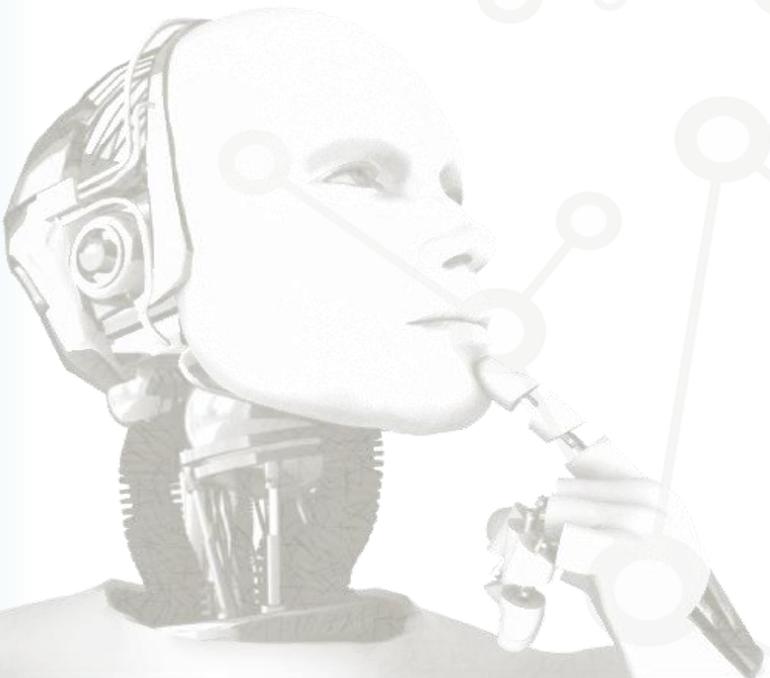
A sentiment analysis of the customer was done so that the company can understand the sentiments of the customers better towards Grofers. The data extracted from twitter, and the tweets were classified as positive or negative. A word cloud was generated to identify what the customers are talking about.

An analysis of the reports and the sentiment analysis was shown, which helped us to infer that reschedules was causing distress to customers. Grofers needs to minimize the number of reschedules from their end. This can be done by improving capacity planning at the stations.

LEARNINGS

Insights about Redash tool used for data visualization.

Understanding the roles of technologies in order to optimize last mile operations.



POOJA KUMAR
PAIKERA
1828149

EXPLORATORY DATA ANALYSIS AND VISUALIZATION OF CLIENT DATA AT CIEL HR SERVICES

CIEL HR Services is into the HR Services industry powered by technology and analytics to deliver full range of Recruiting services from Executive Search, Recruitment Process Outsourcing and Staffing (Permanent as well as Temporary roles) through its offices in India.

In recent years, The Global HR and Recruitment Services industry has grown by 5.3% to reach revenue of \$523 Billion in 2018. HR industry is witnessing a lot of development due to increased adoption of technology in the hiring process. This includes the smart use of people analytics in hiring process and the software systems that assist in making the life of HR team easier.

The cloud deployment is being increased at a rapid pace in the industry.

Problem statement

So, CIEL is trying to compete with the other competitors to expand its client base in this industry. It has 150 clients to work with across various industries and by filling various roles. These clients are the only income source to CIEL. So, the more they analyze clients on how they are performing, the greater the revenues would be for the organization. So, with the past three years' client data, the main problem was to analyze each client's performance over the years, how frequently the company is losing clients and how much revenue can they make with the client in the next few years. Hence, the problem statement is focused across the performance of the clients in the past and the forecasting.

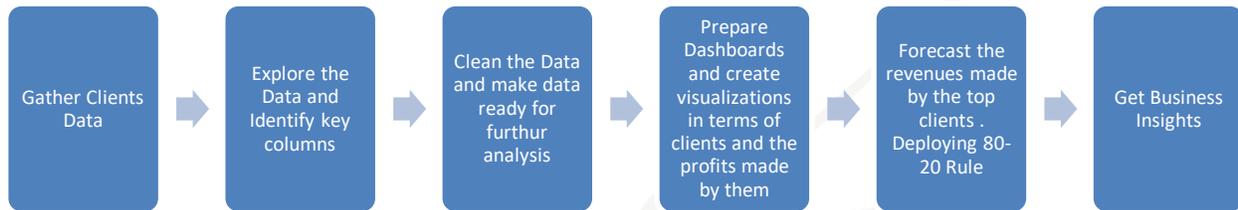
Objectives

The main objectives of study at CIEL HR Services are as follows:

- 1) To understand the pattern of client behaviour and performance, and identify the top clients at CIEL HR Services Pvt. Ltd.
- 2) Enhancing and reporting the web analytics insights using Google Trends and Google Analytics.

Tools used: MS Excel, Google Trends, Google Analytics

The approach in solving the above problem is as follows:



Findings

The client base of CIEL is 150 and the occurrence of the clients from the past 3 years is very low. We can say the company fits for 80-20 rule i.e. 80% of the profits are made by the 20% of clients. So, retaining the valuable clients helps them to stay in profits.

Coming to the website visitors, the company is not getting much clients from the website. Website visibility should be boosted. Company's ranking in terms of visitors is drastically coming down.

Recommendations

- CIEL should concentrate on top 10 valuable clients.
- CIEL can avoid retaining clients from which they aren't making any revenues
- CIEL website's chatbot should be made more interactive by training and allowing it to answer all user queries.
- They should concentrate on website visibility, boost the digital marketing.

LEARNINGS

1. During the internship in CIEL HR Services as a Data Analyst, the main learnings were on how to handle the data, maintain and clean it thoroughly in order to ensure that the end result will be as accurate as possible.
2. Apart from this, it gave an exposure to tools and techniques that helped in expediting various business processes.



SUNIL

1827111

FRAUD DETECTION IN E-COMMERCE INDUSTRY AT KILOBYTE TECH SOLUTIONS

My internship as a Business Analyst at Kilobyte tech solutions, Chennai started on April 2019 for a period of two months. This venture was started by a team of young minds from VIT Chennai, who are passionate about web applications, automation with the use of Artificial Intelligence and Machine Learning. They emphasize on making the task simpler and also help the business to make decisions with better efficiency. They focus on many industries like agricultural sector, mobile applications etc. One such focus is on the E-commerce industry to detect the fraudulent transactions based on the customer's behavior.

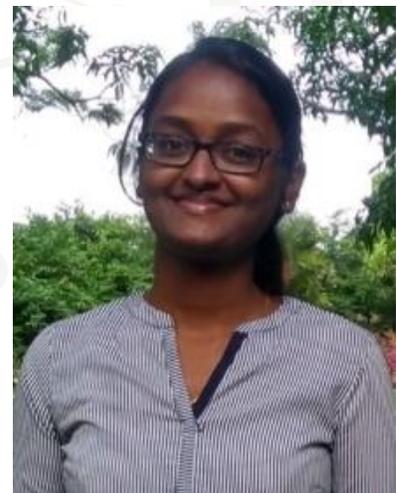
Initial days at Kilobyte started off by analyzing the E-commerce industries like Amazon, Flipkart and Walmart and their approach towards detecting the frauds. The process of data collection and the information about the customers and their preferences was examined. The various challenges faced by the business while a customer places an order, and suggestions on the useful parameters to be taken into consideration while handling frauds were provided. This helped me to gain the knowledge about this particular industry.

Further, I was given a problem statement on how should I approach and derive conclusion to detect frauds. The data had information about the personal details, session details and purchase details of the customers. Almost 60% on my work involved data preprocessing, to make it a structured format to get better insights. The objective of the project was to analyze the parameters that had a significant impact on predicting the customer to be a normal or suspicious or fraudulent.

The data was merged based on the primary key like customer ID, purchase ID and session ID of the customer. The data is cleaned and visualization was done using Python library files like matplotlib, scikit learn and sea born to get insights from the data for better understanding. The feature selection is made using Select K best algorithm and Extra tree classifier method to find the top 20 independent variables. The data was trained by the hyper parameter tuning technique and the model was built on 6 classification algorithms. The accuracy of the models are compared to get a better understanding of the conceptual knowledge. It also helped me to realize the importance of data to make better predictions and solutions. The internship has given me better chances on how to groom myself to meet the industrial expectation in the real world of data science.

LEARNINGS

1. In depth Knowledge about data cleaning and analyzing unstructured dataset
2. Basic understanding about classification models and various algorithms used for the same.



SHWETHA SP

1827048

ANALYSIS OF CUSTOMER SENTIMENT TOWARDS DIGITAL MARKETING AND MODELLING OF TOPICS OFTEN DISCUSSED

India provides an excellent opportunity for education management sector with its population of 500 million in the age group of 5-24 years. In the year 2018, the education sector was estimated at US\$ 91.7 billion, and is expected to reach up to US\$ 10.1 billion in 2019. Skill Velocity, a start-up that belongs to education management industry was founded in early 2019. Skill Velocity offers short, high-intensity hands-on workshops in areas such as Agile, Cloud, DevOps, and IoT & Blockchain. The company reaches its target customers through digital marketing campaign via freshersworld.com.

Problem Statement:

The company is concerned about its customer perception towards its digital marketing campaign. Therefore, identify the most often discussed topics and social sentiment about the campaign.

Objective:

The study aims to analyze customer's perception towards digital marketing campaign of skill Velocity by achieving the following objectives:

- To help the business understand the social sentiment towards their digital marketing campaign.
- To discover the topics that run through the customer reviews by analysing the words of the texts.

Approach and methodology to solve the problem:

Market research was conducted for duration of a month and data was collected through a questionnaire. Questionnaire includes questions that dig into options that reveals the customer's perception towards Skill Velocity's digital marketing campaign. Cross Industry standard process for data mining (CRISP-DM) methodology was followed for executing the project in a transparent and an organized manner. Tools used were Excel, R and Python.

Beginning with understanding the nature of business and data (i.e. text responses), data preparation was carried out by verifying the quality of data, removing stop words, converting the cases into 'bag of words', building document term matrix and finally exploring the data through

visualization by constructing word cloud. Sentiment analysis using Afinn lexicon and topic modeling by Latent Dirichlet Allocation (LDA) were the modeling techniques applied. However, the business objective of this project did not require deploying any model, hence the project was completed at the stage of evaluation.

Findings:

Sentiment Analysis results indicated that from a total of 300 responses, only 29 responses (approx. 10%) were classified with negative sentiment. Three most often discussed topics were extracted through topic modeling and by mapping the negative sentiment responses to the extracted topics, the findings are as follows:

- Topic 1: Customers have talked about jobs and workshops.

There were questions raised on what are the companies that would offer job at skill velocity

- Topic 2: Customers have talked about attractiveness of the poster.

Few customers have questioned the credibility of the company as no website details were mentioned in the poster.

- Topic 3: Customers have talked about information and usefulness to student Community with regard to workshops and jobs.

Customers would prefer to know what skills the workshop is offering to benefit the students/fresher's and also mention the same in their promotional poster.

As per the observations recoded, implementable suggestions were provided for improving the company's digital marketing campaign in its effectiveness to attract the target customers.

LEARNINGS

1. Understanding the basic concepts of python and powerful libraries used in text analytics.
2. In depth understanding about various statistical techniques used in text analytics such as Naïve Bayes classification.



SHYALAJA S

1827956

IMPLEMENTATION OF MACHINE LEARNING ALGORITHM IN OPTIMIZATION OF SERVICE REQUEST ALLOCATION IN SERVICE DESK QUERY OPTIMIZATION AT UNISYS

Industry profile

The Information Technology & Information Technology Enabled Services (IT-ITeS) sector is a field which is undergoing rapid evolution and is changing the shape of Indian business standards. This sector includes software development, consultancies, software management, online services and business process outsourcing (BPO).

Market size of IT industry:

India's IT & ITeS industry grew to US\$ 181 billion in 2018-19. Exports from the industry increased to US\$ 137 billion in FY19 while domestic revenues (including hardware) advanced to US\$ 44 billion. According to the NASSCOM's report the year 2018 is considered to be the year of growth because of the new technology which is changing the businesses. The image below provides information of IT industry's contribution towards global economy, Indian economy and global tech spending.

Problem Statement

Service desk in Unisys handles all the queries related to hardware, software and network issues which are received from across the globe. They are required to resolve the issue quickly so as to not cause any work delay that affects the day to day business operations. The current situation is such that when the query is been raised it is redirected to the team in general which is time-consuming.

Objectives

- To analyse the various service request received from the customers across the globe.
- To categorize the service requests and to optimally allocate the service request queries to the appropriate support team.

Approach and Methodology to solve the problem

Methodology followed to solve the problem is CRISP-DM. CRISP-DM stands for Cross Industry Standard Process for Data Mining, this model describes the common approaches used by the data mining experts and it is also the most widely used analytics model.

Business Understanding

Unisys is a service-based company which will provide service-centric solutions to business and governments. When the service is been provided and if there are any queries which have to be resolved the customers will contact the service desk department for their query to be addressed. The service desk department will receive customer queries from various parts of the globe and will receive nearly three thousand queries from Bangalore alone every month. Wherein each query is been addressed within 30 minutes. This is a very critical process, where the query has to be directed to the appropriate support team to resolve the query. The time taken to resolve the query is been optimized by allocating the queries to the correct department.

The project deals with developing a model which will automatically categorize the service desk queries to its appropriate categories.

Data Understanding & preparation

Six months of data from October to March has been collected which consists of eleven attributes and 46,394 records. Out of these eleven attributes, two attributes – Short description and category were taken into consideration due to their more insights, information and value addition to the model building. The main issue with the data was that the information tickets were not categorized to their right categories and hence the query was rightly categorized to its categories.

Modeling

The tool used to build a model is Microsoft Azure. Microsoft Azure machine learning studio has a specific experiment workflow, this workflow is been followed to build a classification model. The data was imported and the text pre-processing was done such as removal of stop words, lemmatization, detect sentences, normalize the cases to lower case, remove numbers, remove special characters, remove an email address. Feature hashing is the next step which is used to convert variable-length text to equal length numeric feature vector. The data was split in the ratio of 80-20 which is used for training and testing of the data model. The Algorithm chosen was multiclass neural network.

Findings

With respect to the analysis of the service desk queries, some of the findings are:

- Most of the queries which the service desk received were under the category Application, Hardware and Network related query.
- Most queried subcategory was accessed connectivity and password reset.

With respect to the categorization of the queries, the findings are:

- The output obtained from the model are: The overall accuracy of the model is 81.32% and average accuracy is 96.26%, which implies that the model predicts 0.8132 times correctly the right category for the query described.
- With the help of this model, the time taken to find the right category for the service desk query is reduced. Earlier the time taken to resolve the query was 30 minutes with the implementation of this model the time taken to resolve the query will be reduced to 20 minutes or less, which in turn will contribute to the efficiency of the business i.e.33.33%.

LEARNINGS

- Learned most of the aspects of R programming for text analytics such as text mining, tm (text mining) package and word cloud.
- Learned how to work on Microsoft Azure Machine Learning Studio wherein the model was deployed.



KAVYA S

1827943

REVENUE ESTIMATION AT WIPRO TECHNOLOGIES

The IT-BPM sector in India expanded at a CAGR of 11.14 per cent to US\$ 155 billion in FY17 from US\$ 74 billion in FY10. The size of the industry will grow to US\$ 350 billion by 2025. India is a prominent sourcing destination across the world, accounting for approximately 55 per cent market share of the US\$ 173-178 billion global services sourcing business in 2016-17.

WIPRO Technologies being in trend, have introduced consulting services which has been in incubation for the last 5 years and is growing rapidly because of its industry growth. Now that the team is handling seven Business Units and seven Service lines, for which they have to outsource the services for the two project types they have: 1. Fixed Price Projects 2. Time and Material projects. The team has to claim their revenue which was generated by their sources. They have been undergoing a process of accepting the revenue which is estimated by the core team (Bu's and SL's) for every project, in terms of percentage, which is leading to changes in the revenue yield. So, now the team requires a forecast model to cross check the estimates which is given by the core team and to prepare a dash board which would give the valuable customers for consulting team

Revenue Estimation is a process of preparing and understanding the flow of business in the four quarter, where any business could prepare for the expenses, unexpected hindrances, and forecast the same with accuracy becomes highly important. To do the same for the consulting team, the collection process and bucketing them on the basis of business unit and service line was done and the models were designed:

1. Project type wise - Fixed Price and Time & Material wise the billable man months, the rates at which they are being charged for the onsite and offshore work, the allocation percentage for revenue generated by Time and Material was estimated. BMM was estimated using FORECAST functions, rates for FPP was of previous quarter, For T&M rate was average of previous quarters and the Final calculation was $\text{sum}((\text{BMM} * \text{rate}) = \text{revenue})$ estimated.
2. Grade wise – grade wise revenue generated was considered and that the estimates for the next quarter was calculated using BMM by applying the rate of change in the BMM for previous quarters and Rates also with rate of change in the rate in previous quarters. Allocation percentage in the revenue calculated for T&M wise estimates

- ✓ The model was automated with 90% accuracy (tested on previous quarters) using excel as recommended by the team. the model with project type revenue generation was recommended strongly because of its accuracy. Dash board was created to understand the valuable customers with slicing and dicing of the customers according to the business unit and Service line with top 10% revenue generated to make sure that men can be deployed without any delay. Dashboard also shows the revenue generated for region wise, revenue wise, BU wise, SL wise etc.

Recommendations such as

- ✓ All the business units and service lines must be consulted every time to know the additions and deletions in the process in order to know the pass through, removal of project
- ✓ This estimation process gives 90% accuracy and should be deployed for all other departments in order to make the estimation a little more convenient.
- ✓ The data collected must be in a granular level so that the accuracy increases and this can be achieved by using grade wise
- ✓ As we can see the Business grow, deployment must be ready in hand to avoid turnover ratio were given.

LEARNINGS

1. The working of business from SBU to all other domains and how well business line should reciprocate with service line in IT-BPM sector.
2. How the pricing strategy and revenue forecasting are planned and implemented.



**HARITHA
PONNALAGI R
1827742**

MATERIAL REQUIREMENT PLANNING USING SAP AT ROOTS INDIA PVT LTD

Industrial manufacturing is a major growth sector for the Indian economy with diverse companies including those engaged in manufacturing of machinery and equipment, electrical and metal products etc. The project focuses on shifting their material resource planning completely on SAP.

In a manufacturing operation, developing a plan for resources within the operation is of utmost significance. Resource planning enables to manage various areas of supply chain much more effectively, such as inventory, production, and output. Without raw materials being transferred throughout the supply chain adequately, will lead to complications within production such as being unable to fulfill customer orders or either having potential inventory issues. Since the MRP in SAP is not accurate and the forecasting of MRP of 2020 is also done.

Material requirement planning can be improved by reducing the rejection percentage of each component by identifying two important reasons the components get rejected by using root cause analysis, and thereby improving the efficiency and reduce wastage. After the analysis it was found that blow holes and dimensional problem is the factor that causes the rejections of the most of the components. These defects occur when an excessive evolved gas is not able to flow through the mould. The rejection percentage for 6 months and the factors of each individual components. The comparison of the rejection percentage of 6 and 3 months was done. It is observed that the rejection that happens in the greater run is more than the shorter run.

To calculate and analyze the lead time of all the operations of each component, and to identify the operation that has the highest lead time and reduce it. With the analysis of lead time it was found that the operation of die casting has the highest lead time. It is because the operation of die casting takes more time, certain components lead time is increasing drastically. It was found that pre-machining and machining takes the most time, that is non value-added time frame. The total transportation timing taken for out sourced products. Total time of 7 hours is taken for each of the out sourced process.

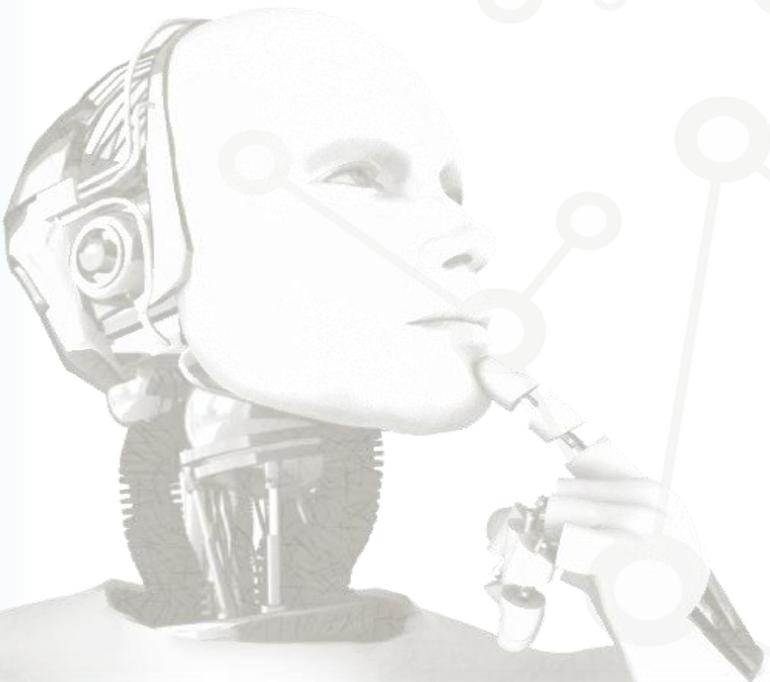
MOQ is calculated by using lead time and rolling stock for material requirement planning. Compare and analyze the FG (Finished Goods) plan VS actual production to identify the gap and forecast the production. Analyzing the data, it was found that the dates of both the Purchase

order and documentation were different from the invoice and the SAP data for about 10% of the data. Therefore, the accuracy of the lead time is not accurate. With the collected data it was found that the suppliers send the stocks randomly based on their production. The reason for the changes in the lead time of the components is because the PO orders are not returned back in the original PO numbers.

The accuracy of the previously available MRP is 70%, and the accuracy of the calculated MRP is 90%, an improvement of 20% in the MRP has been achieved. The recommendations given are Complete MRP using SAP, more concentration towards factors of rejection so that the cost of re-work can be reduced. Since PO dates and 101 dates are entered manually to SAP, the lead time is inaccurate. Since the FG plan is completely based on the data form SAP, there is a drastic difference between the schedule and dispatch.

LEARNINGS

1. How to implement material requirement planning in SAP, so that forecasting of production is done with more accuracy.
2. The concepts of operational management been implemented in the company for more efficient product line.



MADHUMITHA S

1827546

WHY VOICE IS THE NEXT DISRUPTION IN COMPUTING?

INTERVIEW WITH SOHAN MAHESHWAR, ALEXA EVANGELIST

Sohan Maheshwar is a developer, and is the Alexa Evangelist at Amazon. He has more than 8 years of experience in the Information Technology and Services Industry. He has expertise in technical support, mobile application development, developer relations and technical writing.

During the interview, he shared his insights on, “Why voice is the next disruption in computing” making the interview an overwhelming interaction. The discussion also had another point in view: Alexa, its present state and the fast-evolving phase of it. It helped in understanding the importance of subtle nuances when it comes to complexity.

Following is the excerpt from the interview:

1. What utility does voice offer when compared to visuals?

Essentially, Alexa is a cloud service that powers a bunch of devices that let your voice control your work. As humans, speaking comes naturally to us, and so far, we have interacted with tech through screen-based devices. But now we are at this stage. Thanks to technology. It is actually possible to do this. Hence, we came up with Alexa, with which you can do anything, from getting cricket scores, to music, to control the lights, to order pizza, to booking cabs, and all of it just via your voice. Its advantage is the ease and intuitiveness which we have never seen before. Right now, something as simple as adding a reminder on your phone, one has to look for the phone and type it out, save, etc. But here, one can just say something like “Alexa, remind me to buy groceries”. Something as simple as that. And that’s the same way in booking a cab. For instance, “Book me a cab for work”, and its done. So that ease of use where one is not tethered to mobile phone or laptop and they can just speak it out is what I think is the intuitiveness which we had wished for.



2. Typing can be specific at one go. But, technology like Alexa which is a back and forth interaction can be time consuming. Don't you think this is a disadvantage?

It actually depends on the use case. For example, in search, one needs to open a laptop and type it out, where there might be errors and then hit enter. Instead you can open Alexa and can say something like “Hey Alexa, how many centuries did Virat Kohli make?” which makes it easy. So, it is a lot of things like ordering food when you are a picky eater. May be people say that screens are better, and I agree to it that screens are better. But say if you have to reorder a thing, say “Alexa, I want to repeat my last order”. So, in such cases you can go with Alexa.

3. In what way and how fast do you think Alexa's technology will impact the businesses, and which industries will it affect?

It is actually interesting. Last year we have launched something and is right now available only in the US. It is called ‘Alexa for Business’. It is going to revolutionize the work, where you go to a meeting room, you have to set up the TV, dial people and stuff. Right now, there are offices in the US where they use Alexa in the office room. I walk inside the room and say “Alexa, start the meeting”, and it figures out that right now it is 3 pm and dials by itself. All of it happens automatically. In future, it might even be able to take meeting notes and give you minutes of the meeting. And this is just one-way Alexa enables all of this value chain. From a fresher joining, who wants an employee orientation, to filing a ticket by asking Alexa at the desk, to the CEO who wants a daily sales report. Instead of opening the system and drawing three tableau reports which takes around 10 minutes, one can just say “What are my top sales today?” or “What are the sales in quarter 4 looking like?”. I think, throughout the entire employee organization there is some sort of value added.

4. What are some of the security aspects that one has to consider while talking about such a technology?

We take security and privacy very seriously. Anyone who builds a skill for something specific, at no point will anyone get the transcript of what the user has sent. This is something we take very seriously, and not going to change. To be a bit technical, there is something called intents and slot values. Intent is an action which a user wants to do. So,

the user says “What’s the weather?”, and the intent is to get the weather, and the slot value is any variable within that sentence. So, I say “Get me the weather in New York” or “Get me the weather in Bangalore”. So, the intent is the action and the slot which is either Bangalore or New York. You will never get the full sentence of what the user is saying. That’s just one step. The other is, all these devices have something called as wake word, and Alexa won’t process anything until you say the word ‘Alexa’ which is a wake word. Alexa understands that, and now is when it processes it. Without this, it might record everything in the background, which is not what we are doing.

5. How does Alexa understand that two sentences mean the same even when all of the words are different in both the sentences?

Essentially, in any conversational interface, there are different ways of saying things. There is something called as Natural Language Processing in technology which is a subset of AI, which is how computer would understand human speech. So, human speech is largely unstructured, whereas computers can understand structured data. And, by structured I mean array and such things. So, what NLP does is, it takes the sentence that the human says and converts it to something structured which a computer can understand. And that structure is essentially what an intent is and what the slot values are. So, for any chatbot or conversational interface, this is the transaction that is happening. A human sentence is taken and there is an algorithm for NLP which pulls out what the intent is and the slot values are. It creates and builds a structure and feeds that into a program which can then do something.

6. How important is it to understand the emotion in a sentence?

So, that’s a tricky part because it is very important and that is what distinguishes a human from a computer or laptop for instance. It is important as we read news reports recently, about Alexa being able to understand emotions. There is something in computers called as sentiment analysis which sort of sees how the words you are using the way you are using, and can recognize the sentiment. It will give you a percentage. So, it will tell you it is 100% positive or 50% positive. And this is something we are working on. It is also very difficult because two different people can express the same emotions slightly in a different way, like how they can say the same thing, they can also express the emotions differently. One being angry versus someone else being angry is going to be a different sort of interaction. It is

7. Which stage according to you is the most difficult stage in the development of the voice component in Alexa?

I don't know which is most complex, but I can tell which one is the most important, which is the voice designing phase. This is because the technology aspect is something that we are good at, and we can all figure it out. But unfortunately, people don't pay much attention to the voice designing stage. Designing conversation is something that is so different from what we have been doing till now, which is designing websites and apps. Those are screen-based devices which are visually good for the eye. So, I think you should spend more time on this, and this in turn leads to good skills and the best skills come out.

8. Can you give an advice for students who are aspiring to reach great heights in the field of business analytics?

At present, technology and analytics are something in the process of becoming big. And the playing field is completely open. Sitting in India, you can build a system for your client in US. Thanks to technology. So, there is tremendous scope here, and there is always a first mover advantage, and figuring out the right use case will really be good in the future. So, that's my advice to you.

9. As you are working in a field where creativity is of prime importance, what priority must we give for it in the field of analytics?

I am not an analytics expert, but I feel there is always room for creativity. That is what differentiates the insights in business analytics. There are some creative ways of getting data or what you do with the data. Now, access to data is not the problem, it is about what you do with it. And, that's where creativity plays a big role.

INTERVIEWERS

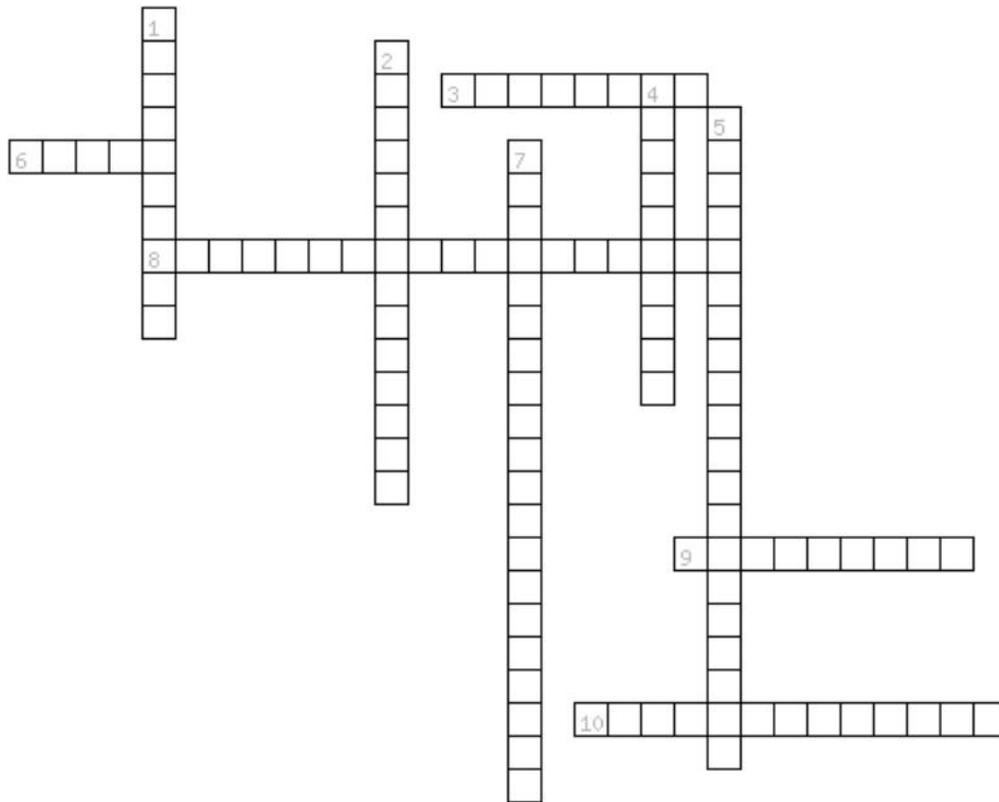


ALLENA CHACKO K
1828034



AISHWARIA PAUL
1827834

CROSSWORD



Down

1. Distributed, decentralized, public ledger maintained across several computers that are linked in a peer-to-peer network.
2. Interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos.
4. Process of exploring and analyzing large amounts of unstructured text data aided by software that can identify key attributes in the data.
5. Simple-to-use BI tool that enables professionals to extract data generate reports and run queries with no or nominal IT support, using various tools and software.
7. Process of associating each pixel of an image with a class label.

Across

3. Unstructured, untagged and untapped data that is found in data repositories and has not been analyzed or processed.
6. Open-source distributed general-purpose in memory cluster-computing framework based on Hadoop MapReduce
8. Use of statistical and linguistic(NLP) technologies to improve data management performance, from data analysis to data sharing and business intelligence.
9. Graphical user interface which often provides at-a-glance views of key performance indicators (KPIs) relevant to a particular objective or business process.
10. A distributed computing paradigm which brings computation and data storage closer to the location where it is needed, to improve response times and save bandwidth.

ANSWERS

1. **BLOCKCHAIN**
2. **COMPUTER VISION**
3. **DARK DATA**
4. **TEXT MINING**
5. **SELF SERVICE ANALYTICS**
6. **SPARK**
7. **SEMANTIC SEGMENTATION**
8. **AUGMENTED ANALYTICS**
9. **DASHBOARD**
10. **EDGE COMPUTING**



DATA GEEK CREW



ARCHANA SCARIA
1827431



GAUTAM DEKA
1827208



P JEYALAKSHMI
1827442

DESIGN & CREATIVITY



HARITHA T
1827236



Dr LAKSHMI SHANKAR IYER
Head of Specialization
Business Analytics

EDITOR

“How you gather, manage and use information will determine whether you win or lose.”

- Bill Gates

