**NAME: OGUNDIPE OMOYENI, SHITTU OLAYEMI**

**PROJECT TOPIC: ARTIFICIAL INTELLIGENCE IN THE INSURANCE INDUSTRY**

With the power of the cloud and today’s rapidly changing technologies, forward-thinking insurers can leverage Artificial Intelligence to drive faster and more personalized customer experiences, increasing satisfaction with claimants and generating significant efficiencies in insurance underwriting.

The insurance industry has only begun its foray into AI, and companies are already experimenting new ways to incorporate AI into their day-to-day operations in anticipation of further technological development. The vision is that by 2030, AI could make more impact in the services they provide, and this evolution will shift insurance from its current state of “detect and repair” to “predict and prevent,” transforming every aspect of the industry in the process.

AI and its related technologies will have a seismic impact on all aspects of the insurance industry, from distribution, underwriting and pricing to claims. Some key areas include:

* Claim processing
* Client Journey
* Fraud Detection
* Customer retention and churn prediction
* Operation process optimization
* Risk Assessment

As we move forward in this research, we will streamline our focus to one of these areas and try to understand better how Artificial Intelligence can be fully optimized in that area and how it can positively impact the insurance industry as a whole.

**Fraud Detection**

Insurance fraud is an intentional deception against or by an insurance company for financial benefits. Claimants, policyholders, third-party claimants, or professionals providing services to claimants could commit fraud at different stages during insurance. Insurance agents and company employees may also engage in insurance fraud. Some of the common frauds perpetuated include "padding" (inflating claims), misrepresenting facts on an insurance application, submitting claims for injuries or damage that never occurred, and staging accidents.

Insurance fraud has been a Multi-Billion Dollar Fraud Problem in the insurance industry for decades now. A 2022 study by The Coalition Against Insurance Fraud (CAIF) showed that insurance fraud can cost U.S. consumers $308.6B yearly. That amount includes estimates of annual fraud costs across several liability areas, including Life Insurance, Property insurance, Workers Compensation, and Auto insurance.

**Traditional ways of detecting fraud by Insurers**

Historically, some of the ways insurance companies detect frauds are through:

1. **Analysis of claim history:** Insurers usually do a deep dive into the insurance claims of individuals, carrying out scrutiny to find patterns, frequency through historical claims. They do all sorts of data analysis to get information from the data they have and if your claim doesn't match the typical pattern, they’ll notice.
2. **Checklist of "Suspicious Loss Indicators":** Insurance agents look for "suspicious loss indicators" when they suspect a claim may be false. For example, looking to determine when and how a home fire started.

The National Insurance Crime Bureau (NICB) has developed a list of 23 “Suspicious Loss Indicators” that can signal that a claim may be fake, bogus or a rip-off.

Some of the suspicious loss indicators insurance agents look for:

* A claimant who's totally calm and unflustered after submitting a large claim
* A claimant who submits handwritten receipts for repairs on a covered item
* A claimant who adds or increases homeowners or auto insurance coverage shortly before submitting a claim
* A fire-damage claim for a home or auto where the fire started immediately after a family argument, or shortly after family members left the home/car
* A medical claim submitted by an employee whose job is ending

1. **Using Private investigators** to ensure claims aren’t false and leveraging social media for investigations.
2. Looking for Evidence of Personal Injury.
3. **Using sophisticated computer systems to ensure legitimate Billing:** Billing is one way people use to defraud their insurance. Often, they'll work with their auto repair, or health center as the case may be, to pad the bill to cover things like the deductible. In situations like this, insurance companies' computer systems can pull up claims where repairs appear inflated, or don't match with information provided about the claim.
4. **Handing Cases to Special Investigation Unit:** Many insurers have special investigation units, and they usually consist of well-trained individuals who have experiences working as detectives, police officers, medical personnel, etc. They're able and well trained to perform a variety of tests and checks to detect fraud. For example, they can:

* Conduct burn-pattern analyses and computer simulations on cars and homes damaged by fire to determine if the fire was intentionally set or accidental.
* Determine if a claimant's injuries match a reported accident.
* Investigate damaged vehicles to see if the resulting dents and scratches are from the reported accident. Also, use rust analysis and wear patterns to see if your car's damage is indeed from an old accident or not.
* Conduct financial reviews on claimants. Auto or homeowners claims from those who are behind on car or mortgage payments are immediately flagged as potentially fraudulent.

1. **Evaluate Prospective Employees' Credit Histories:** Claimants are not the only ones committing insurance fraud. This also happens with insurance agents, hence employees often check their credit before being hired. Insurance agents may commit fraud by "stealing" a customer's insurance premiums. In this common scam, an agent can take money for insurance and keep it without ever underwriting the policy for you. Insurance companies try to prevent such fraud by doing a credit check on all potential employees. Applications from people with credit or financial problems are flagged as most likely to commit fraud.
2. **Reporting claims in an online anti-fraud information system:** Thousands of insurance companies, self-insured entities and third-party administrators report all their claims to ISO Claim Search, an anti-fraud information system. The system was created by Insurance Services Office, Inc., and covers auto, property and liability claims. Cross-checking a new claim against all of those in this database (1 billion-plus) is one of the easiest ways for insurers to catch fraud.

**How AI is Transforming Insurance Fraud Detection**

Insurance fraud is a very big problem in the United States. And the people who commit it are increasingly creative, outsmarting some of the traditional checks and balances put in place, causing a large amount of insurance frauds to go undetected thereby constantly increasing the costs of insurance premiums.

It is becoming increasingly challenging for financial institutions and insurance companies to provide superior customer services, comply with regulatory requirements, and manage fraud, all at scale and in a cost-efficient manner. Traditional approaches are reactive, manual in nature, and ineffective, prompting institutions to consider leveraging Artificial Intelligence.

Some of the ways AI can be used to detect and prevent fraud in insurance are:

1. **Predictive analytics:** The first line of defense against insurance fraud is predictive analytics for early detection and fraud prevention. Based on historical behavior, predictive analytics can gain insights to an insured’s potential fraud risk, thereby taking the traditional system of “detect and repair” to “predict and prevent,” saving billions of dollars for insurance companies in the process. One predictive algorithm that is worth looking further into for this case is the logistic regression.
2. **Using NLP to analyze historical data:** In addition to processing mountains of information around the clock, Natural Language Processing can analyze historical data of past fraudulent claims and learn how to classify claims into groups of fraudulent or non-fraudulent claims by evaluating recorded conversations and other types of textual data, such as emails. Without using AI to detect claim fraud, this would be ineffective if not impossible to replicate with human labor alone. By tracking the historical trends in a person's claim history, the algorithms understand an individual's claim history and whether a particular claim seems normal or suspicious.
3. **Data mining:** Data mining can help third-party payers like insurance companies to extract useful knowledge from thousands of claims and identify a smaller subset of the claims or claimants for further assessment and scrutiny for fraud.

Combining automated methods and statistical knowledge led to a newly emerging interdisciplinary branch of science that is named Knowledge Discovery from Databases (KDD). Data mining is the core of the KDD process. In the domain of health care fraud and abuse detection, supervised data mining involves methods that use samples of previously known fraudulent and non-fraudulent records. These two groups of records are used to construct models, which allow us to assign new observations to one of the two groups of records. Supervised methods require confidence in the correct categorization of the records. Furthermore, they are useful in detecting previously known patterns of fraud and abuse. Some of the algorithms that are currently being used for fraud detection include clustering for outlier detection which can detect anomalous non-compliant fraudulent claims, decision tree, neural networks, Support Vector Machine (SVM).

1. **Real-time notifications:** With an artificial intelligence system that works around the clock and continuously monitors the habits and behavior of claims and policyholders, algorithms can easily flag potentially fraudulent activity and provide real-time alerts to the business when complaints require further investigation. The earlier insurers can be alerted to potentially fraudulent activity, the better protected they are against paying the requested amount and corresponding loss.

AI is a useful tool that improves a firm’s resource efficiency and saves insurers millions of dollars each year. With better early fraud risk detection, NLP to analyze historical claims data, advanced data mining, and real-time alerts, insurance companies can leverage AI to better manage fraud and resulting losses.

**Customer Churn**

It can be difficult enough to generate insurance leads and turn potential clients into customers. The fact that some of the already existing customers could also leave makes it considerably harder to expand an insurance company. The pandemic in 2020 also changed churn trends in the insurance industry in shocking ways. Today, consumer churn is impacted far more by the customer experience.

How AI is Transforming Insurance Customer Churn

1. Reducing human error: The distribution chain in the insurance industry is very complex and complicated. According to Breen, there are several middlemen who review information between the insured and the carrier, which causes a lot of human error and manual labor that slows the process. However, AI is beginning to resolve that issue. With AI and machine learning, both customers and insurers benefit, because insurers can develop better products based on more accurate assessments and will not have to go through the rigorous manual processes involved in onboarding a new client or recommending a product to an existing. This in turn reduces the chances of manual error. AI is skilled at noticing behaviors and trends, and insurance companies are going to be able to take advantage of that data.
2. Customer and Employee Service: To improve customer experience, many insurance companies are investing in virtual assistance like chatbots. A chatbot is a form of digital service that can interact with people in conversations that seem natural and also do specific objectives, such answering questions. Chatbots are accessible to respond to queries, check billing information, and address frequent enquiries and transactions in addition to providing basic guidance.

AI chatbots provides convenient and frictionless customer service that can address everyday inquiries, complaints, and transactions. Without the chatbots, customers might have a tough time getting through to their insurers to ask questions about the type of products offered or to get basic recommendations.

AI chatbots are up to a thousand times faster than humans and always accessible with no hold times. True AI chatbots are built on Natural Language Process (NLP) and Machine Learning (ML) that help them replicate human learning capabilities. One current example of this is Lemonade Insurance and their Chatbot, AI Maya, a playful onboarding and customer experience bot, and AI Jim, a claims bot that handles the “first notice of loss” for all their claims and can manage an entire claim through resolution without any human involvement for almost half of all claims. A two-minute chat with AI Maya is all it takes to get a personalized quote, sign up for a policy, and facilitate payments. This experience from start to finish is a core part of what makes their company Lemonade unique—giving customers a seamless, delightful experience.

**METHODOLOGY**

We made use of historical insurance claim data including normal and fraudulent ones, to investigate the normal/fraud behavior features. Based on ML techniques, and using these features, we checked if a claim is fraudulent or not. After this, a comparative study was performed to decide which Machine Learning classifier performed best.

We used 12 Machine Learning algorithms to build models for fraud detection.

The algorithms are:

* Support Vector Classifier
* KNN
* Decision Tree Classifier
* Random Forest Classifier
* Ada Boost Classifier
* Gradient Boosting Classifier
* Stochastic Gradient Boosting (SGB)
* XgBoost
* Cat Boost Classifier
* Extra Trees Classifier
* LGBM Classifier
* Voting Classifier

**DATA ANALYSIS**

The data used for this study is an open data gotten from [Kaggle](https://www.kaggle.com/datasets/buntyshah/auto-insurance-claims-data) and has been extracted from insurance claim settlement. The dataset contains 1000 rows and 40 columns.

After carrying out exploratory data analysis we saw that some columns had ‘?’ and ‘missing values’ (Figure 1) which we treated using mode fill.

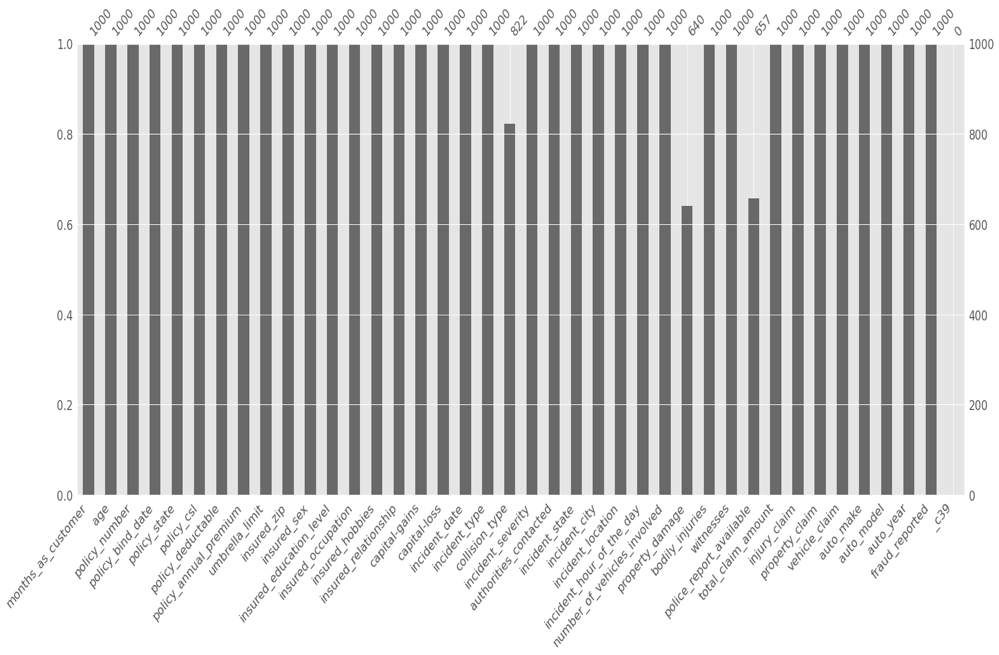


Figure 1: Visualization of missing values

Using a heatmap to visualize the correlation between numerical columns in the dataset

Chart, treemap chart

Description automatically generated

Figure2:Correlation between numerical variables

We see from (Figure 2) that there is multicollinearity in the numerical data, so we removed some of the numerical columns that are not necessary for prediction and to remove the problem of multicollinearity. After doing this, we were left with 27 columns.

Using a heatmap to check for multicollinearity again.

Chart, bar chart, waterfall chart

Description automatically generated

Figure3:Correlation between numerical variables

From the above plot (Figure 3), we can see that there is still multicollinearity in the data and there is a high correlation between **age** and **months\_as\_customer**, and between **total\_clam\_amount, injury\_claim, property\_claim,** and **vehicle\_claim.**

As total claim is the sum of all others, we will drop the total claim column and the age column. After doing that, we are left with 25 columns as seen below.

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 months\_as\_customer 1000 non-null int64

1 policy\_csl 1000 non-null object

2 policy\_deductable 1000 non-null int64

3 policy\_annual\_premium 1000 non-null float64

4 umbrella\_limit 1000 non-null int64

5 insured\_sex 1000 non-null object

6 insured\_education\_level 1000 non-null object

7 insured\_occupation 1000 non-null object

8 insured\_relationship 1000 non-null object

9 capital-gains 1000 non-null int64

10 capital-loss 1000 non-null int64

11 incident\_type 1000 non-null object

12 collision\_type 1000 non-null object

13 incident\_severity 1000 non-null object

14 authorities\_contacted 1000 non-null object

15 incident\_hour\_of\_the\_day 1000 non-null int64

16 number\_of\_vehicles\_involved 1000 non-null int64

17 property\_damage 1000 non-null object

18 bodily\_injuries 1000 non-null int64

19 witnesses 1000 non-null int64

20 police\_report\_available 1000 non-null object

21 injury\_claim 1000 non-null int64

22 property\_claim 1000 non-null int64

23 vehicle\_claim 1000 non-null int64

24 fraud\_reported 1000 non-null object

Moving on, we went ahead to encode the categorical variables, by turning categorical variables into numerical variables. The columns and variables that were encoded were:

policy\_csl: ['250/500' '100/300' '500/1000']

insured\_sex: ['MALE' 'FEMALE']

insured\_education\_level:['MD' 'PhD' 'Associate' 'Masters' 'High School' 'College' 'JD']

insured\_occupation:

['craft-repair' 'machine-op-inspct' 'sales' 'armed-forces' 'tech-support'

'prof-specialty' 'other-service' 'priv-house-serv' 'exec-managerial'

'protective-serv' 'transport-moving' 'handlers-cleaners' 'adm-clerical'

'farming-fishing']

insured\_relationship: ['husband' 'other-relative' 'own-child' 'unmarried' 'wife' 'not-in-family']

incident\_type:['Single Vehicle Collision' 'Vehicle Theft' 'Multi-vehicle Collision' 'Parked Car']

collision\_type: ['Side Collision' 'Rear Collision' 'Front Collision']

incident\_severity: ['Major Damage' 'Minor Damage' 'Total Loss' 'Trivial Damage']

authorities\_contacted: ['Police' 'None' 'Fire' 'Other' 'Ambulance']

property\_damage: ['YES' 'NO']

police\_report\_available: ['YES' 'NO']

**Reference:**

Accenture. (2022, September 5). *Transforming claims and underwriting with AI*. https://www.accenture.com/us-en/insightsnew/insurance/ai-transforming-claims-underwriting

Accern, T. (2022, September 12). *How AI is Transforming Insurance Claims Fraud Detection*. Accern. https://accern.com/blog/how-ai-is-transforming-insurance-claims-fraud-detection/

*Background on: Insurance fraud | III*. (n.d.). Retrieved October 25, 2022, from https://www.iii.org/article/background-on-insurance-fraud

Balasubramanian, R., Libarikian, A., & McElhaney, D. (2021, July 1). *Insurance 2030—The impact of AI on the future of insurance*. McKinsey & Company. https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2014). Using Data Mining to Detect Health Care Fraud and Abuse: A Review of Literature. *Global Journal of Health Science*, *7*(1). https://doi.org/10.5539/gjhs.v7n1p194

McManus, M. R. (2022, April 29). *10 Ways Insurance Agents Spot Fraudulent Claims*. HowStuffWorks. https://money.howstuffworks.com/personal-finance/auto-insurance/10-ways-insurance-adjusters-spot-fraudulent-claims.htm

Uzialko, A. (2022, September 19). *Artificial Insurance? How Machine Learning Is Transforming Underwriting*. Business News Daily. https://www.businessnewsdaily.com/10203-artificial-intelligence-insurance-industry.html