

# Insurance Claim Analysis Using Machine Learning Algorithms

Rama Devi Burri, Ram Burri, Ramesh Reddy Bojja, Srinivasa Rao Buruga

*Abstract—Now a day's Data is playing a central role and is carrying the big asset in the insurance industry. In today's journey insurance industry has a vital role. Insurance transporters have access to more information than ever before. From the past 700+ years in the insurance industry we can consider the three major eras Starting from 15th century to 1960, industry followed the manual era, from 1960s to 2000 we are in the systems era, now we are in digital era i.e. 2001-20X0. The highest corporate object in all three eras is that the fundamental insurance industry has been determined by believing the data analytics in adopting the changing technologies to better and keep the ways and keep capital together. In advanced analysis the main challenge is the analytical models and algorithms which are being insufficient to support insurers; only by machines we can overcome this challenge.*

*Key Words — Machine Learning, Digital Era, Client Lifetime Value (CLV).*

## 1. INTRODUCTION

The most important advantage of Machine Learning (ML) to use in Insurance Industry is to facilitate data sets. Machine learning (ML) can be successfully useful across Structured, Semi Structured or Unstructured datasets. Machine learning can be used accurate across the value chain to identify with risk, claims and customer actions, by means of advanced predictive accurateness. The probable applications of machine learning in insurance are plentiful from perceptive risk appetite and premium leakage, to expense administration, subrogation, proceedings and fraud detection. Machine learning is not a novel technology; this technology is following from the last few decades. There are 3 main categories of learning they are supervised learning, Unsupervised Learning and reinforcement learning. The greater part of the insurers are following Supervised Learning from last few decades for assessing the risk by means of known parameters in dissimilar combinations to acquire the preferred outcome. Present age insurers are motivated to unsupervised learning, in this predestined goals are clear. If there are any modifications in the variables, the method identifies those modifications and tries to change as per the goals. For example according to traffic the GPS Suggests different routes dynamically based on traffic conditions. In insurance industry also the learning is adopted for usage based insurance. Reinforcement learning is mostly depends on ANN (Artificial Neuron Network), Target/ Goal

can be modified dynamically depending on objective. Reinforcement learning is used for IOT applications.

## 2. TRANSFORMING MACHINE LEARNING TO THE INSURANCE SECTOR:

We have three ways to transform machine learning techniques in to insurance industry; they are 1. Automated and Personalized product offerings, 2. Improved Risk Assessment, 3. Enhanced Fraud Detection.

### 2.1. Automated and personalized product offerings:

Automated processes have a important impact on the insurance industry, to analyze large portion of data and gain more specific perspective of the client activities. Clients allow insurers to go into standard product offerings with common pricing and features. Insurers can offer personalized products and solutions which are based on the specific needs of narrow segments. Insurers benefit from accepting the progress patterns of clients activities.

### 2.2. Improved Risk Assessment:

When compared to humans, Machine learning delivers additional perfect predictions. Insurance companies can construct secure predictions on coverage changes, and possible losses for policies and manage risks more effectively. Advanced machine learning algorithms can develop risk management. Insurers can assess customer lifetime value (CLV) to compute the client's productivity for their insurance corporation.

### 2.3. Enhanced Fraud detection:

Fraud is a growing in every sector and in the insurance sector estimates billions of dollars to the industry. Machine learning algorithms can simply eliminate human errors and separate unobserved fraud patterns by identifying exceptions. Insurance companies rely on predictive models use the preceding cases of falsified actions. The machine learning algorithms with new data have much more perfect fraud detections.

## 3. THE DYNAMIC FACTORS THAT IMPACT THE IMPLEMENTATION OF AI AND ML:

The dynamic factors that implement Artificial intelligence and machine learning from 2018.

### 3.1. Smart Everything:

Advanced machine learning algorithms are used to force smart and automated applications in field of health care analysis, predictive maintenance, customer services, self

Revised Manuscript Received on April 12, 2019.

Rama Devi Burri, LBR College of Engineering, Mylavaram, A.P, India.

Ram Burri, Whirlpool Corporation, Michigan, USA.

Ramesh Reddy Bojja, LBR College of Engineering, Mylavaram, A.P, India..

Srinivasa Rao Buruga, LBR College of Engineering, Mylavaram, A.P, India.

driving cars smart homes and Insurance industry.

3.2. Open source everywhere:

Data becomes universal, open source protocols come in view to make sure that data is used and shared across public and private entities, these can create for sharing data on multiple frame works.

3.3. Harnessing (IOT) Internet of Things:

From IOT we can automate the generation of actionable approaching by means of advanced machine learning tools. According to Garter, by the year 2020, twenty percent of enterprises will employ enthusiastic citizens to monitor and guide machine learning tools.

3.4. Ability to talk back:

NLP (Natural language processing) algorithms are endlessly advancing. Artificial intelligence is capable to understand the facial recognition and spoken language, and helping to make more useful and intuitive, Like Google translator which is more effective.

4. MACHINE LEARNING IN INSURANCE

The resistance to machine learning surrounded by the industry is multi-faceted –starting from the underwriters and the information technology teams to operational management. A variety of reasons contain the following steps:

- Loss of jobs
- Focus on current issues like building the scalable foundation for their current journey
- Non-availability of resources and dearth of required skill-set
- Financial constraints in form of funding the development/POC/research stages
- Insurers want clear evidence of success before adoption
- Regulatory restrictions and privacy norms
- Large diversity of training for real-world operation and simulation of various scenarios
- Initiation challenges
- Infrastructural challenges faced by a neural network designer include filling millions of database rows for its connections.

Figure 1: This figure illustrates the improvement of the machine learning market in diverse geological regions over 10 years. It shows the accelerating approval of Artificial intelligence (AI) and the critical significance of technology trend. Global AI market, by geography 2017–2024 (in US\$ M).

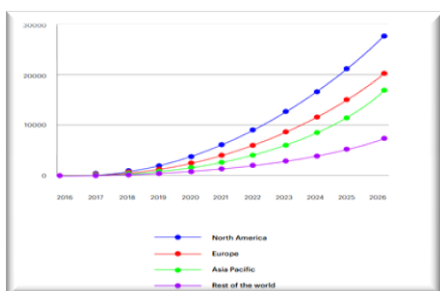
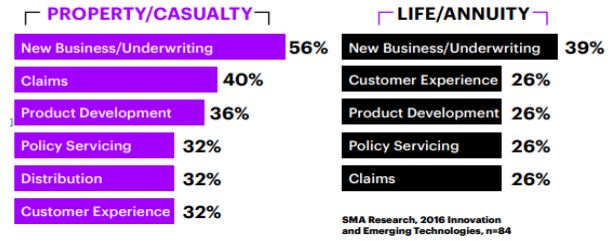


Fig1: Machine learning for insurance business Areas

Machine learning is extensively used across the insurance value chain.



5. POTENTIAL USE CASES

5.1. Insurance Advice

Machines will play a important role in client service, from organizing the preliminary communication to determine and cover a client necessities. According to a latest survey, a greater part of clients are satisfied to take delivery of computer-generated insurance recommendations. Clients are looking for tailored solutions which are made by machine learning algorithms that evaluate their profiles and recommend tailor-made products. PROPERTY/CASUALTY SMA Research, 2016 Innovation and Emerging Technologies,.

From the above figure n=84, Insurance business areas are machine learning and can be leveraged Machine learning is expansively used across the insurance value chain. One example is all state, which combined with (EIS) Early Information Science to develop a virtual assistant, called the Allstate Business Insurance Expert (ABIE).

5.2. Claims Processing

Insurers are using machine learning to progress operational competence, from claims registration to claims settlement. Many carriers have already started to automate their claims processes, thereby enhancing the customer experience while reducing the claims settlement time. Machine learning and predictive models can also equip insurers with a better understanding of claims costs. These insights can help a carrier save millions of dollars in claim costs through proactive management, fast settlement, targeted investigations and better case management. Insurers can also be more confident about how much funding they allocate to claim reserves. Tokio Marine has an AI-assisted claim document recognition system that helps to handle handwritten claims notice documents using a cloud-based AI optical character recognition (OCR) service. It reduces 50 percent of the document input load as well as complies with privacy regulations. AI is used to read complicated, ambiguous Chinese characters (Kanji), and the “packet like” data transfer system protects customer privacy. The results: over 90 percent recognition rate, 50 percent reduction in input time, 80 percent reduction in human error, and faster and hassle-free claims payments.

5.3. Price optimization

Price optimization procedure is a complex notion. Therefore it uses numerous combinations of various methods and algorithms. Despite the fact that it is still the

disputable issue of applying this procedure for insurance, more and more insurance companies adopt this practice. This process supposes combining the data not related to the expected costs and risk characteristics and the data not related to the expected loss and expenses, and its further analysis. That is, it takes into consideration the changes in comparison to the previous year and policy. Thus, price optimization is closely related to the customers' price sensitivity. In other words, historical costs, expenses, claims, risk, and profit are projected into the future. Special algorithms give the insurers the opportunity to adjust the quoted premiums dynamically. As a key positive feature, price optimization helps to increase the customers' loyalty in long perspective. Along with this, comes the maximization of profit and income.

#### 5.4. Personalized marketing

The customers are always willing to get personalized services which would match their needs and lifestyle perfectly well. The insurance industry is not an exception in this case. The insurers face the challenge of assuring digital communication with their customers to meet these demands. Highly personalized and relevant insurance experiences are assured with the help of the artificial intelligence and advanced analytics extracting the insights from a vast amount of the demographic data, preferences, interaction, behavior, attitude, lifestyle details, interests, hobbies, etc. The consumers tend to look for personalized offers, policies, loyalty programs, recommendations, and options. The platforms collect all the possible data to define the major customers' requirements. After that, the hypothesis on what will work or won't work is made. Here comes the turn to develop the suggestion or to choose the proper one to fit the specific customer, which can be achieved with the help of the selection and matching mechanisms. The personalization of offers, policies, pricing, recommendations, and messages along with a constant loop of communication largely contribute to the rates of the insurance company.

#### 5.6. Customer segmentation

Modern technologies have brought the promotion of products and services to a qualitatively new level. Different customers tend to have specific expectations for the insurance business. Insurance marketing applies various techniques to increase the number of customers and to assure targeted marketing strategies. In this regard, customer segmentation proves to be a key method. The algorithms perform customers' segmentation according to their financial sophistication, age, location, etc. Thus, all the customers are classified into groups by spotting coincidences in their attitude, preferences, behavior, or personal information. This grouping allows developing attitude and solutions especially relevant for the particular customers. As a result, target cross-selling policies may be developed and personal services may be tailored for each particular segment.

#### 5.7. Fraud Prevention

Insurance companies lose an estimated US\$30 billion per year to falsified claims. Machine learning helps them identify potential fraudulent claims faster and more accurately, and flag them for investigation. Machine

learning algorithms are superior to traditional predictive models for this application because they can tap into unstructured and semi-structured data such as claims notes and documents as well as structured data, to identify potential fraud. Chola MS, one of India's fastest-growing insurance companies, has adopted mobile technology for its claims survey process. The company's vehicle surveyor application uses the voice, camera and data connectivity capabilities of the Samsung Galaxy Tablet to capture and store auto survey data in one database. In the past, loss adjusters had to manually match survey notes with e-mail and photos saved in other databases before making a decision on a claim. This initiative helped to speed up the claims settlement process, increased surveyor productivity and improved fraud prevention.

#### 5.8. Risk Management

Insurers use machine learning to predict premiums and victims for their policies. Detecting risks near the beginning in the process enables insurers to make improved exploit of underwriters' time and gives them a vast competitive improvement. Progressive Insurance is evidently leveraging machine learning algorithms for predictive analytics based on data collected from client drivers. The car insurer claims that its telematics mobile app, Snapshot, has collected 14 billion miles of driving data. To encourage the use of Snapshot, Progressive offers "most drivers" an auto insurance discount averaging US\$130 after six months of use.

## 6. OTHER APPLICATIONS& RESULTS

These are just some examples of potential use cases. Insurers are also seeing significant benefits from using machine learning across functions such as direct marketing, audits, claims prediction and customer retention.

## 7. CLAIMS PREDICTION

The insurance companies are tremendously interested in the prediction of the future. Accurate prediction gives a probability to decrease financial loss for the company. The insurers use rather complex methodologies for this purpose. The major models are a decision tree, a random forest, a binary logistic regression, and a support vector machine. A great number of different variables are under analysis in this case.

The algorithms involve detection of relations between claims, implementation of high dimensionality to reach all the levels, detection of the missing observations, etc. In this way, the individual customer's portfolio is made.

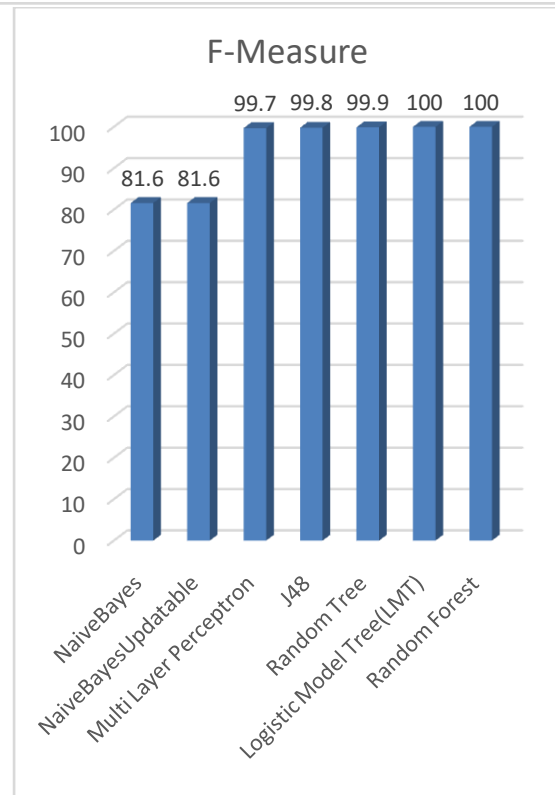
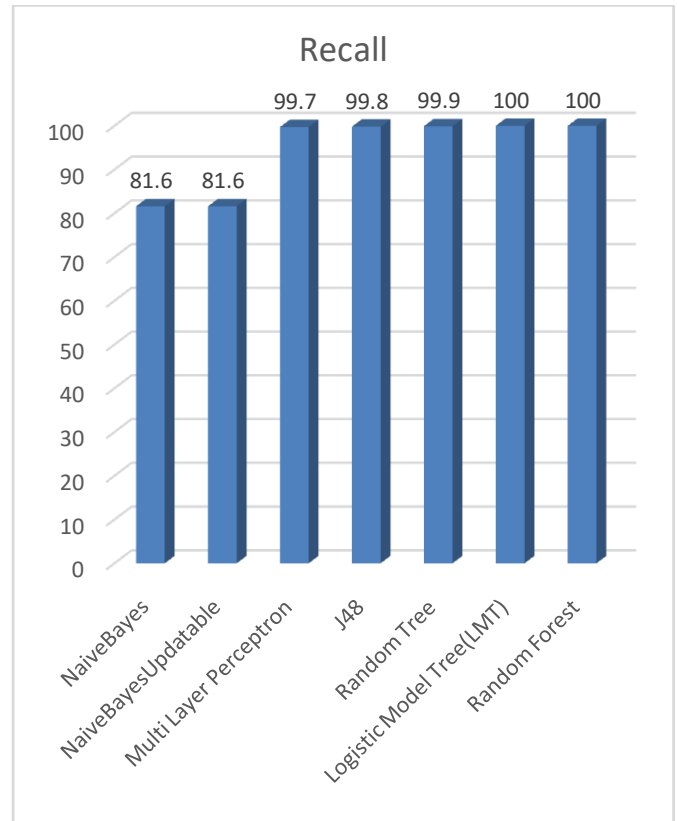
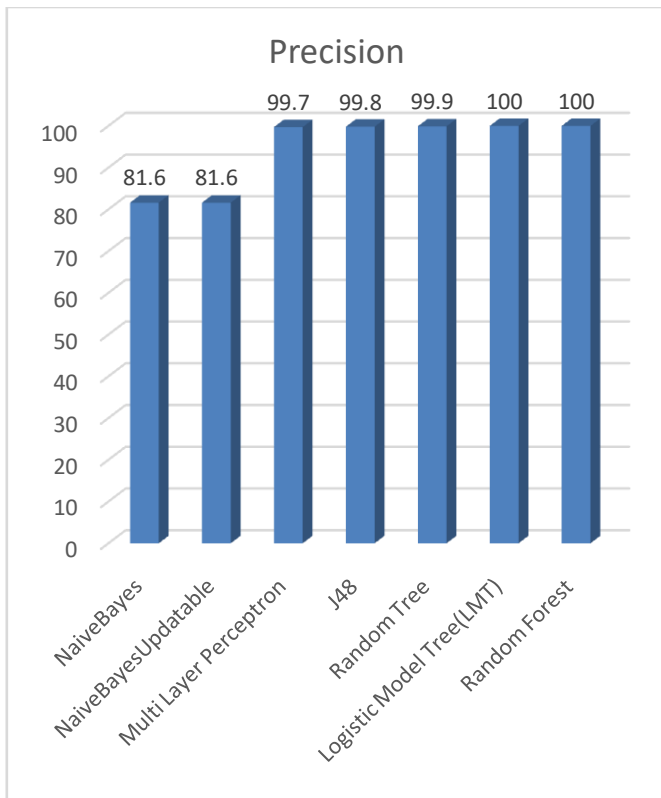
Forecasting the upcoming claims helps to charge competitive premiums that are not too high and not too low. It also contributes to the improvement of the pricing models. This helps the insurance company to be one step ahead of its competitor.

## 7.1 Datasets are collected from the website

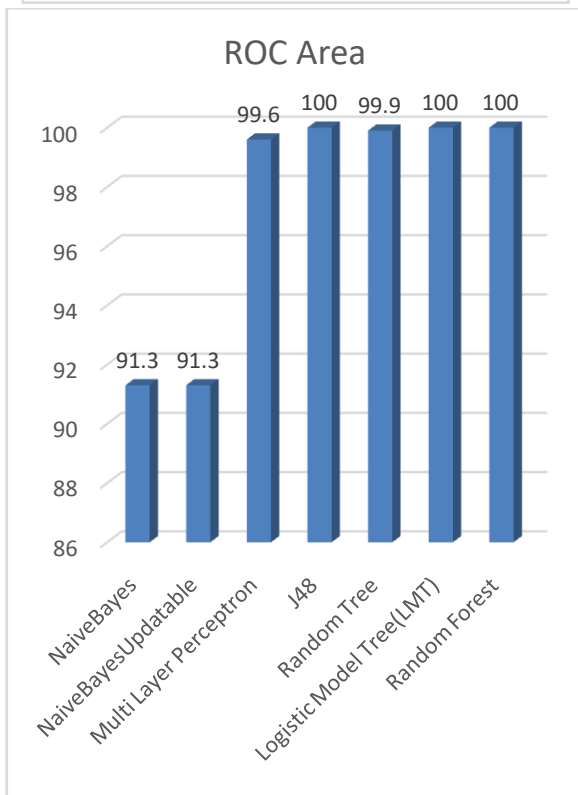
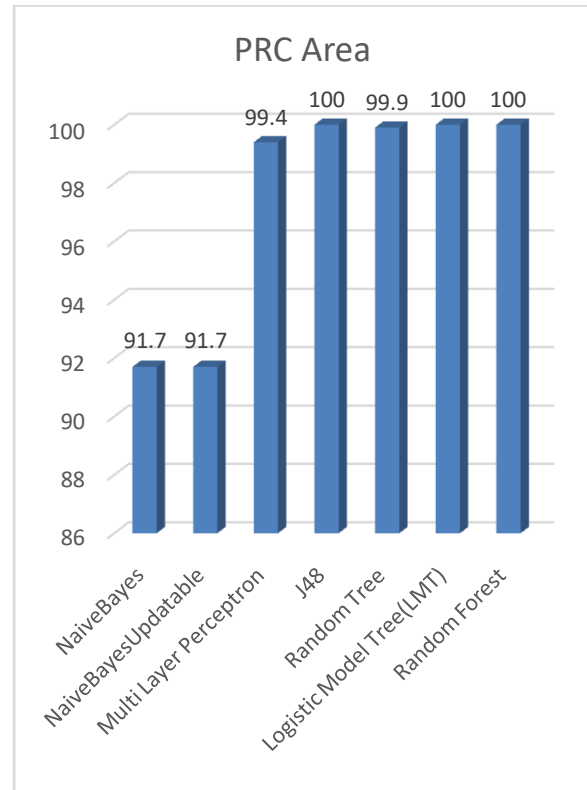
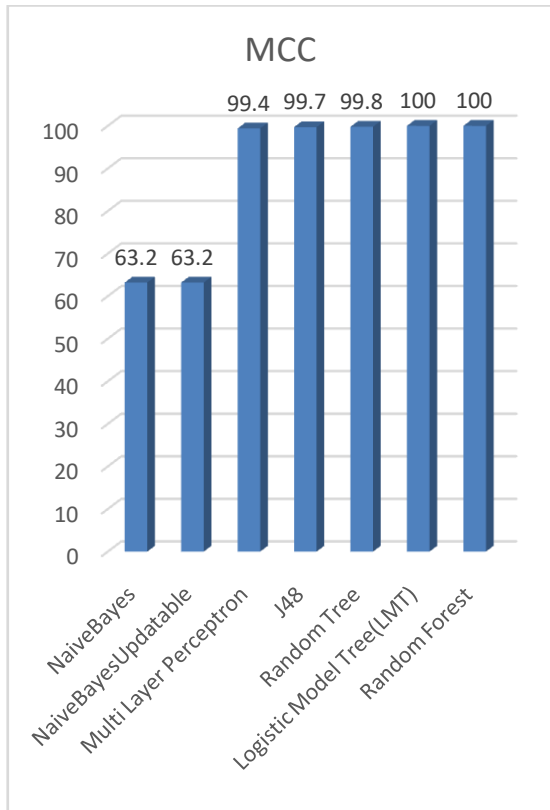
<https://www.kaggle.com/easonlai/sample-insurance-claim-prediction-dataset#insurance2.csv> which has 8 attributes/features with 1240 samples.

**Results Analysis:** To evaluate the performance of various machine learning algorithms (*Naïve Bayes Updatable, Naïve Bayes, Multi-Layer Perceptron, J48, Random Tree, Logistic Model Tree (LMT), Random Forest*), following metrics were used: *Precision, Recall, F-Measure, MCC, ROC Area, PRC Area*.

Logistic Model Tree (LMT), Random Forest algorithms have given better results when compared with Naïve Bayes Updatable, Naïve Bayes, Multi-Layer Perceptron, J48, and Random Tree Algorithms.







## 8. CHALLENGES IN IMPLEMENTING MACHINE LEARNING:

Most insurers recognize the value of machine learning in driving better decision-making and streamlining business processes. Research for the Accenture Technology Vision 2018 shows that more than 90 percent of insurers are using, plan to use or considering using machine learning or AI in the claims or underwriting process.

Some of the challenges insurers typically encounter when adopting machine learning are

### 8.1. Training requirements

AI-powered intellectual systems must be trained in a domain, e.g., claims or billing for an insurer. This requires a separate training system, which insurers find hard to provide for training the AI model. Models need to be trained with huge volumes of documents/transactions to cover all possible scenarios.

### 8.2. Right data source

The quality of data used to train predictive models is equally important as the quantity, in the case of machine learning. The datasets need to be representative and balanced so that they can give a better picture and avoid bias. This is important to train predictive models. Generally, insurers struggle to provide relevant data for training AI models.

### 8.3. Difficulty in predicting returns

It's not very easy to predict improvements that machine learning can bring to a project. For example, it's not easy to plan or budget a project using machine learning, as the funding needs may vary during the project, based on the

findings. Therefore, it is almost impossible to predict the return on investment. This makes it hard to get everyone on board the concept and invest in it.

### 8.4. Data security

The huge amount of data used for machine learning algorithms has created an additional security risk for insurance companies. With such an increase in collected data and connectivity among applications, there is a risk of data leaks and security breaches. A security incident could lead to personal information falling into the wrong hands. This creates fear in the minds of insurers.

## 9. CONCLUSIONS:

Modern technologies are moving extremely fast making their ways into various fields of the business. In this respects, the insurance industry does not lack behind the others. The application of statistics in the insurance has a long history. Thus, the fact that insurance companies are actively using data science analytics is not amazing. In essence, the aim of applying data science analytics in the insurance is the same as in the other industries—to optimize marketing strategies, to improve the business, to enhance the income, and to reduce costs. In this paper, we presented several machine learning techniques to analysis the insurance claims efficiently and compare their performances using various metrics.

## REFERENCES

1. Belhadji, E., G. Dionne, and F. Tarkhani, “A Model for the Detection of Insurance Fraud, Geneva Papers on Risk and Insurance Theory”, 25: 517-538, may 2012.
2. Crocker, K. J., and S. Tennyson,” Insurance Fraud and Optimal Claims Settlement Strategies: An Empirical Investigation of Liability Insurance Settlements” The Journal of Law and Economics, 45(2), April 2010.
3. Kajiamuller, “The Identification of Insurance Fraud – an Empirical Analysis Working papers on Risk Management and Insurance” no: 137, June 2013.
4. S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” Informaticavol 31, pp 249-268,May 2011.
5. Sivarajah U, Kamal M, Irani Z, Weerakkody V (2017) Critical analysis of big data challenges and analytical methods. J Bus Res 70:263–286
6. Mishr K (2016) Fundamentals of life insurance theories and applications. In: 2nd ed, Delhi: PHI Learning Pvt Ltd
7. The Kaggle Website. [Online]. <https://www.kaggle.com/c/prudential-life-insurance-assessment/data/>
8. The Accenture website <https://www.accenture.com>.