Artificial Intelligence Based Agentless Financial Model

Krishanu Prabha Sinha
Graduate Student
College of Science and Engineering
University Of Houston Clear Lake

Mehdi Sookhak
Assistant Professor
School of Information Technology
Illinois State University

Abstract— Since past couple of years, Agents have been a crucial part of the financial sector, primarily focusing on the Auto Insurance sector, whose key responsibilities are centered around finding new prospective customers and maintaining a relationship with existing customers. But with every other company streamlining their business processes with the latest Technology, Insurance Industry is not too far behind. Currently, Insurance Industry has dived and started exploring the online space. Prospective customers can now get online insurance quotes, chat with an online robot and even purchase an Insurance policy online. Digitalization, Automation, and Streamlining are key buzzwords in every type of business sector. Given the above trends, Insurance Agents seem to be an unnecessary expense. In this paper, we propose an Artificial-Intelligence driven approach that eliminates the need for a human Insurance Agent that will ultimately reduce the overall cost for the end customer. As part of our contribution to the above problem statement, we have proposed a Software Application where four Statistical Models are deployed. These Models are tasked with determining prospective customers who will likely buy an Insurance Policy, identifying customers who are likely to cancel a policy so that we can provide them with something better, identifying customers submitting fraudulent insurance claims and finally a Recommendation System Model to recommend updates to current policy to existing policy of Customers. In our Experimentation Results, we identified a cluster of customers who were most likely to buy a product using an Unsupervised Statistical Machine Learning model.

Keywords— Artificial-Intelligence driven, Insurance Agent, Digitalization, Streamlining, Automation, Statistical Machine Learning Model.

I. INTRODUCTION

To gain a deeper understanding of the Agent-Based Model we must understand what role Agents play in it. To gain a clearer picture of the responsibility of an Insurance Agent, we conducted informal interviews and also studied the job duties of an Insurance Agent posted in various online job portals. We found that currently most Insurance Agents are involved in finding prospective new customers and also maintaining relationships with the existing set of customers. Typical strategies used by Insurance Agents to find new Customers include a referral process as part of which the Agent requests their current set of customers to refer them to newer and likely customers. Cold Calling, which refers to making unsolicited calls to someone, usually through telephone or in-person, to sell policies is another strategy that is commonly used. Insurance Agents are often given the authority to sell and, sometimes, adjust the coverage on behalf of one or more Insurance Carriers, that he or she may represent. This is usually done to maximize the sale and profit of an Insurance Carrier. Agents can either be dedicated to a single Insurance Carrier or Independent. Dedicated Agents represent a single Insurance Carrier whereas Independent Agents represent multiple carriers. All agents operate on behalf of the Insurance Carrier and not on the Customer. Their main aim is to match the customer with coverage from Insurance Carrier. They usually earn a stable commission from the Insurance carrier they are contracted with. In some cases, Dedicated Agents are salaried [1].

Many insurance carriers have traditionally sold their products through local Insurance Agents. Around a decade ago, 80 percent of personal auto policies were put with an agent (the figure for homeowners and small commercial policies was closer to 100). For insurance carriers, these were simpler times. They focused on the strength of their actuarial pricing models, their ability to quickly return a quote and manage claims and their agency force consistency. Distribution management discussions focused on increasing the Agency's efficiency, making it easier for them to do business and creating incentives to reward their entrepreneurial zeal [2].

Agents used to be involved in the claim registration and handling process in the past, and also served as the primary source of contact for different customers when seeking assistance. However, with huge investments made by Insurance Carriers, in harnessing the power of data and technology over the years, massive Advertising Campaigns and simplification of the claim registration and handling process (most of which can be easily handled using a Mobile Phone Application), companies have finally reached a state where substantial number of customers find it less cumbersome to take matters in their own hands and directly deal with the carrier rather than relying on an Agent for a claim. This has typically been the case for bigger Insurance Carriers who hold the majority of the Insurance Market Share (for example Geico) in the United States and still have Agents representing them in the current moment. The Insurance

Carriers who are entering the market in the present (for example Root) are heavily armed with Artificial Intelligence and some of the latest technology available to do the usual Insurance-related work. For Example, Root Insurance a recent player in the Insurance Market reaches out and acquires new customers through Targeted Advertisements in various mediums without any human agent involvement. This changing dynamic in the Insurance Market with the addition of newer players who are armed with more technically advanced and less labor-intensive ways for doing the same work that the bigger players have traditionally got done utilizing surplus human capital that resulted in bigger spending is leading them to reconsider their dependence on Agents. Dependence on Insurance Agents has proved to be an expensive affair for most old carriers who are driving operational costs upwards forcing them reconsider. As a result of the expensive operational cost, customers of Agent Driven Insurance Carrier has to pay a higher monthly premium for a service that could have been significantly lower with an Agentless Insurance Carrier [2].

To compete with the newer players, bigger carriers have increased their investments to reduce their dependency on Agents. Going by some of the current trends in terms of Marketing expenditure to reduce the dependence on Insurance Agents, the McKinsey report of 2013, states that the average spending of bigger insurance carrier has only seen an increase from 1.7 billion dollars in 2002 to 5.9 billion dollars in 2011 and has seen a steeper increase in 2019 [2].

Our main contribution in this paper is to propose a software application that can remove the need for an insurance agent to help the larger carriers minimize their operating costs and thereby remain competitive by reducing their customer base's total monthly premiums. Our proposed software application shall have four highly trained Statistical Machine Learning or Deep Learning Algorithms trained and deployed to accomplish four specific tasks:

a) Predict the likelihood of certain customers to buy Insurance Policy. Specific predictors such as Monthly Income, Marital Status, Credit Score, Recent Automobile Purchase can be very helpful in determining how likely a customer will buy certain Insurance Policy. Data can be collected from advertisements, visitors to websites and google searches. This will allow us to narrow down our problem to a few targets. The targeted customers can then be attracted to buy a policy through specifically designed offers.

b) Determining the customers that are likely to leave. Type of Policy Holder, Average Monthly Premium, feedbacks, and Engagement Time can be indicative of a particular customer leaving our policy. The targets identified from this method will narrow our focus to specific customers that are likely to cancel our policy. Such customers can be engaged with special discount offers that can potentially help them reconsider canceling our policy.

c) Determining the fraudulent claims filed by customers. Fraudulent claims cost the company huge losses. Fraud history, Coverage Amount and Credit Score can be highly indicative of determining whether an Insurance claim is Fraudulent or not. Filtering out fraudulent claims can help save on operational costs and increase profitability.

d) Recommend new and updated policies to customers. For engaging our customers constantly and selling new and updated policies, certain eligibility criteria must be defined to determine eligible customers.

The below table summarizes the contributions that is relative to what currently many Insurance Carriers already possess:

	Human Agents	AI Based Software Architecture System that specifically tries to automate the responsibilities of an Agent in the Insurance Industry
Root Insurance	YES	NO
GIECO Insurance	NO	NO
American Family Insurance	NO	NO
Our Proposed Agentless Model	**YES**	**YES**

Table 1: Contributions of our proposed Agentless Model

Organization. The rest of the paper is organized as follows. Section 2 presents an overview the influence of Artificial Intelligence in Insurance and Financial Sector. Section 3 discusses the current architecture of the Agent Based Model. Section 4 discusses our contribution in detail as well as the pseudo code implementations. Section 5 discusses the Experimentation Results and Metrics Used. Section 6 discusses key challenges of substituting an Agent with Agentless Model. Finally, we conclude the paper in Section 7.

II. RELATED WORK

This section reviews the influence and usage of Artificial Intelligence in financial and insurance domain with focus on specific merits and demerits.

Yamasaki et al [14] discuss the application of artificial intelligence in Japan's health sector that can largely influence life and death matters. This paper proposes to build an artificial intelligence application to influence the general public health-related habits and further discuss its viability using a case study. This paper addresses the privacy and existing ethical concerns about the use of artificial intelligence in a patient's health data. The healthcare AI market is projected to grow at 39.4% CAGR (Compound Annual Growth Rate) and generate ten Billion in worldwide revenue by 2024. One of the high-value applications of AI would be Robot-Assisted Surgery, which has the highest potential. Not only will it bring down the cost of surgery but also increase the patient's overall health. All the above projections make this paper highly valuable. However, this paper does not discuss the obstacles surrounding the adoption of AI in healthcare such as:

a) The integration of healthcare data is complex, which results in missing disparate data and a lack of curated data.

b) Talent Shortage and High Initial Capital Invest- specific skills and knowledge are needed to succeed with AI which is certainly lacking currently [25].

Zhang et al [15] discuss the enthusiasm surrounding artificial intelligence in the Information Technology (IT) sector. This paper briefly discusses the emerging trends in artificial intelligence for a wide range of business needs such as chatbots as a substitute for customer care, fraud detection, AI in insurance. This paper is highly informative about the utility of AI in finance. However, the paper only scratches the surface concerning each topic it covers and abstains from providing many details. Even from the perspective of an informative research paper highlighting the core areas in finance where AI is playing a major role, this paper does not discuss challenges in the adoption of Artificial Intelligence. Challenges surrounding the AI in Finance include high cost involved in production and maintenance, centralization of power in the hands of few controlling it, unemployment due to replacement of workforce [26].

Abdelrahman et al [16] discuss an approach of detecting driver behavior and subsequently providing a measure of the actual driving risk score. Two machine learning algorithms, which are support vector regression (SVR) and decision tree regression (DTR) are trained to reflect the driver's score. The paper proposes to develop an artificial intelligence mobile phone application that makes use of high-frequency sensor data from instruments that already exist in all smartphones such as Global Positioning System (GPS), Accelerometer, Gyroscope, Magnetometer. Apart from the above factors, the customer's age, the type of vehicle (such as make, model, year, etc.) can be used to determine the monthly premium. However, this proposal has already been implemented by a few insurance carriers such as Root Insurance. This paper does not propose any new predictor in the already known feature list [27].

Bellomarini et al [17] propose the usage of knowledge graphs as the reference technology for enterprise AI to govern the proliferation of smart AI-driven applications. The primary contribution of this paper is to propose software architecture principles for governing the proliferation of AI-driven applications. The paper further focusses on the Vadalog system, a successful knowledge graph middleware to show knowledge graph in action. However, the proliferation of smart AI-driven is potentially distant possibility and not an existing problem.

Brigo et al [18] discuss the applications of Robotics, AI to financial services and insurance in particular. Concerning artificial intelligence (AI) and cognitive computing, in particular, the author discusses extremely advanced artificial intelligence (AI) applications such as an AI technology that has achieved human-level interaction capabilities. This paper states that such technology can act as virtual brokers for tailored life and car insurance policies. The author argues that only when we achieve this level of sophistication, we would be able to leverage the complete power of artificial intelligence. This paper only discusses distant possibilities in an age when AI has achieved highly advanced capabilities. However, this paper does not discuss a key challenge surrounding AI in the Insurance sector. Insurance is a highly regulated sector and this gives rise to risk profiling of customers. Insurance is a business deeply rooted in understanding the risk profile of customers. In our paper, we discuss how using certain variables one can identify a fraudulent claim. Once a customer filing a fraudulent claim is identified, the risk score associated with such customers must be accordingly updated [28].

Linthicum et al [19] discuss the applications that will benefit the most from AI capabilities such as fraud detection, predictive marketing. However, this paper does not discuss the immediate challenges that one will face when such applications replace human counterparts. In our paper, we propose to leverage customer churning model, fraudulent claims, propensity models and recommender system models to eliminate the role of an Agent. We also discussed the potential challenges and business impacts of substituting an agent-based model with an agentless model.

Nguyen et al [20] argue that most recent papers, within the topic of AI in Healthcare, focus on developing prediction algorithms that can identify symptoms for patient’s diseases' such as cancer, cardiology, breast cancer. However, there are rarely research papers that train patients and doctors to know more about their surgery or diseases. This paper proposes to build an AI and Block chain mobile application that has three discrete AI assistants. Firstly, AI as an advisor inspects free appointments from available time for doctors and patients. Second, AI advises the physician's schedule of upcoming surgery list and preparing patients before surgery. Third, under the doctor's supervision, deep learning AI analyzes the health care history of patients and recommends the options for patients following surgery. The users of this proposed mobile application will not only be doctors but also patients, drug companies, insurance companies, and hospitals. This paper sets a very ambitious and interesting target for the application of AI in healthcare. However, a common application that has users with such different domain expertise seems to be very challenging. Such an application would require a very high capital in the initial phase of development and its success cannot be guaranteed. Comparatively, our proposed application would not be as expensive since most machine learning models are prebuilt. We only need to customize these prebuilt models to fit in our use case.

Kang et al [21] address the issue of training a statistical model on an imbalanced or biased dataset using traditional

techniques. Such models end up being biased in their outlook since they don't encounter minority class in the training dataset. This paper proposes the following steps to deal with the imbalance in the dataset. First, we use the Synthetic Minority Over-sampling technique to reject inference to employ imbalanced learning for the accepted applicant data. Second, we conduct rejection inference on the basis of a graph-based semi-supervised learning algorithm, called label propagation, for rejected applicants. Second, to train the combined training data, we use tree-based ensemble learning models as the base classifiers. Finally, we are giving an exact assessment experiment using a Chinese auto loan company's data. Comparatively, in our paper, while developing the fraud detection model we encounter an imbalanced dataset. To deal with the imbalance in our dataset we focus on the following approaches:
a) Oversampling — **SMOTE**
b) Under-sampling — **RandomUnderSampler**
c) Combined Class Methods — **SMOTE + ENN**

Begel et al [22] propose a nine-stage AI workflow propose which they have acquired through prior experiences while developing AI applications at Microsoft. They further emphasize discuss data science tools specific to Microsoft. However, the development tools and environment available are very specific to Microsoft's internal environment is not what developers and data scientists at other organizations have available to them. In our paper, we propose to develop, train and deploy our AI application at Amazon Web Services (AWS) Sage maker which is much more pervasive.

Sandu et al [23] discuss the overall emphasis on the acceptance and push to develop more AI-based applications across all domains. This paper emphasizes the importance of Chatbots. The author states that Chatbots can revolutionize the education industry by replacing a human teacher. Further, the author proposes a technique to determine the factors that can lead to the acceptance of Chatbots as a substitute for teachers. A major flaw in this paper is that the entire paper is written based on the assumption that Chatbots can entirely replace human teachers without any argument. However, in our paper, we clearly understood the responsibilities of an insurance agent and then suggested AI models that could effectively replace an agent. Further, we discussed the immediate business impacts that would have when replacing an agent-based system with an agentless system.

To the best of our knowledge, this is the first attempt to use AI to develop an agent-less insurance system. Based on our understanding AI can help larger insurance carriers mitigate the issue of overhead expenditure and provide better and cheaper services to their existing and new customers.

III. ARCHITECTURE

This section provides an overview of the architecture of an Agent Based Model and discusses its demerits.

How the Agent Based Architecture looks right now?

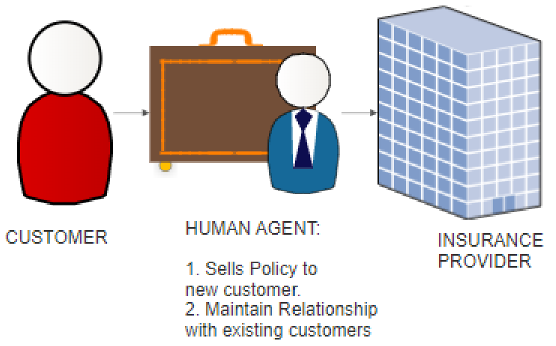

Fig. 1. Agent based Insurance Architecture

Why should we go towards an Agentless Model?

A. *Advancement of Predictive Modelling and Information Technology:*

According to the Mckinsey Report entitled "Agents of the Future: The Evolution of Property and Casualty Insurance Distribution," there are several reasons that cast a shadow over the very role of an insurance agent in the current Insurance Business Model with the advances in predictive modeling and IT. Perhaps most disruptive of the traditional agent-value model, auto insurance–which accounts for 70 percent of personal line premiums–is rapidly becoming commodified [2].

B. *Availability of Multiple Channels:*

While customers previously expected to shop with their insurer for insurance and file claims and seek answers to their questions in the same manner, today they are increasingly preferring to connect with their insurance provider on their schedules at all times and through multiple channels (e.g., mobile, online self-service, click-to-chat). Also, customers expect a reliable, enjoyable experience every time they connect. As a result, the walls are collapsing between conventional distribution channels. This is most evident in personal cars, where recent efforts by McKinsey to map the customer decision travel reveal the extent to which shoppers jump from one channel to another as they move from information collection to purchase and beyond [2].

In light of this growing channel of "agnosticism," carriers are making investments in heavy technology to reduce customers ' and data friction as they switch between channels. We open central contact centers for dealing with queries and making transactions for policy changes, billing, claims, and all other communications. Carriers will soon be able to mimic the experience of the office of an agent–with greater consistency and willingness to test new strategies (e.g., segment-based service-to-sale scripts, "next product to purchase" or "likelihood to move" analytics). As more customer experiences flow through these central contact centers, there will be less incentive for agents to connect with the customer and add value [2].

C. *Shrinking agent force:*

Only a subset of current agents make a successful transition. The total number of agents competing to capture a share of the available "commission pools" has declined by about 10 per cent since 1995. If any of these pools diminish without offsetting growth in other business lines, it is likely to exacerbate the decline in agent counts. Many factors are likely to cause a drop in car insurance commissions–the largest pool at about $12 billion per annum such as:

• Total auto gross written premium (GWP) is expected to grow slowly at best or even stagnate or decline with increased usage-based insurance adoption, continued market competition, the downward trend in accidents and increased car safety.

• Direct channel's market share of auto GWP (where few if any commissions are paid) continues to grow at the expense of the agency channel.

• Carriers must continue to spend on ads, technology and physical infrastructure to satisfy consumer demands for multi-channel capabilities at the rates available to paid agents as commissions [2].

D. *Scope of Automation in the area of Insurance:*

The affordances of AI applied to the insurance industry can be claimed to magnify profitability and drastically cut costs. The robe-life agent can all gain the opportunity to source and develop life insurance portfolios, promote underwriting, and track policies. Such an approach would often be considerably more effective, a way to deliver superior solutions and considerably less costly [3].

According to the 2017 McKinsey Global Institute report, insurance and finance industries have 43 percent of automation potential. By 2025, up to 25 percent of the insurance industry task force may be consolidated or replaced, particularly in operations and administrative support. While AI, Machine Learning & Cognitive Systems are combined with Robotic Process Automation (RPA) to put efficiencies into existing processes and rising operational costs, the insurance industry is now actively looking at use cases for intelligent automation [4].

DIFFERENCE BETWEEN AGENT AND AGENTLESS ARCHITECTURES

	Agentless Based Model	**Agent Based Model**
Agent Involvement:	No Agents Involved	Agents Involved
Techniques to identify new customer:	Retrieves data from various Vendors and then applies techniques to find how likely they are to buy Insurance Policy.	Using Cold Calling and other human techniques
Techniques to recommend policies to new customer	Using Recommendation Algorithms	Using Human Approach
Techniques to maintain relationship with existing customers	Uses Customer Churning Classification Techniques to identify customers who might leave	Using Human Approach

Table 2- Agent vs Agentless

IV. OUR CONTRIBUTION

We propose to build a Software-based Application that can have a set of four Machine Learning Models deployed in them that do the exact set of activities that an Insurance Agent thereby reducing or eliminating the dependency on a Human Agent. Below are the simple set of responsibilities that we have identified for automation using our proposed Agentless Model:

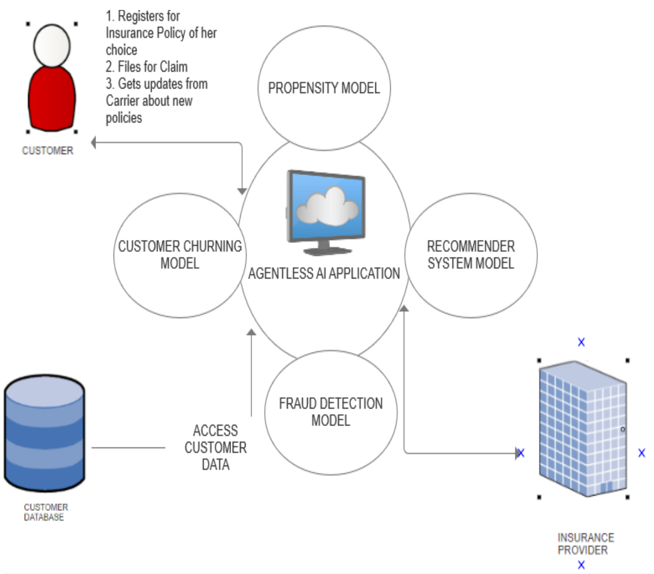

Fig. 2. Agentless Insurance Architecture

A. *Finding new Customers Using Propensity Models:*

1. What is a Propensity Model?

Propensity modeling aims to determine the probability that some acts such as purchasing your product will be undertaken by travelers, leads, and clients. It is a mathematical method that accounts for all of the independent and conflicting factors influencing said behavior.

So, for example, propensity modeling can help a marketing team predict the likelihood that a lead will convert to a customer.

The obtained propensity score from a propensity model is the probability that the visitor, lead, or customer will perform a certain action.

Following steps should summarize the steps to build a Propensity Model:

- Selecting your features;
- Constructing your propensity model;
- Calculating your propensity scores.

We can use a highly interpretable Regression Algorithm such as Logistic Regression to predict the likelihood of a new customer buying Insurance policy.

How we can leverage it in our business problem?

In our proposed Agentless Model, the Insurance carrier first gets/purchases data about potential customers of its carrier. Relevant fields about the customer, such as Annual Income and net worth of the Customer can be used in our Propensity Model to determine the Purchasing Power of the customer. Based on the prediction results obtained from the propensity model it can be determined whether to reach out to the customer or not. Hence we recommend deploying the Propensity model to predict new customers [5][8].

Pseudo Code Implementation:

Algorithm 1: Pseudocode description of propensity model to predict which the customer is likely to purchase.

Input: Income_level, Properties_Owned, Marital_Status, Policy_Premium_Amount.

Output: Predict which the customer is likely to purchase our Insurance Policy.

$$\ln\left(\frac{e(Xi)}{1-e(Xi)}\right) = \ln\left((\Pr(z=1)\,|Xi)\right) = \alpha + \beta(Xi)$$

where:
$e(Xi)$ = b0 + b1(Income_level) + b2(Properties_Owned) + b3(Marital_Status) + b4(Policy_Premium_Amount)

and
b0, is the intercept
bi, is the regression coefficient
Xi, the treatment variables and covariates (random variables)

We propose to deploy the above trained model in a Cloud Based Application.

B. *Customer Churning Model to determine which Customer is likely to leave*

2. What is the Customer Churning Model?

If consumers avoid doing business with a company, customer churn, also known as customer attrition, happens. Because the price of acquiring a new customer is usually higher than maintaining the old one, the companies are interested in defining parts of these customers. For example, if Root Insurance knew a group of customers at risk of churning, they might approach them proactively with special offers rather than simply losing them.

Choice can be between a linear and a non-linear model to determine whether a particular customer is about to churn. Some useful features that are normally considered while building a Customer Churning Model are:

- number of sessions before buying something,
- average time per session,
- time difference between sessions (frequent or less frequent customer)

How we can leverage a Churning Model in our business problem?

In our proposed Agentless Model, from the existing fields of the customer, we can determine which customer is likely to leave or cancel his/her policy using relevant fields such as type of policyholder, Income Level, Age of the customer in a Customer Churning Model. Various strategies can be devised to retain such customer base such as giving them a discount coupon. [6][9]

Pseudo Code Implementation:

Algorithm 2: Pseudocode description of applying ID3 Algorithm to identify customer attrition.

Input: Change in Consumption, Customer Care Calls, Age, Insurance Policy Price.

Output: Predict which the customer is likely to leave.

1. /* Compute the Entropy for Dataset. */

p=Count of Customers who left
n=Count of Customers who did not leave
Attribute_List = [‘Change in Consumption’, ‘Customer Care Calls’, ‘Age’, ‘Insurance Policy Price’]

$$Entropy = -(\frac{p}{p+n})\log 2\left(\frac{p}{p+n}\right) - \left(\frac{n}{p+n}\right) log2\left(\frac{n}{p+n}\right)$$

2. /*Repeat until we get desired tree */

while (Desired_Tree=True) {

2.1. /* For every Attribute/Feature */

for attribute in Attribute_List {

2.1.1. /*Calculate Entropy for all Attributes */

$$Entropy = -(\frac{p}{p+n})\log 2\left(\frac{p}{p+n}\right) - \left(\frac{n}{p+n}\right) log2\left(\frac{n}{p+n}\right)$$

2.1.2. /* Take Avg Info Entropy for the Current Attribute */

$$I(attribute) = \sum \frac{pi+ni}{p+n}(Entropy(attribute))$$

2.1.3 /* Calculate Gain for current Attribute */

Gain = Entropy(S) – I(attribute)
}

2.2 /* Pick the highest Gain Attribute. */
Max_Gain = Find max value of Gain
}

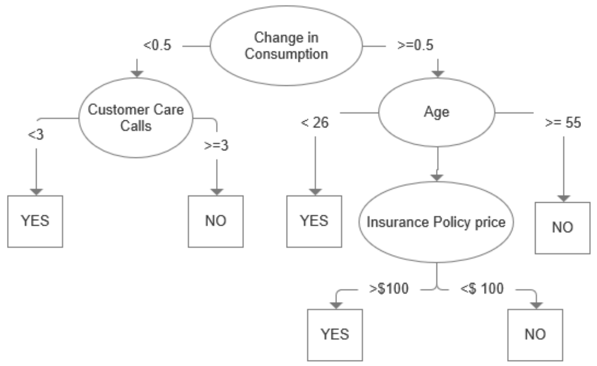

Fig. 3. An anticipated sample decision tree output from applying ID3 Algorithm on the feature set mentioned above

We propose to deploy the above trained model in a Cloud Based Application.

C. *Fraud Detection Model to determine financial fraud committed by a customer*

3. What is a Fraud Detection Model?

Fraud is a business worth a billion dollars, which develops every year. The 2018 PwC Global Economic Crime Survey found that half (49 percent) of the 7,200 surveyed businesses have experienced some sort of fraud. This is an improvement from the PwC 2016 report in which economic crime had been reported by slightly more than a third of the organizations surveyed (36 percent).

Opportunities for fraud co-evolve with technology, especially in Information technology. Business reengineering, reorganization or downsizing can weaken or eliminate control, while new information systems can offer additional opportunities to commit fraud.

The greatest challenge here that we have to deal with is Imbalanced Data i.e., the Majority Class vehemently outnumbers the Minority class (In most cases 99:1 ratio)

Techniques to deal with Imbalances in Data:

There are many ways of dealing with imbalanced data. We will focus on the following approaches:
Oversampling — **SMOTE**
Under sampling — **RandomUnderSampler**
Combined Class Methods — **SMOTE + ENN**

Which Statistical Machine Learning Model to use in such a case?
A non-linear Machine Learning Algorithm such as a Random Forest would be most favorable to use in this case .

How we can leverage a Fraud Detection Model in our business problem?

In our proposed Agentless Model, once a customer submits a claim, it can be determined if the customer committed fraud by using relevant fields pertaining to the customer such as previous customer history [7] [10] [11].

Pseudo Code Implementation:

Algorithm 3: Pseudocode description of the ensemble learning Algorithms based on feature subspace and counter voting rule for classifying imbalanced multiclass Fraud Detection Multiclass Data.

Input: Training set T; Feature set C= (SalesChannel, TransactionStartTime, Amount, FraudResult);
Size of feature space K; Size of feature subspace D;
Number of classes C; Number of feature subspace L;
Baseliner I ; One test sample x'

Output: h(x') which is the class label of the test sample x'.

```
(1) for i=1 to C {
(2)      Label the samples of the ith class as
         Positive and the rest samples as Negative;
(3)      External L diverse training subsets by
         feature subspaces generation algorithm;
(4)      for j=1 to L {
(5)          Train imbalance base classifier Iij
             by training subset Tij using Decision Tree,
             Random Forest, ANN and Naïve Bayes,
             respectively.
                 }
(6)   for i=1 to C {
(7)          for j=1 to L {
(8)                  Use Iij to classify the test
                     sample x';
(9)          Calculate the value of Counter_i
                     }
(10)  Output h(x') by 10  }
```

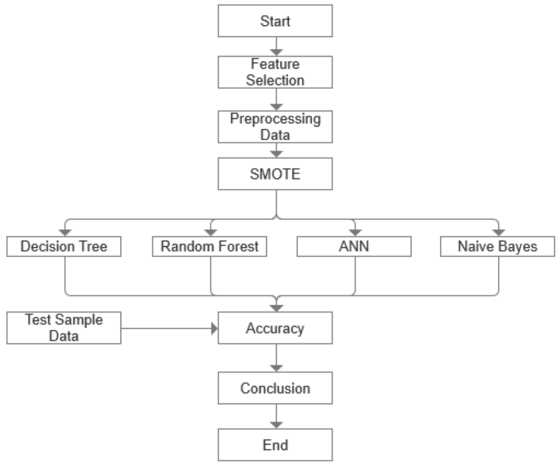

Fig. 4. Set of steps ranging from Feature Selection followed by Preprocessing Data to applying set of classifiers on Training Data.

D. Recommend Policies to Customer:

4. What is Collaborative Filtering for a recommendation system?

The central idea here is that in the future, the users who have agreed in the past tend to agree too. It usually expressed in two types, in terms of user preference. Explicit Rating is a score on a sliding scale given by a consumer to an object, like 5 stars for Titanic. This is the users ' directs input to demonstrate how much they like an object. Implicit ranking, implicitly indicates user choice, such as page views, clicks, buy records, whether or not to listen to a music track, and so on.

How to leverage this technique in our current problem?
Various Recommendation Algorithms such as content and Collaborative Filtering can be deployed to recommend suitable policies to eligible and potential customers.
Using the above deployed Algorithms, we have presented the high level architecture of an operational Insurance Model that has totally automated the tasks of an Insurance Agent thereby securing the Insurance Carriers a way to deal with the upcoming challenges and reducing the dependency of Insurance Agent [12].

Pseudo Code Implementation:

Algorithm 4: Pseudocode description for Identifying Similarity between two customers using Jaccard Similarity Coefficient for Recommendation Model.

Input : InsurancePolicyID, CustomerID, CustomerFeedback

Output: Jaccard Similarity Score between every pair of customers

A = Customers who subscribed to Policy 1 (P1)
B = Customers who subscribed to Policy 2 (P2)

/*We compute the Jaccard Similarity between two customers*/
/* The Jaccard coefficient measures similarity between two customers.*/

$$J(\mathrm{A}, \mathrm{B}) = \frac{|A \ \cap \mathrm{B}|}{|A \ U \ B|}$$

($0 < J(\mathrm{A}, \mathrm{B}) \leq 1$)

(If A and B are both empty, define J(A, B) = 1)

Recommender System for a Bipartite network:

This paper proposes a recommender system for bipartite network as demonstrated in the below figure 5(a). There are no specific criteria for a bipartite graph. Nodes in our bipartite network for Insurance Sector can be divided in two sets – Insurance Customers and Insurance policies. While there are connections between two nodes from different node sets, there is no direct connection between each set. The connections can be any action between customers and Insurance Policies, such as: buying, viewing, and rating [29].

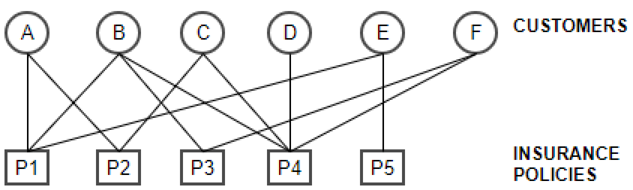

Fig.5(a). Bipartite Projection

The only way to interpret and study the above bipartite graph is to split it by its unipartite properties such as customers, Insurance policies and transactions from Bipartite Projection (figure 5(a)), into two individual unipartite graphs- Customers and Insurance Policies [29].

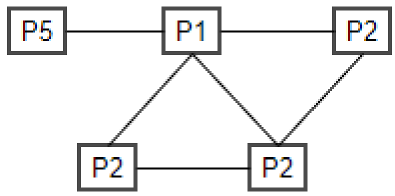

Fig.5(b) Unipartite Projection of Insurance Policies

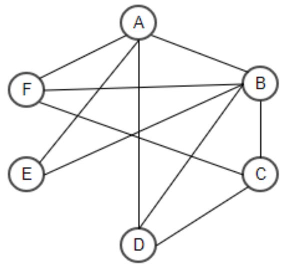

Fig.5(c) Unipartite Projection of Customer Similarities

Fig.5(b) and Fig.5(c) represent unipartite projected graphs. Once two nodes in Set V (or U) connect with one or more nodes in Set U (or V), then we successfully connected two nodes in the projected graphs. With this arrangement, we can calculate certain metrics. For Example: a. The Clustering Coefficient is 0.5 and density is 0.67 etc [29].

V. CODE IMPLEMENTATION AND METRICS

In order to demonstrate the feasibility of developing a statistical model that determines or groups prospective customers (that are likely to buy our product) from available sales data from a Mall, we have developed an Unsupervised Machine Learning Model based on available fields such as Customer ID, age, gender, annual income and spending score. Spending Score is something that has been assigned to the customer based on parameters like customer behavior and purchasing data. Since in case of an Unsupervised Machine Learning Algorithm, there is no feedback mechanism no metric has been used. Below the customers groups have been plotted based on their likeliness or probability to buy a specific product [24].

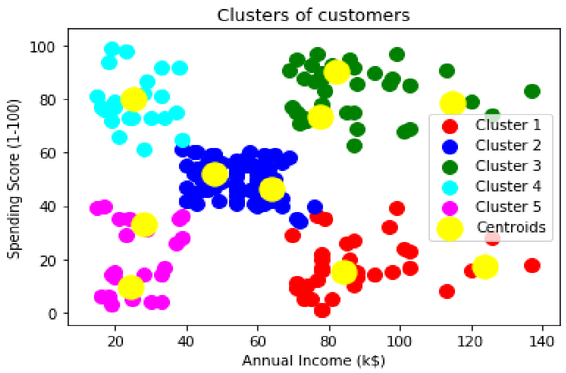

Fig.6. Output of Kmeans Clustering Algorithm on Sales Dataset from Mall

Interpretation of the Model:

Cluster 1 (Red Color) -> earning high but spending less
Cluster 2 (Blue Color) -> average in terms of earning and spending
Cluster 3 (Green Color) -> earning high and also spending high [TARGET SET]
Cluster 4 (cyan Color) -> earning less but spending more
Cluster 5 (magenta Color) -> Earning less, spending less
We can put Cluster 3 into some alerting system where email can be send to them on daily basis as these re easy to converse wherein others we can set like once in a week or once in a month.

In another experiment performed on the same dataset, we first try to segregate high spending customers from low spending customers. For this purpose, we have applied the Random Forest Classification Algorithm. We have separated the customers based on spending scores and applied a label for each customer. We have used Precision, Recall, F1 Score and Confusion Matrix as a feedback mechanism for our machine learning model. The following are the results of our experiment.

```
Random Forrest Accuracy: 100.0
Mean Absoulte Error: 0.0
Classification Report:
             precision    recall  f1-score   support

          0       1.00      1.00      1.00        30
          1       1.00      1.00      1.00        20

avg / total       1.00      1.00      1.00        50

Confusion Matrix:
         predicted
                0   1
actual 0       30   0
       1        0  20
```

VI. CHALLENGES OF AN AGENTLESS MODEL:

Our proposed agentless model in Insurance does come with its fair share of flaws. If agents are eliminated, the consumer will face an array of problems. A subset of those problems are listed below:
Who shall address the Insurance Questions of the customer?
How will the customer know that they have the right kind of insurance?
Who advocates for the consumer when an insurance company denies a claim?
Who helps the consumer make changes as life changes?
The agent is not just a sale representative. An agent is the partner of the client. An agent is responsible for ensuring the customer has the correct coverage, the right policy terms, and the right price. The agent takes the time to consider what the consumer needs and the right insurance company suit them [13].

VII. CONCLUSION

While many may argue that Agent is here to stay, we don't dispute the fact that humans like to deal with humans, especially something that impacts a large part of our life such as Insurance for their favorite car. But if the trend is anything to go by then technology is paving its way into every aspect of our lives and customers are getting accustomed to it. In the foreseeable future, the millennial generation might be making big purchases without interacting with any human. Disruptive technologies such as Artificial Intelligence are driving changes very fast and the most intelligent thing to do would be to streamline, digitalize and improve all processes before they become bottlenecks and are exploited by competitors.

REFERENCES

[1] Lindeman, J. (2018). Insurance brokers vs agents: Key differences explained.PacificPrime.https://www.pacificprime.com/blog/insurance-brokers-vs-agents.html

[2] McGranahan, D., & Gaeta, G. (2013). Agents of the Future: The Evolution of Property and Casualty Insurance distribution https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/financial services/latest thinking/insurance/agents_of_the_future_the_evolution_of_property_and_casualty_insurance_distribution.ashx

[3] Furness, D. (2019). Today's AI May Replace Tomorrow's Life Insurance Agent. Emerj. https://emerj.com/ai-future-outlook/todays-ai-may-replace-tomorrows-life-insurance-agent/

[4] Shiv, N. (2018). How Intelligent Automation is Changing the Shape of Insurance Industry in Digital Age. MindTree A Larsen and Toubro Group Company. https://www.mindtree.com/blog/how-intelligent-automation-changing-shape-insurance-industry-digital-age

[5] Artun, Ömer Levin, D. (2015). How to Use Predictive Purchase Behavior Modeling to Understand Consumers. Hubspot. https://blog.hubspot.com/agency/predictive-analytics-buy

[6] Tausend, F. (2019). Hands-on: Predict Customer Churn. Towards Data Science. https://towardsdatascience.com/hands-on-predict-customer-churn-5c2a42806266

[7] Karaman, B. (2019). Predicting Next Purchase Day. Towards Data Science. https://towardsdatascience.com/predicting-next-purchase-day-15fae5548027

[8] Mullin, S. (2019). Propensity Modeling: Using Data (and Expertise) to Predict Behavior. CXL. https://cxl.com/blog/propensity-modeling/

[9] Roman, O. (2019). Churn prediction. Towards Data Science. https://towardsdatascience.com/churn-prediction-770d6cb582a5

[10] Pierre, R. (2018). Detecting Financial Fraud Using Machine Learning: Winning the War Against Imbalanced Data. Towards Data Science. https://towardsdatascience.com/detecting-financial-fraud-using-machine-learning-three-ways-of-winning-the-war-against-imbalanced-a03f8815cce9

[11] Wikimedia Foundation. (2010). Data analysis techniques for fraud detection.InWikipedia.Wikipedia. https://en.wikipedia.org/wiki/Data_analysis_techniques_for_fraud_detection

[12] Luo, S. (2018). Introduction to Recommender System. Towards Data Science. https://towardsdatascience.com/intro-to-recommender-system-collaborative-filtering-64a238194a26

[13] Hockaday, J. (2018). Are Insurance Agents Really Necessary? New Horizons Insurance Marketing. https://blog.newhorizonsmktg.com/are-insurance-agents-really-necessary

[14] Yamasaki, K., & Hosoya, R. (2018, October 4). Resolving asymmetry of medical information by using AI: Japanese people's change behavior by technology-driven innovation for Japanese health insurance. PICMET 2018 - Portland International Conference on Management of Engineering and Technology: Managing Technological Entrepreneurship: The Engine for Economic Growth, Proceedings. https://doi.org/10.23919/PICMET.2018.8481824

[15] Zhang, X. P. S., & Kedmey, D. (2018). A Budding Romance: Finance and AI. IEEE MultiMedia, 25(4), 79–83. https://doi.org/10.1109/MMUL.2018.2875858

[16] Abdelrahman, A., Hassanein, H. S., & Abu-Ali, N. (2018). Data-driven Robust Scoring Approach for Driver Profiling Applications. 2018 IEEE Global Communications Conference (GLOBECOM), 1–6. https://doi.org/10.1109/GLOCOM.2018.8647971

[17] Bellomarini, L., Fakhoury, D., Gottlob, G., & Sallinger, E. (2019). Knowledge Graphs and Enterprise AI: The Promise of an Enabling Technology. 2019 IEEE 35th International Conference on Data Engineering (ICDE), 26–37. https://doi.org/10.1109/ICDE.2019.00011

[18] Lamberton, Chris; Brigo, Damiano; Hoy, D. (2017). Impact of Robotics, RPA and AI on the Insurance Industry: Challenges and Opportunities. Journal of Financial Perspectives, Vol. 4, No. 1, May 2017, 13. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3079495

[19] Linthicum, D. S. (2017). Making Sense of AI in Public Clouds. IEEE Cloud Computing (Volume: 4 , Issue: 6 , November/December 2017), 70–72. https://doi.org/10.1109/MCC.2018.1081067

[20] Le Nguyen, T., & Do, T. T. H. (2019). Artificial Intelligence in Healthcare: A New Technology Benefit for Both Patients and Doctors. 2019 Portland International Conference on Management of Engineering and Technology (PICMET), 1–15. https://doi.org/10.23919/PICMET.2019.8893884

[21] Kang, Y., Cui, R., Deng, J., & Jia, N. (2019). A novel credit scoring framework for auto loan using an imbalanced-learning-based reject inference. 2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), 1–8. https://doi.org/10.1109/CIFEr.2019.8759110

[22] Begel, A. (2019). Best Practices for Engineering AI-Infused Applications: Lessons Learned from Microsoft Teams. 2019 IEEE/ACM Joint 7th International Workshop on Conducting Empirical Studies in Industry (CESI) and 6th International Workshop on Software Engineering Research and Industrial Practice (SER&IP), 1–1. https://doi.org/10.1109/CESSER-IP.2019.00008

[23] Sandu, N., & Gide, E. (2019). Adoption of AI-Chatbots to Enhance Student Learning Experience in Higher Education in India. 2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET), 1–5. https://doi.org/10.1109/ITHET46829.2019.8937382

[24] Choudhary, V. (2018, December). Mall Customer Segmentation Data, Version 1. Retrieved January 01, 2020 from https://www.kaggle.com/data/46091

[25] Le Nguyen, T., & Do, T. T. H. (2019). Artificial Intelligence in Healthcare: A New Technology Benefit for Both Patients and Doctors. 2019 Portland International Conference on Management of Engineering and Technology (PICMET), 1–15. https://doi.org/10.23919/PICMET.2019.8893884

[26] Mali, G. P. (2018). Pros and Cons of Artificial Intelligence (AI) in Banking. Small Business Bonfire.https://www.smallbusinessbonfire.com/artificial-intelligence-banking/

[27] Root, T. (2018). Driver or passenger? Here's how Root makes the call. Root. https://www.joinroot.com/blog/how-root-knows-driver-or-passenger/

[28] Pyshkov, S. (2017). The real opportunities (and limitations) of AI for insuranc. LinkedIn. https://www.linkedin.com/pulse/real-opportunities-limitations-ai-insurance-sasha-pyshkov/

[29] Yao, L. (2015). STUDY ON BIPARTITE NETWORK IN COLLABORATIVE FILTERING RECOMMENDER SYSTEM [The Pennsylvania State University]. https://etda.libraries.psu.edu/files/final_submissions/11112