REVA UNIVERSITY
Bengaluru, India

A Project Report on

Reducing Business Compliance Risk with help of Analytics

Submitted in Partial Fulfilment for Award of Degree of

Master of Business Administration
In Business Analytics

Submitted By
Lalit Aggarwal
R19DM003

Under the Guidance of
Phaneendra Akula
Senior Manager of Data Science, Sunrise Ltd.

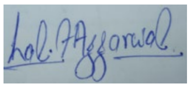
REVA Academy for Corporate Excellence - RACE
REVA University
Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064
race.reva.edu.in

August, 2022

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, **Lalit Aggarwal** hereby declare that I have completed the project work towards the first year of **Master of Business Administration in Business Analytics**at REVA University on the topic entitled **"Reducing Business Compliance Risk with help of Analytics"** under the supervision of **Phaneendra Akula**, **Senior Manager of Data Science, Sunrise Ltd.** This report embodies the original work done by me in partial fulfilment of the requirements for the award of a degree for the academic year **2022**.



Place: Bengaluru

Date: 20th August 2022

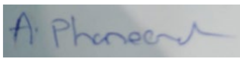
Name of the Student:**Lalit Aggarwal**

Signature of Student

REVA UNIVERSITY
Bengaluru, India

Certificate

This is to certify that the project work entitled "**Reducing Business Compliance Risk with help of Analytics**" carried out by Lalit Aggarwal with SRN R19DM003, is a bonafide student of REVA University, and is submitting the project report in fulfilment for the award of Post Graduate Diploma in Business Analytics during the academic year 2022. The Project report has been tested for plagiarism and has passed the plagiarism test with a similarity score of less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said Degree.



Signature of the Guide
Name of the Guide: **Phaneendra Akula**
Guide

Signature of the Director
Name of the Director: **Dr. Shinu Abhi**
Director

External Viva
Names of the Examiners

1. Harsh Vardhan, Chief Digital Technology Architect, Capgemini
2. Pradeepta Mishra, Director – AI, L&T InfoTech
3. Rajib Bhattacharya, Director – Data & Analytics, Cargill

Place: Bengaluru
Date: 20^{th} August 2022

Bengaluru, India

Acknowledgement

Among many people who helped my research journey on virtual try-on to be a success, without whom I would not have managed to complete this research nor would I have accomplished my MBA.

I would like to acknowledge the support provided by the founder and Hon'ble Chancellor, **Dr. P Shayma Raju**, Pro-Vice Chancellor, **Dr.Dhanamjaya Madiraju,** Registrar, **Dr. N. Ramesh.**

My special thanks go to **Dr. Shinu Abhi, Director of REVA Academy of Corporate Excellence** for believing in me and for providing me with all the guidance and support throughout my MBA program.

I am grateful to my mentor **Phaneendra Akula, Senior Manager of Data Science, Sunrise Ltd** for all his help and guidance in the project and for showing me a path I could walk.

The team of mentors at REVA Academy of Corporate Excellence, including **Dr. Bharatheesh Jaysimha, Ratnakar Pandey, Mithun D.J. a**nd others, taught us some awesome methodologies throughout the academic year. I would like to extend my gratitude to the educational department at Reva Academy of Corporate Excellence, REVA University campus, for providing ample opportunities and guidance.

Throughout the project work, I was grateful for the support, suggestions, and friendly advice offered by my classmates.

Lastly, I want to thank my family for supporting me through this research process and for believing in me.

Place: Bengaluru

Date: 20th August 2022

REVA UNIVERSITY
Bengaluru, India

Similarity Index Report

This is to certify that this project report titled **"Reducing Business Compliance Risk with help of Analytics"** was scanned for similarity detection. Process and outcome are given below.

Software Used: **Turnitn**
Date of Report Generation: **19-Aug-2022**
Similarity Index in %: **8**
Total word count: **6860**
Name of the Guide: **Mr. Phaneendra Akula**

Place: Bengaluru
Date: 20th August 2022

Name of the Student: **Lalit Aggarwal**
Signature of Student

Verified by: Dincy Dechamma

Signature
Dr. Shinu Abhi,
Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	LR	Logistic Regression
2	KNN	K-Nearest Neighbours
3	NB	Naive Bayes
4	RF	Random Forest
5	GB	Gradient Boosting
6	FCA	Financial Conduct Authority
7	MLPNN	Multi-Layer Perceptron Neural Network.
8	CHAID	Chi-square Automatic Interaction Detector
9	CART	Classification And Regression Trees

List of Figures

List of Tables

Abstract

In the present time, compliance risk assessment is very important for the organization as well as its reputation in the market, Compliance risk is broadly defined as any action made by a financial organisation or individual that has a negative impact on clients, the stability of the market, or the ability of rival businesses to compete successfully. Failure to control compliance risk can result in regulatory action, penalties, and reputational loss that can hurt a company for years after the event. Due to compliance-related regulatory measures, there has been a huge financial impact on businesses, and it all started with an individual's conduct. Because compliance risk violations are of great public interest, it is crucial to adopt a comprehensive approach for a successful defence.

The objective of this study is to identify the employees who are deviating from organization compliance policy. So the organization reduces the financial and reputational losses that could happen due to violation of market standards set by the Financial Conduct Authority (FCA). If the organizations find any malpractice happening in the business they can take corrective action to stop or minimize them in the early stages.

Different machine learning techniques have been used on the given client data and found that the majority ofaround 60% of employees are on medium risk, whereas it has been predicted that around 17% of employees could do malpractices or not follow compliance rules, on other hands 23% of employees are diligently following the organizational policies and they are on very low risk. Recommended better pieces of trainingare also given, so they can make aware their staff for the consequences of any misconduct or malpractice.

In this project, the assessment of Compliance risk of Employees with the help of modern machine learning techniques has been classified into three categories high, medium, and low based on different features. So organizations can take some preventive measures to improve their customer handling, in that way Organization may prevent any tentative loss in the future.

***Keywords: Compliance, Logistic Regression, K- Nearest Neighbour, Random Forest, Gradient Boosting,* Naïve Bayes.**

Contents

Chapter 1: Introduction

Since the financial crisis of 2007–2009, the financial services sector has gone through considerable changes. The regulation discussion has centred on preventing unfair behaviour among other things at the national, international, and regional levels. (Resti, 2017) this led to an examination of the business techniques used by financial organisations. Regulators and policymakers have brought up the topic of dealing with "securitization abuse," which has seriously eroded trust in the financial sector, in this context. After that, other cases of misconduct were known: the incorrect sale of budget items and the control of LIBOR. While the former identifies with the distorted use of trade practices, the latter is concerned with the deceptive use of the global reference cost of credit.

The sales practices of financial companies are now in the spotlight. In particular, the mis-selling of payment protection insurance items in the UK and the Wells Fargo account misrepresentation case in the US were more reflective of the crippled use of strategically offered work to include offering various items to existing clients,(Mehrotra, 2016).

The broad use of the term "compliance risk" by financial services companies, and banks, is to describe the risks associated with the way organizations and their employees harm customers or negatively affect market stability. Compliance risk means the risk that a firm's behavior will result in bad outcomes for customers. Recent developments have shown how crucial it is to stay focused in both the direct retail and discount marketplaces. Organizations that fail to balance potential clients' risks may face administrative action, fines, and reputational damage, all of which can hurt the company's bottom line long after the position has ended. Regulators also focused on other compliance risks, including compliance with financial sanctions, client capital management, and mortgage availability.

The basic goal of banking supervision is to encourage the ongoing fair and open provision of financial services to individuals and enterprises. To accomplish this goal, supervisors conduct both individual and group examinations of financial institutions to aid in boosting their resilience to a variety of potential hazards, including credit, market, liquidity, and operational risk. Employee misbehavior, or the potential for compliance or commercial operations that are unlawful, unethical, or in opposition to the company's ideals, values, policies, and procedures, is one risk that stands out.

Employee misconduct has repercussions that go beyond the individual and may have a more profound influence on the company as a whole, the economy, and the financial sector. Employee misbehavior can reduce a company's adaptability by, among other things, diverting the board's attention, obstructing business, causing changes in personnel arrangements, and depleting the company's money. In the larger economic and monetary business sectors, an insult can directly hurt buyers and agents. After some time, market members may lose confidence in the budget area in general and antagonistically affect its core part in monetary intermediation.

For instance, Gallup discovered that trust in the financial industry has decreased by 50% over the past ten years. (Astuti, 2021) "The prevalence of wrongdoing in the financial industry has increased to a level that has the potential to cause systemic concerns by weakening confidence in financial institutions and markets," recently stated Mark Carney, the governor of the Bank of England. These occurrences may lead to conflict and raise intermediation costs, which would restrict the flow of financial services. Such externalities, which spill over to different firms, shoppers, and organizations, are the fundamental reason for authentic area interventions identified with worker misconduct.

Supervisors typically manage the risk of misconduct by monitoring the efficiency of internal control procedures including compliance and auditing as well as corporate risk management. This is unquestionably not a new issue; issues with vicarious harm power, for instance, have existed since the Banking Act of 1933. The Federal Reserve and the OCC were given the authority to fire bank officers and directors as a result of that statute. and forbid them from engaging in risky and harmful banking practises inside the sector (FSB, 2016).

But regardless of the guidelines, the administrative center, and the firm's efforts, the risk of wrongdoing continues. For instance, high-profile instances of misbehavior involving reference rates, foreign exchange trading, and retail banking sales practices have occurred recently (HEAKAL, 1933) . Large corporations have paid fines totaling $320 billion for representative bad behavior worldwide. (Tracy, 2017).

What justifies this tenacity? One theory holds that the current issues are simply a string of eccentric occurrences, or random selections of "bad apples". Another is that certain companies have operational flaws, or "poor processes," which allow for these results. However, root cause analyses of numerous recent incidents of misconduct in the financial sector point to the fact that misconduct is often the outcome of widespread organisational breakdowns rather than just a few bad actors or ineffective procedures. Many workers and managers took part in the misconduct, supported it, or chose to ignore the troubling behaviour. This suggests a different issue.

Despite the fact that there are probably numerous factors that contribute to this type of broad breakdown, academic writing is beginning to highlight the hierarchical culture of society as a significant influence on behaviour and the ensuing risk of wrongdoing. According to this model, culture is a collection of attitudes, convictions, customs, and values that either reduce or raise the risk of compliance. Although it may be challenging to define these social elements ex-bet, the ex-post influence can be clearly shown in subsequent actions and results.

Firms and their boards have the primary responsibility for improving corporate culture and reducing the risk of misconduct, but people and businesses are likely to invest less in risk mitigation than is socially desirable due to the possibility of externalities to various aspects of the financial system. The risk of wrongdoing endangers the primary goal of supervision, which is to ensure the effective delivery of financial services to the economy, from both a prudential and financial stability standpoint. It implies that managers make a commitment to supporting ethical internal practises that reduce the likelihood of misconduct and improve the culture.

There area lot of incidents where an organization faced financial and reputational losses due to their employee misconduct and not adhering to the company's compliance policies. In the present scenario, the use of modern machine learning techniques can analyze past data and predict the probabilities of their future occurrence. In this way, the employer can analyze employee past data and can predict whether the employee is adhering to organization compliance policies, if not organization can take timely corrective actions.

Because it is a crucial component of the occupational health and safety management strategy, risk assessment is very significant. They help:

- Sensitize people to dangers and risks.
- Determine who might be in danger. (eg staff, cleaners, visitors, suppliers, public, etc.).
- Establish whether a control scheme is necessary for a specific hazard.
- Assess if the current controls are effective or whether more needs to be done.
- Avert illness or injury, particularly if this is done during the design or planning stages.
- Prioritize the risks and preventative measures.
- Obey all necessary legal requirements.

Chapter 2: Literature Review

The paper presented by Francesco (Francesco., 2018) underlined the misdeed stake in the cross-selling in the banking sector. A series of malpractice incidents in the financial services sector have suggested that there was a potential of fraud through cross-selling. In this report, Francesco highlighted the regulatory key to the malformed use of financial practices. The author also highlighted the risk of misconduct that can pose systematic threats to the community's financial strength. To stop the wrongdoing in cross-selling, the author prefers intervention in the conduct, culture, and governance framework of financial institutions. It is an inner issue of the community, so it should be addressed from the interior.

In the paper Hennessy,(Hennessy et, 2017)studies the wrongdoing in financial institutions and the role of banking rules and managers to mitigate risk by analyzing risk administration, interior controls, and governance. For instance, mischief can prevent this fundamental intermediation activity by changing the manager's evaluation, shattering a company's reputation, causing a change in the workforce of the community, depleting its capital, and making a firm less resilient. Again, wrongdoing can break the budgetary space by reducing faith and belief in ways that inhibit intermediation. By removing from the developing literature about the primary causes of wrongdoing and the supporting justifications of a strong culture in financial businesses, managers can help average slight risk and make the financial drive more dropped.

Their report(Choi, 2019)studies the impact of dishonest reports against an institution on their workers. If dishonest financial reporting happens against any organization, it may affect Employees' wages and their turnover before and after the incident. They have discovered that, on average, employee wages fell by about nine percent and the attrition rate increased up to twelve percent. Organization employment transition also goes down after the happening got reported. In their paper,they pointed out whistle-blower reporting to outside entities rather than utilizing internal channels for reporting, about the misconduct occurring in the organization. When they feel that they can be retaliated or their complaint will not be handled properly if they report it to the internal channel. Their article largely focuses on these types of problems and mechanisms that should be followed by corporations so it reduces the chances that an employee use outside entities rather than inside one to report the matter against their organizations.

Their paper, (Namasivayam, 2006)discussed linking between employee misconduct and consumer satisfaction. Employee misconduct decreases buyers' impression of control and their fulfilment with a trade. To improve purchaser fulfilment, examining likely reasons for misconduct is critical. Their study gives arguments against the many linkages that are suggested and emphasises the contribution that hierarchical work makes to remembering and strengthening for enhancing customer fulfilment.

The study presented by Mark Egan(Mark Egan, 2019)on financial Adviser misconduct and their associates. As per them around seven percent of advisers are usually involved in misconduct and they do it repeatedly, it also depends on the organization size in some of the bigger firms it was noticed that up to fifteen percent of counsel do misconduct. It's a common practice when some firms retrench them due to compliance some other firms rehire them. So they continue to misconduct consumers.

In the paper Young, R. F.(Young, 2017) explained how very much planned HR rehearses related to performance evaluation and remuneration can be linked to workers' deplorable behaviour. Because of a normal decision perspective on moral conduct, various HR forms will likely build the obvious costs or benefits of worker offence. This paper joins clear HR deals with both saw costs and benefits of worker offense. At long last, this article ends up with offers designed to offer both effective work execution and boundary points of usual deplorable behaviour.

In the paper presented by Sharma (Sharma, 2018) on the improvement of organisational versions of training and development, on-the-job training, and training design and delivery styles. This mainly focuses on employee retention in the organization as Human resource is the numerous important part of it. Associations require the representatives to be exceptionally talented and experienced with the right psyche for smooth working and progress. This paper centers around the recent works on being persisted in companies for preparing and improving the representatives in enterprise with dissecting the effect of readying and advance developers on agents' work performance in the enterprises.

In the job done by Ngo, F. T. (Ngo, 2015) on forecasting Inmate misconduct – they used Regression and another variety of processes which contain –Logistic Regression (LR) and three Classification techniques–Type and deterioration Tree (CART), Chi-square automatic interchange detection(CHAID), and Multilayer Perceptron Neural network (MLPNN). A sample of prisoners housed in state and federal prisons was employed for these procedures. Their overall precision varies from 0.60 to 0.66 with AUC in the range of 0.60 to 0.70. They saw Logistic Regression and Multilayer perceptron neural networks gave better results.

These papers explore employee behaviour and its effect on the organization's financial health and reputation. In some papers, researchers use some techniques to predict it but the result was not so encouraging. There is still scope for using modern machine learning techniques to improve predictability. In this paper, present machine learning techniques have been used on the client dataset and created different machine learning models with the help of these models employee behaviour can be predicted with better accuracy.

Chapter 3: Problem Statement

The purpose is to predict the compliance violators based on their past data in the organization and classify them as high, medium, and low risk. So the organization can take corrective measures in advance.

Considering that data needs to be used, there are five main features to create the machine learning model in the first phase:

- **Sales Pattern:** If there are potent sales by Bank employees during specific months or periods, this will trigger suspicious activity. This variable can be considered one of the independent variables in /the model.
- **Misaligned incentives:** Banking outcomes are typically based on the customer profile and the worthiness of the customers. Occasionally, it is followed that banking employee accomplishes over commitment to complete their mark and offer something to the customer who does not fit a specific consumer part. This is a case of misaligned incentives and can be regarded as one of the variables in the compliance risk prediction.
- **Client Feedback:** It is also an essential element incorporated in Compliance risk prediction.
- **Account Use Data:** There are instances when a Banking employee has shared his credentials against the organization's policies, and sometimes multiple logins can be observed at once.
- **Employee Performance:** Bank charges all the pertinent data for the Employee Performance evaluation, which can be considered an essential factor for predicting compliance risk violations.

Chapter 4: Objectives of the Study

This study's goal is to classify each organisation employee into a different segment according to how likely they are to comply with business regulations, so in the case, they are following compliant as per anticipation or standard defined by Financial Conduct Authority (FCA) in that case, the employer can educate them by providing proper training or make them aware about the consequence of the bad compliance to the business or for organizational reputation. That way, an organization can avoid unnecessary monetary or reputational loss.

The objectives of the study are:

1. Collect the previous transaction history of the employee and the data about indicators of compliance violations
2. Understanding the repeated pattern in the dataset can help for the reoccurrences of the incident in the future also.
3. With the help of modern machine learning techniques developing statistical models to analyze and draw inferences from the data patterns.

Chapter 5: Project Methodology

In this report, the Customer, Employees, and Products data of the Financial Organization\ a Bank has been taken, and multiple machine learning models have been developed to predict the employee's compliance risk probability, how the employees conduct in dealing with customers and doing business. Based on the probability, it has been categorized into different segments.

Figure 5.1: Project Methodology followed

In Figure 5.1 following steps are followed:

1. Collect the Data from the client.
2. Pre-process the data: New features have been created.
3. Make the different Machine learning models, e.g., Logistic Regression, K-Nearest Neighbor, Naïve Bayes, Random Forest, Gradient Boosting
4. Analyze the result of different models based on the ROC curve and select the best one.
5. Predict the Compliance risk probability of Employees
6. Classify all Workers into High, Medium, and Low-risk categories.

Chapter 6: Business Understanding

Compliance risk is defined as "client responsibility" that results from unsuitable items ending up in the wrong hands and the inconvenience that results from people not having the opportunity to access the appropriate products.

Associations that ignore lead concerns risk legal action, financial loss, and reputational damage, all of which can harm a company long after the function. Organizations have had significant financial effects as a result of conduct-related regulatory action, which might result from a person's actions. The current report from the Fixed Payment, Currencies and Commodities Markets Standards Board (FMSB) gauges banks have paid some $375 billion in lead damages in the most recent long term.

Compliance risk in a financial services organisation can be defined as the risk that decisions and procedures result in poor or unfavourable outcomes for their clients and the risk that the company fails to uphold high standards of market conduct and purity. Compliance risk contacts in all parts of a venture arrangement.

There is no one-size-fits-all system for assessing compliance risk profiles for businesses because each one will be different. In budgetary management, firm-explicit meanings of compliance risk ought to have been made and provided related to characterized hazards that need fixing for compliance risk. As associations become more progressive and embrace a more comprehensive scope of inventions, the way to deal with determining and reducing compliance risk must advance.

Discussionis necessary to stimulate and create a convergence of meanings. Behavioural risk is a new concept that emerged as a result of the failure of collective behaviour in the financial services industry. As such, it lacks a specific definition because a universal approach in its management is not possible because firms have different behavioral risk profiles. However, it is also argued that convergence in terms of its characterization of the risk affecting consumer protection and market integrity is sufficient to provide a solid knowledge of this risk. This agreement is only sufficient to justify the existence of behavioral risk. This article argues that how behavioral risk is interpreted by regulators is far from offering clarity. Until now, regulators have been reluctant to bring any specific definition. This may depend on several reasons, including the fact that behavioral risk has emerged very recently, so further research is needed. At this stage there exist only general definitions that indicate how regulators interpret or consider this risk. Even these general understandings are not definitive as to the nature and extent of behavioural risk.

Critically, the paper shows that behavioural risk is trapped between being considered a subset of operational risk or as a separate risk. In the first dimension, the behavioral risk is included in another dimension risk; while the second dimension gives behavioral risk its autonomy. It seems that both arguments have strengths and weaknesses and therefore nothing conclusive. However, they represent a starting point for expanding the debate on behavioral risk identity. Maybe multifaceted approaches to behavioral risk management, but the question of whether action risk costs need an answer. This question cannot be ignored for several reasons.

Primarily, coordination across the financial services industry is urgently required to mitigate behavioral risk management. That is, a clear understanding of the nature and scope of the active risk is fundamental to achieving this goal. Second, it's clear deconstruction makes better predictions of how this risk will develop. As a result, awareness of the existence of action and the risk needs to be supplemented by an answer to the basic question of where it is located.

Chapter 7: Data Understanding

Data used for this project was provided by the client organization (Financial/ Bank). In this data,21 independent featuresand a target variable is available, where the target is the dependent feature which defines whether an employee is involved in any suspicious activities like insider trading/ mis-selling/collusion or the employee adheres to company compliance policies.

List of features:

1. Emp. Id
2. Account Usage data: 0- Single login, 1- Multiple Login
3. Employee Performance: 1- Lowest, 5-Highest
4. Customer Feedback: 1-lowest/completely unhappy; 10-highest/extremely happy
5. Customer segment:
   1 Lowest net worth customer,
   2 Lower medium net worth customer
   3 Medium net worth customer
   4 Upper medium net worth customer
   5 Highest net worth customer
6. Misaligned Incentives
   1. Lower range incentives
   2. Lower medium-range incentives
   3. Medium range incentives
   4. Higher medium-range incentives
   5. Higher range incentives
7. Target (Dependent Variable)
8. An employee involved in suspicious activities like insider trading/mis-selling/collusion
9. Employees following compliance risk policies properly
10. Sales (in volumes)
11. Balance
12. Income_levels
13. Tenure_with_bank (in years)
14. Credit_score (assuming CIBIL score)

15. Age (in years)
16. Points_of_improvement: 1-most/ lots of improvement needed, 10-leats improvement needed
17. Compliance: 1least compliant to companies policies/work guidelines, 5-highly compliant
18. Retail_acc_setup_errrate (in %): 2-30 (2 implies least error rateand 30 highest error rate)
19. Avg_time_to_close_issues (in hours): 8-72(8 - least time taken to close issuesand 72 max time taken)
20. Prod_id: 1-7(different products of a bank;1-3: low price, 4-5: med price, 6-7: high price)

    1-performing well

    0-not performing well
21. Prod_sell_count: 0-product being sold once, 1-product being sold multiple times
22. Prod_sell_perf: 0-not performing well even after being sold multiple times, 1-performing well after being sold multiple times
23. Product_price

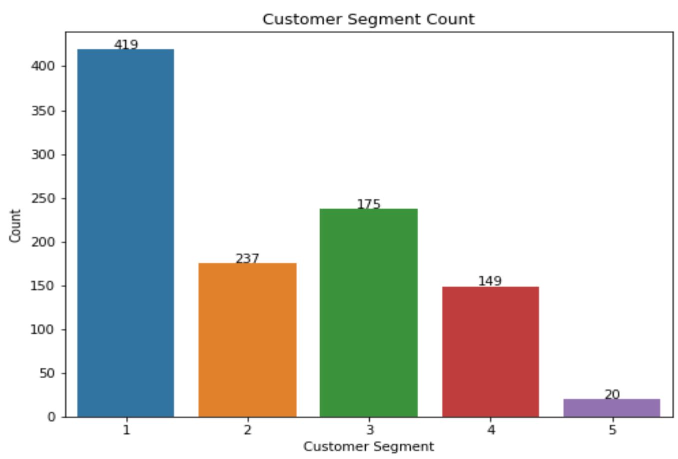


Figure 7.1. Different Customer Segments

In Figure7.1 it is depicted the majority (419) of customers belong to the Lowest net worth and the Highest net worth have just 20, the rest belongs to the medium range. Customer segments are :

1 Lowest net worth customer,

2 Lower medium net worth customer

3 Medium net worth customer

4 Upper medium net worth customer

5 Highest net worth customer

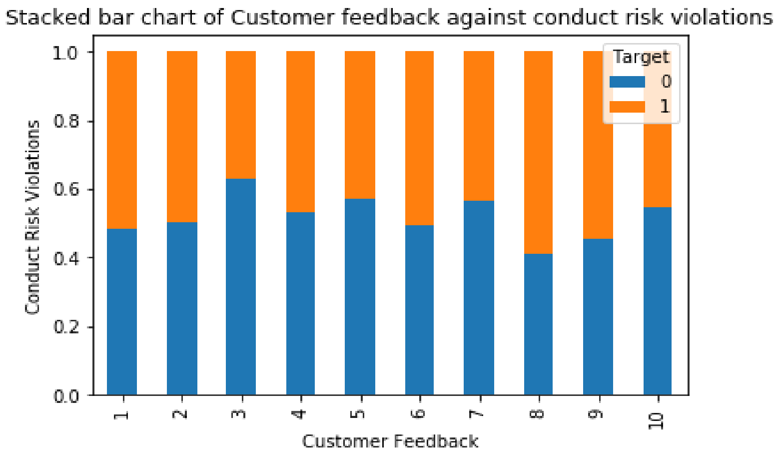


Figure 7.2. Customer feedback vs.Compliance risk Violations

In Figure 7.2 Customer feedback has 10 segments where segment 1-completely unhappy, and segment 10-extremely happy, this figure depicted those employees who are violating the rules to satisfy the customer who got good feedbackfrom them.

1 An employee involved in suspicious activities like insider trading/mis-selling/collusion

0 Employees following compliance risk policies properly

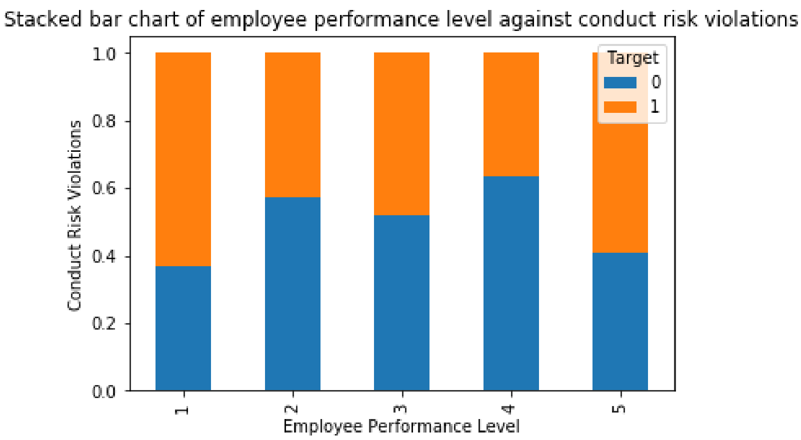


Figure 7.3. Employee performance (1- Lowest, 5-Highest)vs.Compliance risk Violations

In Figure 7.3 it has been noticed that the employees whose performance is very good or bad in those categories have more violators, whereas in the case of avg. performing employees are not involved much in those activities.

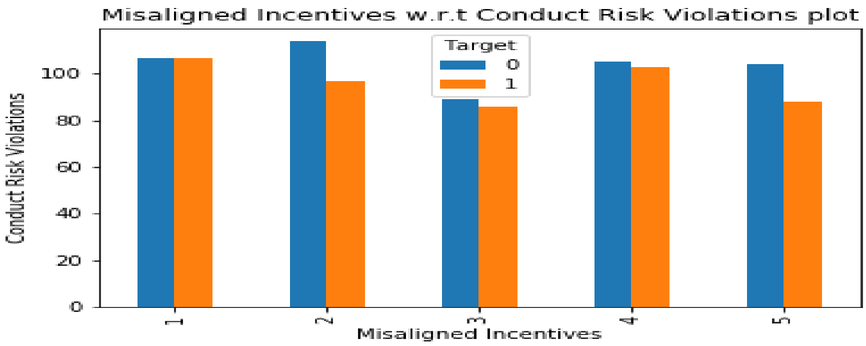


Figure 7.4. Misaligned Incentives (1-lower, 5 higher)vs.Compliance risk Violations

In Figure 7.4 it has been noticed that in the lower incentives 1 and higher incentive 4 segments violators are high as compared to other segments.

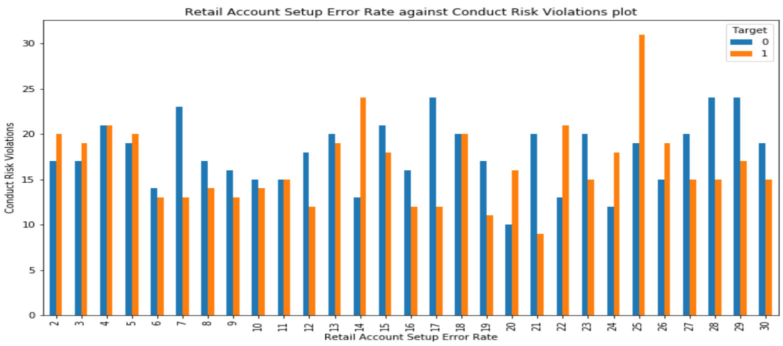


Figure7.5. Retail Account Setup Error Rate (2-30)vs.Compliance risk Violations

In Figure 7.5 it has been noticed that violators are equally distributed in the whole range, but the maximum is25. So it means all the employees are making errors during retail account setup

Some factors were identified as crucial to predictingcompliance risk violations are: -

1. **Sales Pattern** - Sales patterns are graphically represented collections of data that a company collects to review its sales information. This data focuses on the sales of a specific product or collection of products during a specific period.

2. **Misaligned Incentives** - Misaligned incentives usually occur in the absence of well-designed rules governing rewards or sanctions for participants. The basic idea is that unless the rules incentivize them to repeat, both people and organizations tend to act in their self-interest, which may not always be what was desired.

3. **Customer feedback** - Customer feedback is information provided by customers about their experience with a product or service. Its purpose is to reveal their satisfaction levels and help product, customer, and marketing teams understand where there is room for improvement.

4. **Account Usage data** - You can track your current data usage and the number of days left in your billing cycle, so there are no surprises on your next bill.

5. **Employee performance** - Employee performance is defined as how the employee fulfills his job duties and performs the required tasks. It refers to the effectiveness, quality, and efficiency of their output. The performance also contributes to our assessment of how valuable an employee is to the organization.

6. **Product Performance** - Product Performance KPI is one of the most used key performance indicators in companies across all sectors. It allows you to rank products based on their performance in terms of sales volume and revenue generated.

Chapter 8: Data Preparation

Making sure that the raw data is accurate and consistent before processing and analysing it is one of the main goals of data preparation in order to assure the validity of the output from BI and analytics tools. Missing numbers, inaccuracies, and other errors are frequently present in data, and disparate data sets frequently have different formats that need to be reconciled when joined. A significant portion of data preparation tasks involves erasing data mistakes, confirming data quality, and unifying data sets.

To ensure that analytics programmes deliver useful information and practical insights for business decision-making, data preparation also entails locating pertinent data. To make data more informative and usable, it is frequently enhanced and optimized—for instance, by combining internal and external datasets, developing additional data fields, removing outliers, and addressing unbalanced datasets that could skew analytical results.

Any business should be aware of all the risks it faces and the need to address them all at once because reducing risks can be costly and tax organisational resources. An organisation should instead rank risks and concentrate its time and energy on preventing the most significant hazards. Prioritizing risks can be accomplished by creating a risk assessment table.

The foundation of a risk assessment chart is the idea that risk has two main dimensions: likelihood and impact, each of which is depicted on a single axis. These two metrics can be used by an organisation to graph hazards, which enables us to set priorities and distribute resources.

Guidance has been provided to the customer segment based on different customer-related parameters, e.g., Account Balance, income classes, age, term with the bank, and Credit score. The customer has been broadly classified into five categories 1-5. Where one is the most down, those who are youngsters hold low income, low credit, and less than five years with the bank. All high-value consumers have been put in the 5th Segments who are high experts, have long tenure with banks, and have very increased income and fund balance with banks. The rest of the customers come in central components based on different elements.

The five elements have been defined for the workers, established on parameters associated with them e.g., means to progress, mistakes in retail Account setup, Average time they are bringing to approach the problem, and their Compliance ratings. The lower rating has been given to the employees who require a lot, taking 8-9 days to complete the problem, making more errors while putting up the Accounts and their adherence ratings are also the lowest.

Product performance has been segmented into two types, 1 or 0, based on their cost, the Customer who is taking it, and whether it has been taken single or numerous times.

Chapter 9: Data Modeling

Divided the dataset into train and test sections in 70: 30 ratios, train the model on the Train data set, and test it on the test data set.
Here to find the probability of the Employee compliance risk with various machine learning techniques:

1. Logistic Regression (LR): LR is a classification algorithm for supervised learning that is used to estimate the likelihood of a target variable. One of the simplest ML classification algorithms is this one.

2. K-nearest Neighbours (KNN): Both classification and regression prediction issues are addressed by KKN. Nevertheless, it is more frequently used in classification-related concerns in the sector.

3. Naïve Bayes: It's a classification strategy based on the Bayes Hypothesis with suspicion of autonomy among indicators. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. It's easy and fast to predict the class of test data set. It also performs well in multi-class prediction. When an assumption of independence holds, a Naive Bayes classifier performs better compared to other models like logistic regression and you need less training data.

4. Random Forest: It is like bootstrapping algorithm with a Decision Tree (CART) model. Random forest tries to construct different CART models with distinctive tests and diverse introductory factors. Random Forest gives much more accurate prediction when compared to CART\CHAID or regression models in many scenarios. This is because it captures the variance of the input variable at the same time and enables a high number of observations to participate in the prediction. As the sample size is less the overfitting issue has been detected with it, to overcome the cross-validation technique has been used.

5. Gradient Boosting: To provide the final predictions, it aggregates the predictions from many decision trees. All the weak learners in gradient boosting machines are decision

tree. Subsequently, they can capture distinctive signals from the information. Boosting algorithms plays a significant part in managing with Bias Variance trade-off. Not at all like Boosting algorithms, Boosting handles both the views (Bias and Variance) and is seen to be more successful than controlling for excessive variance in a model.

Name of Algorithims	Accuracy	Precision	Recall	F1-Score
Logistic Regression	**80**	**78**	**77**	**77.5**
Random Forest	77	68	70	68.98
KNN	69	66	63	64.47
Naïve Bayes	77.6	72	76	74
Gradient Boosting	78	70	72	70.98

Table No 9.1: Result metrics of all models

In Table No. 9.1 the Accuracy, Precision, Recall, and F-score of the Models against the test data have been summarized, out of all the algorithms Logistic Regression is giving the best results with 80% accuracy Gradient Boosting also givesa good result but its precision, recall, and F1 Score are not so good.

Chapter 10: Data Evaluation

The process of evaluation research consisting of data analysis and reporting is a rigorous, systematic process that involves the collection of data about organizations, processes, projects, services, and/or resources. Evaluation research improves knowledge and decision-making and leads to practical applications. In the figure given below, there are the true positive and false positive rate of receiver operating characteristics which involves logistic regression, random forest, KNN, naive Bayes, and gradient boosting.

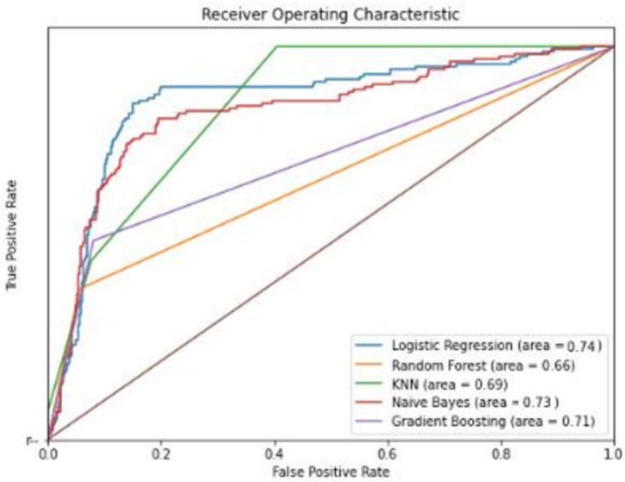


Figure 10.1. ROC curve of the different algorithms.

As per the ROC curve, the best score has been given by the Logistic Regression algorithm (0.74)

Based on the probability score by Logistic Regression, All employees are classified into three categories.

Probabilities	**Segment**
1.0 to 0.7	High Risk
0.7 to 0.3	Medium Risk
Less than 0.3	Low Risk

Table 10.1: Employee segments based on probabilities score.

Based on Table no 10.1 all employees have been segmented based on their probabilities score, as per that, 17.2 % of employees' risk assessments are on High Risk, 59.6% are on Medium Risk, and 23.2 % are on Low Risk.

Chapter 11: Deployment

The models have been provided to the client and suggested some important steps to improve the compliance score of their employees.

1. With the help of this model,an organization can analyze the employee's previous data and predict if the employee is following the compliance or not.
2. For improvement,an organization can take preventive steps to correct the behavior of the employee and avoid possible future losses.
3. Organizations can arrange important pieces of training for their employees to increase awareness of compliance in the business.

Chapter 12: Analysis and Results

Based on Employee possibility scores, employees are categorized as a high risk where possibility lies between 1.0 to 0.7, the medium risk for possibility between 0.7-0.3, and low riskwith probability less than 0.3.

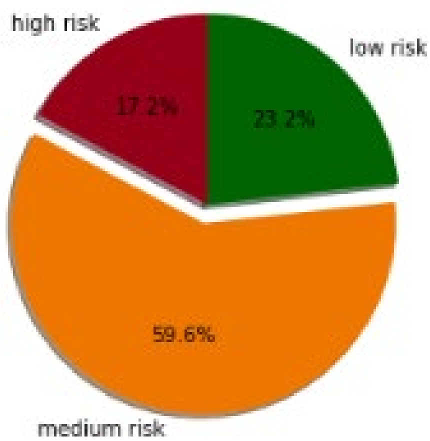


Figure 12.1: Segmentation of Employees based on probability score

Figure 12.1 with the use of the Logistic regression model, All employee data is segmented into three classes. Most of the employees are on medium risk, around 60%, 23% are on lower risk, and only 17 % of employees are categorized as high risk.

Chapter 13: Conclusions and Recommendations for future work

The developed model has been provided to the client thatputs all the employees in different categories.Some suggestion has also been provided for training and understanding courses to enhance their knowledge about the significance of Compliance risk and the results of its losses. A suggestion for some awards and praise for the employees who are doingpretty well in the area of conductwas also provided. Proper risk relief will lead to a happy client result as well as control the number of serious damages and penalties. It also sustains financial integrity and reputation, indicating improved consumer retention as well as precludes big fines. Risk assessment is not a monolithic process or a single method. All risk assessments share some common principles, but their application varies greatly from area to area.

In this report client data was analyzed against the basic machine learning techniques as the data was not so big. In the future, more advanced deep learning techniques can be used if there will be sufficient data to train them and get better accuracy or predictability.

Bibliography

Astuti, Y. (2021). Alfa Value Scalability on Single and Double Exponential Smoothing Comparatives. *IJISTECH (International Journal of Information System and Technology)* , 456-465.

Choi, J. H. (2019). Fraudulent Financial Reporting and the Consequences for Employees*. *SSRN Electronic Journal*

Francesco. (2018). Sales Culture and Misconduct in the Financial Services Industry: An Analysis of Cross-selling Practices. Business Law Review. 39.

FSB. (2016). FSB reports to G20 Leaders on financial regulatory reforms. *https://www.fsb.org/2016/08/fsb-reports-to-g20-leaders-on-financial-regulatory-reforms/* .

HEAKAL, R. (1933). *What Is the Glass-Steagall Act?* Retrieved from Investopedia: https://www.investopedia.com/articles/03/071603.asp

Hennessy et, a. (2017). Federal Reserve Bank of New York Misconduct Risk, Culture, and Supervision.

Mark Egan, G. M. (2019). The Market for Financial Adviser Misconduct. *University of Chicago* .

Mehrotra, K. (2016). Wells Fargo Sued by Shareholders Over Cross-Selling Scandal - Bloomberg. *Bloomberg* .

Namasivayam, K. (2006). Linking employee misbehavior to consumer satisfaction. *Journal of Foodservice Business Research* .

Ngo, F. T. (2015). Assessing the Predictive Utility of Logistic Regression, Classification and Regression Tree, Chi-Squared Automatic Interaction Detection, and Neural Network Models in Predicting Inmate Misconduct. *American Journal of Criminal Justice* .

Resti, A. (2017). DIRECTORATE-GENERAL FOR INTERNAL POLICIES UNION ECONOMIC GOVERNANCE BANKING Fines for misconduct in the banking sector – what is the situation in the EU ? External aut. *https://www.europarl.europa.eu/committees/en/ECON/home/highlights* .

Sharma, S. (2018). The effect of training on employee performance. *International Journal of Recent Technology and Engineering* .

Tracy, S. C. (2017). Federal Reserve Bank of New York Misconduct Risk, Culture, and Supervision. *www.gallup.com/poll/1597/Confidence-Institutions.aspx.*

Young, R. F. (2017). Blowing the whistle: Individual persuasion under perceived threat of retaliation. *Behavioral Research in Accounting* .

Appendix

Plagiarism Report[1]

Reducing Business Compliance Risk with help of Analytics

ORIGINALITY REPORT

8%	5%	1%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

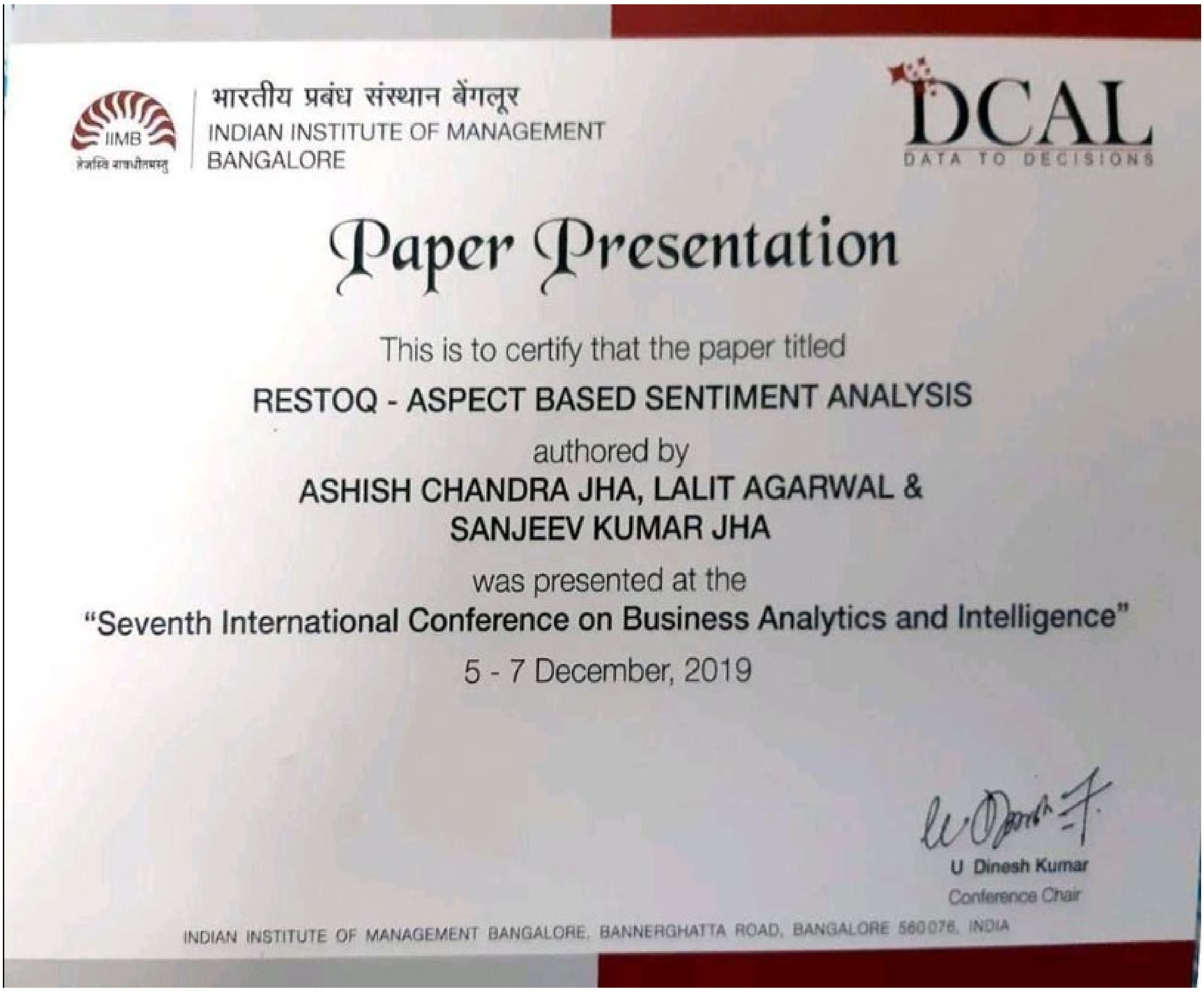
1	www.newyorkfed.org Internet Source	2%
2	www.sfari.org Internet Source	2%
3	medium.com Internet Source	1%
4	Submitted to IDEA Leadership & Management Institute Student Paper	1%
5	Submitted to Southampton Solent University Student Paper	<1%
6	Fawn T. Ngo, Ramakrishna Govindu, Anurag Agarwal. "Assessing the Predictive Utility of Logistic Regression, Classification and Regression Tree, Chi-Squared Automatic Interaction Detection, and Neural Network Models in Predicting Inmate Misconduct", American Journal of Criminal Justice, 2014 Publication	<1%
7	www.coursehero.com	

[1]Turnitn report to be attached from the University.

	Internet Source	<1 %
8	Submitted to Zovio Inc Student Paper	<1 %
9	Submitted to British School of Commerce - Colombo Student Paper	<1 %
10	Submitted to Liverpool John Moores University Student Paper	<1 %
11	Submitted to Indian School of Business Student Paper	<1 %
12	Submitted to Chartered Banker Institute Student Paper	<1 %
13	www.emerald.com Internet Source	<1 %
14	Submitted to University of Greenwich Student Paper	<1 %
15	academic-accelerator.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		

Publications in a Journal/Conference Presented/White Paper[2]



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

DCAL
DATA TO DECISIONS

Paper Presentation

This is to certify that the paper titled
RESTOQ - ASPECT BASED SENTIMENT ANALYSIS
authored by
ASHISH CHANDRA JHA, LALIT AGARWAL &
SANJEEV KUMAR JHA
was presented at the
"Seventh International Conference on Business Analytics and Intelligence"
5 - 7 December, 2019

U Dinesh Kumar
Conference Chair

INDIAN INSTITUTE OF MANAGEMENT BANGALORE, BANNERGHATTA ROAD, BANGALORE 560076, INDIA

[2] URL of the white paper/Paper published in a Journal/Paper presented in a Conference/Certificates to be provided.

RestoQ – Aspect Based Sentiment Analysis

Ashish Chandra Jha [1,†]
REVA University
Bangalore, India
Ashish.ba05@reva.edu.in

Lalit Agarwal
REVA University
Bangalore, India
Lalit.ba05@reva.edu.in

Sanjeev Kumar Jha [2,†]
REVA University
Bangalore, India
Sanjeev.ba05@reva.edu.in

†A.C.J. and S.K.J. contributed equally to this work.

***Abstract* – People love experimenting on their food with different tastes. And when it's about visiting the restaurant or ordering food online, they will definitely look for the reviews which will talk about the aspects like services, ambience and cost along with food quality. This food problem is not a single day problem, it's getting repeated everyday. Some end up with positive reviews and few end up with negative or neutral reviews. In this work a framework is developed called 'RestoQ', which uses text analytics for sentiment analysis at the aspect level to discover and rank the restaurants. The framework analyzes the reviews for the sentiments across four aspects – price, food quality, service quality and ambience. Unsupervised lexicon-based classifier and a naïve Bayesian classifier are used to evaluate and score the sentiments at aspect level. The final score will be a combined sum of each score for the review, which requires further work rank the aspects based on reviews. Surprisingly unsupervised method out performs the supervised method. It is proposed to extend the work with context based methods using word2vect and LSTM.**

I. INTRODUCTION

Before ordering food or booking a table in any restaurant consumers generally check the reviews of the places. Online food ordering sites like Zomato, Food Panda and UberEats do sentiment analysis of the reviews given by the customers and give a rating for these restaurants. The majority of current sentiment analysis approaches try to detect the overall polarity of the reviews or sentence regardless of the target entities (e.g. restaurants) and their aspects (e.g. services, ambience and cost along with food quality). Aspect Based Sentiment Analysis is fine grained sentiment analysis. A sentence may contain multiple opinions about different entities and we need to find each of them. This has to be analysed by model and should give insights. In this work, the research findings of such a system are presented.

II. LITERATURE SURVEY

Sentiment analysis is one of the fastest growing research areas in computer science, making it challenging to keep track of all the activities in the area. It is a case of natural language processing which could mark the emotion or mood of the people about any specific product by analysis. It is a process of automatic extraction of features by mode of notions of others about specific product, services or experience. [1]

Customers as well as ecommerce companies (online food ordering in this case) are looking for the reviews of the restaurants to order food or to check their customer satisfaction ratio. A lot of research has been done on Sentiment analysis on restaurants and their reviews. Reviews are considered to be positive, negative or neutral on the overall score of the sentence. To some extent it is very useful and many customers are using it before ordering their food on daily basis [2][3].

Unlike document level sentiment classification task, aspect based sentiment analysis is a more fine-grained classification task. It aims at identifying the sentiment polarity (e.g. positive, negative and neutral) of one specific aspect in its context sentence. For example, given a sentence "great food but the service was dreadful" the sentiment polarity for aspects "food" and "service" are positive and negative respectively [4].

Aspect Based Sentiment Analysis (ABSA) was introduced as a shared task for the first time in the context of SemEval in 2014; SemEval2014 Task 41 (SE-ABSA14) provided datasets of English reviews annotated at the sentence level with aspect terms (e.g., "mouse", "pizza") and their polarity for the laptop and restaurant domains, as well as coarser aspect categories (e.g., "food") and their polarity only for restaurants (Pontiki et al., 2014). SemEval-2015 Task 122 (SE-ABSA15) built upon SE-ABSA14 and consolidated its subtasks into a unified framework in which all the identified constituents of the expressed opinions (i.e., aspects, opinion target expressions and sentiment polarities) meet a set of guidelines and are linked to each other within sentence-level tuples (Pontiki et al., 2015) [5][6][7].

Aspect Based Sentiment Analysis poses several challenges in processing text data, is a popular area of research in this direction. Several challenges which has not been addressed and people are trying to do some research are implicit aspect detection, mapping aspect words to categories, resolving anaphora references etc. Researchers combine techniques from common sense rules, unsupervised supervised and semi supervised techniques to perform these tasks.

Aspect Based Sentiment Analysis has been done for this particular topic by various researchers [8] [9] [10]. In this paper, the four aspects depending on which the comments will be reviewed. It has been seen that there are lots of aspects which affects the overall sentiment of the review. For example, in restaurants, people give review based on food quality, services, ambience and price. In this work restaurants will categorized based on the customer reviews. The goal is to determine the sentiment expressed toward each aspect on restaurant of Bangalore in English language.

The problem of aspect-based sentiment analysis deals with classifying sentiments (negative, neutral, positive) for a given aspect in a sentence. A traditional sentiment classification task involves treating the entire sentence as a text document and classifying sentiments based on all the words [11].

Labeling of data is a little difficult task to perform automatically. Most of the researcher who are working on new dataset used to label the data manually. The lack of labeled data has led to several researchers to explore unsupervised learning techniques to learn both aspects and their sentiments expressed in plain text. Particularly the fact that aspects are normally described by opinion words and opinion words in turn will have a target aspect can be used to iteratively expand the sentiment and aspect lexicon. The expansion is done with the help of rules to associate aspects and sentiment [12][13].

In this paper, we are trying to do aspect based sentiment analysis on restaurant reviews data from an online food delivery site (Zomato). We have introduce a system based on Text Analytics on the reviews using Supervised Machine Learning with a Naïve Bayes algorithm and unsupervised Machine Learning with Lexicon based algorithm for scoring sentiments

III. Research Methodology

In this section, we will explore the different techniques, methods, and features used in this experiment. We will divide the section into two sections: data exploration and pre-processing and model building. Model building is further divided into supervised ML and Unsupervised ML.

Data access: 2000 restaurant across Bangalore along with their reviews has been collected. The data is from an online food ordering company i.e. Zomato. Labeling of data is the hard part of any new research and is done manually Here the reviews has been labeled based on restaurants name, aspects and sentiment. Positive, negative and neutral sentiments have been used as the three classes.

Data Exploration: Data has the solution to every problem. But one must know how to use that data. Data exploration gives the ability to summarize the main characteristics of a data set, including its size, accuracy, initial patterns, null values, outlier values and missing values. It can use a combination of manual methods and automated tools such as data visualizations, charts, and initial reports to explore the data.

Data Preprocessing: It is the most vital part of any analysis. Considering few important preprocessing steps, below mentioned techniques has been used

- Stopwords Removal - Stopwords are the meaningless and repeated words which do not contribute to the semantic of the statement. It should be removed.
- Symbol Removal – Reviews generally contains symbols like @,#,$ with no contribution towards analyzing the sentiments. So, it should be removed.
- Contractions and Annotation Removal – The contractions and annotation like shouldn't should be removed with 'should not'
- Normalization - Normalization stands for making the word or sentence case insensitive. Data should be normalized.
- Exploration – It is to check the word frequency of the corpus. It gives the idea of what the document is about. We check the word frequency by TF-IDF model. Words with high frequency can be seen using word cloud. In addition conditional exploration, based on sentiments and aspects word cloud has been made.
- For sentiments three word cloud, one for positive, one negative and one for neutral emotions. For aspect four word cloud has been made based on food quality, services, ambience and cost.

Model Building: The model will be created four times with different strategies. Here combination of 90%-10%, 80%-20%, 70%-30% and 60%-40% train/test split along with 10 fold cross validation has been used.

Starting with the supervised learning model then tried unsupervised learning model has been created to compare which algorithm will perform better.

In Supervised Machine Learning stage, label data is used for building classifier using Naive Bayes algorithm. Naïve Bayesian algorithm is a probabilistic ML algorithm, which assume independence among the features.

In Unsupervised Machine Learning stage, label data is used for building classifier using Lexicon based algorithm. The lexicon based approach is based on the assumption that the contextual sentiment orientation is the sum of the sentiment orientation of each word or phrase.

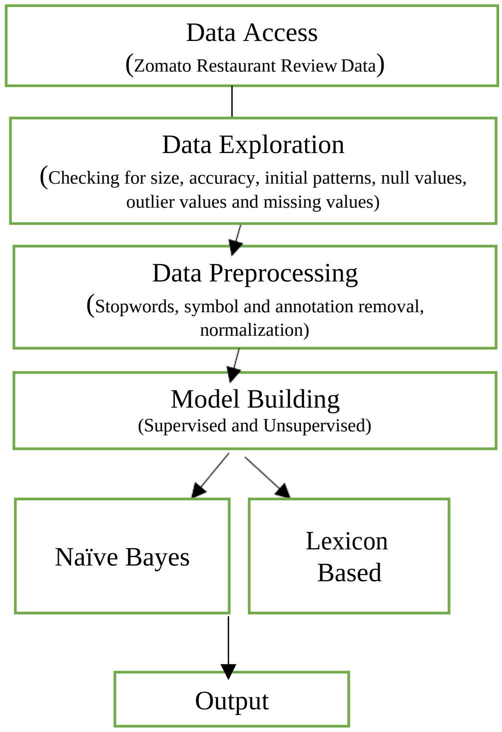


Fig.1: Proposed Approach

IV. Results and Discussion

Supervised (Lexicon based model) and Unsupervised (Naive Bayes model) has been created at different train test split. For each instant, accuracy has been captured and report is mentioned below:

Table 1. Experimental results at review level

		Accuracy	
Training	**Test**	**Lexicon**	**Naive Bayes**
90%	10%	71%	72%
80%	20%	68%	65%
70%	30%	70%	65%
60%	40%	68%	63%

The classification accuracy of all the models are consistent with the results published in literature and hence support the methodology used in this research. From the result, it clearly shows, the unsupervised learning lexicon-based model performs better than supervised learning technique using Naive Bayes. Accuracy at different train test split given an idea of optimum split scoring highest accuracy.

V. CONCLUSIONS

This paper covers the Aspect Based Sentiment Analysis on restaurants reviews dataset for Bangalore restaurants. The ABSA task consists of four aspects namely food quality, services, ambience and cost. For each aspect, sentiments have been analyzed. Supervised and Unsupervised machine learning has been used.

The proposed approaches achieved very good results. The algorithm successfully able to analyze the aspects of the sentiments. Further the restaurants are ranked based on the over all score and the positive score, which can be used by consumers for selection of restaurants. It is proposed to carry out a context-based analysis of the sentiments using word2vec and LSTM to test the improvements in the accuracies.

REFERENCES

[1] "A Study on Sentiment Analysis: Methods and Tools by Abhishek Kaushik1, Anchal Kaushik2, Sudhanshu Naithani3 https://pdfs.semanticscholar.org/c151/dfad8c1bf88b0afc716758c77d533ded7dd0.pdf"

[2] "Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews https://arxiv.org/ftp/arxiv/papers/1709/1709.08698.pdf"

[3] "SENTIMENT ANALYSIS OF RESTAURANT REVIEWS USING HYBRID CLASSIFICATION METHOD http://iraj.in/journal/journal_file/journal_pdf/4-54-140014488817-23.pdf"

[4] "Sethia, A., & Bhattacharyyaa, P. Aspect Based Sentiment Analysis-A Survey. Accessed on August 2019; http://www.cfilt.iitb.ac.in/resources/surveys/aspect-based-sentiment-analysis_survey.pdf"

[5] "Bhoi, A., & Joshi, S. (2018). Various Approaches to Aspect-based Sentiment Analysis. arXiv preprint arXiv:1805.01984"

[6] "Brychcín, T., Konkol, M., & Steinberger, J. (2014, August). Uwb: Machine learning approach to aspect-based sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 817-822) "http://alt.qcri.org/semeval2014/cdrom/pdf/SemEval2014145.pdf

[7] "Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks Binxuan Huang, Yanglan Ou and Kathleen M. Carley https://arxiv.org/pdf/1804.06536.pdf"

[8] 8. "Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., & Hoste, V. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30). https://www.aclweb.org/anthology/S15-2082"

[9] "SemEval-2016 Task 5: Aspect Based Sentiment Analysis Maria Pontiki∗1 , Dimitrios Galanis1 , Haris Papageorgiou1 , Ion Androutsopoulos1,2 , Suresh Manandhar3 , Mohammad AL-Smadi4 , Mahmoud Al-Ayyoub4 , Yanyan Zhao5 , Bing Qin5 , Orphée De Clercq6 , Véronique Hoste6 , Marianna Apidianaki7 , Xavier Tannier7 , Natalia Loukachevitch8 , Evgeny Kotelnikov9 , Nuria Bel10 , Salud María Jiménez-Zafra11 , Gülşen Eryiğit12 https://www.aclweb.org/anthology/S16-1002.pdf"

[10] "SemEval-2015 Task 12: Aspect Based Sentiment Analysis Maria Pontiki*, Dimitrios Galanis*, Haris Papageorgiou*, Suresh Manandhar± , Ion Androutsopoulos◊ * https://www.aclweb.org/anthology/S15-2082.pdf"

[11] "Various Approaches to Aspect-based Sentiment Analysis Amlaan Bhoi Department of Computer Science University of Illinois at Chicago Chicago, IL, USA abhoi3@uic.edu Sandeep Joshi Department of Computer Science University of Illinois at Chicago Chicago, IL, USA sjoshi37@uic.edu . https://arxiv.org/pdf/1805.01984.pdf"

[12] "http://UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis Toma´s Brychc ˇ ´ın NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitn´ı 8, 306 14 Plzenˇ Czech Republic brychcin@kiv.zcu.cz Michal Konkol NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitn´ı 8, 306 14 Plzenˇ Czech Republic konkol@kiv.zcu.cz Josef Steinberger Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitn´ı 8, 306 14 Plzenˇ Czech Republic jstein@kiv.zcu.cz https://www.aclweb.org/anthology/S14-2145.pdf "

[13] "Semi-supervised Aspect Based Sentiment Analysis for Movies using Review Filtering Deepa Ananda Deepan Naorema https://www.sciencedirect.com/science/article/pii/S1877050916300850 "