Social Media Insurance

Precious Mumbi & Stanley Sayianka

INTRODUCTION

Insurance is a risk transfer mechanism. It is a written contract between one party(the insurer) and another(the insured), whereby the insurer promises to cover the insured against financial losses caused by unexpected events. The insurer operates by pooling common risks together to create a portfolio of policies in order to enable the higher costs of the bad risks to be offset by the relatively lower costs of the better risked within the pool. The insured pays to the policy a premium[1], and in exchange for that the insurer promises to reimburse, or compensate the insured in the event of *financial* losses caused by peril, or damage to the insured's interests.

[1] The premium paid is meant to reflect the insured's contribution into the insurer's risk pool, making it an equitable amount.

The insurance system is one of the oldest systems, which was first practised as early as 4000-3000 BCE[2]. Some of the earliest forms of insurance include marine insurance, where merchants who had shipment at sea were granted loans such that, if shipment was lost, they did not have to repay the loan[3]. The insurance system rapidly grew to many other areas of human life to include contracts such as:

[2] A history of insurance

[3] This insurance system was commonly called a Bottomry contract

- Liability insurance: Some of the earliest forms of liability insurance include fire insurance, property insurance etc. with the most notable examples being in England(1666) after the *Great Fire of London*.
- Life insurance: This developed for instance in America in 1759, under the Presbyterian Ministers' Fund. Life insurance is reported to have taken off well by 1910.
- Auto insurance: Also known as vehicle/motor insurance, first emerged after the first world war[4], with a compulsory scheme being offered in America in 1930 under the Road traffic act[5]

[4] Wikipedia: Vehicle insurance

[5] The Road Traffic Act

Since then, the insurance industry has evolved to offer cover for almost any **insurable** risks, with interesting examples being:

- Wedding insurance[6]
- Body part insurance, commonly taken by musicians, sportsmen, and artistes.[7]
- Alien abduction insurance[8] etc.

[6] With one known as Change of heart insurance

[7] With notable ones being: Bruce Springsteen's Voice, Merv Hughes' moustache, Egon Ronay's Taste Buds, Cristiano Ronaldo's legs, bum bum policies,

[8] More info on the same here

Insurable risk

For a risk to be considered *insurable*, the risk must satisfy a given number of conditions, which are outlined below[9]:

[9] see *Wikipedia* for other points

- There must be a large number of units exposed to the same risk. This is to enable the insurer to pool risks together.
- Definite Loss: That the loss caused by the unexpected events, must take place at some known time, in a known place and be caused by a known cause. For certain classes, definiteness may be clear, while for others, it may only exist in theory. This however is at times objective.
- Accidental Loss: That the happenings constituting the trigger of a claim should be accidental and unexpected i.e. losses such as from gambling do not meet this requirement.
- Quantifiable loss: That the loss occurring from a given insured risk must be of a reasonable financial quantifiable nature. This also implies that the probability of loss, should be quantifiable in nature.
- That an affordable insurance premium can be reached at, by the insurer. A premium paid to the insurer is supposed to cover the expected costs of claims, costs of issuing, administrative costs, and an allowance for profit loading. However, a very large premium may turn out to be discouraging to the insured and no one would want to take up such a policy, thus the risks being insured must have an affordable premium.

In this study, we seek to develop an insurance product for covering risks and the financial losses associated with social media usage, some of which include: hacking, phishing, online abuse, online impersonation and web application attacks. Social media has grown to become an integral part of people's livelihood such as to: educators, musicians, politicians, marketers, influencers etc; thus developing a product/insurance framework to cover against social media related risks is of importance to an insurer willing to include such in their portfolio of policies.[10]

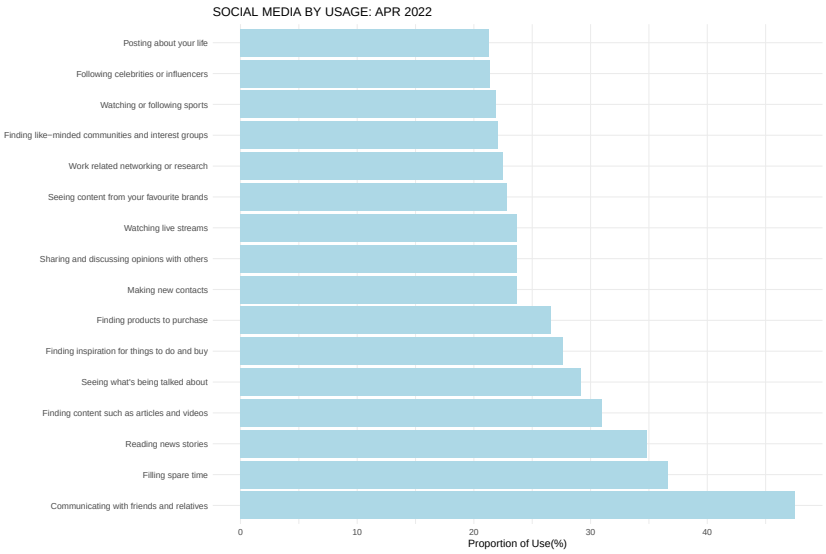
[10] In Kenya, cyber attacks insurance is relatively in its infancy stages, and one such policiy is offered by Britam Insurance

DATA

Social Media growth

The social media has grown to be one of the largest networks of communication, with an active 4.65 billion users globally by April 2022.[11] The most notable applications used include: Facebook, YouTube, WhatsApp, Instagram, Wechat, TikTok etc. ranked according to user base.[12].

As of April 2022, the rankings of social media by use are shown below[13]:



[11] Source

[12] Source

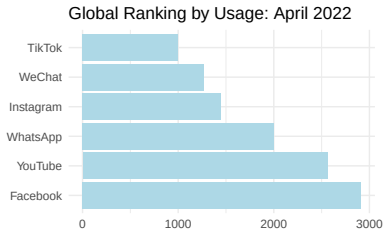


Figure 1: The world's most used social media platforms(Figures in millions)

[13] This is from a study done by DATAREPORTAL

The Internet and Social Media in Kenya

In Kenya, Internet usage has grown at a stable rate over the years, with a user base of 23.35 million in January 2022, which is 42% of the total Kenyan population.[14] Social media usage in Kenya stands at 11.75 million(*21.1% of the total population*), which is an increase by 6.8% from 2021. The historical growth of internet users is shown below:

[14] Source: Kepios

Subscriptions	2015/16	2016/17	2017/18	2018/19	2019/20	2020/21
Mobile data/Internet	26758789	29205204	40743570	49532380	40922499	42806044
Terrestrial wireless data/Internet	13449	47231	122037	66989	88159	91826
Satellite	280	693	1165	1243	1698	1275
Fixed Digital Subscriber Line (DSL)	3063	2715	1254	1014	997	995
Fixed fibre optic	27571	54700	135964	213199	351332	373835
Fixed cable modem (Dial Up)	77319	99971	101508	132072	176589	176081
*Other Fixed Data Subscriptions	NA	NA	7352	7408	804	804
Total	26880471	29410514	41112850	49954305	41452221	43450860

The various cyber attacks in Kenya, are shown below:[15]

[15] Data is fetched from Communications Authority Of Kenya

Cyber Attack Vector	2016/2017	2017/2018	2018/2019	2019/2020	2020/21
Malware	4146435	16306547	40893141	101651143	31842635
DDOS/Botnet	952327	3756334	4852022	1475537	1245451
Web Application Attacks	2656675	3743638	6109184	7662793	2057369
System Misconfiguration	NA	6 158	47913	108596	28482
Online Abuse	61	3295	458	196	134
Online Impersonation	46	368	568	585	220
Totals	7755498	23815972	51903286	110898850	35173937

The data used for modelling is quarterly data as shown below:

Period	Total Internet users	Web app attacks	Online Abuse & Impersonation
Apr - Jun 18	41111850	771518	681
Jul – Sept 18	42204503	1064971	448
Oct - Dec 18	45705440	737289	217
Jan - Mar 19	46870422	1222237	361
Apr - Jun 19	49954305	3084687	532
Jul - Sept 19	52008895	4069671	354
Oct - Dec 19	39657090	1908001	233
Jan - Mar 20	39394702	582281	510
Apr - Jun 20	41452221	1102840	441
Jul - Sept 20	43450860	2057369	348
Oct - Dec 20	44391490	7847457	222

The definitions of the cyber attack terms are as shown below:

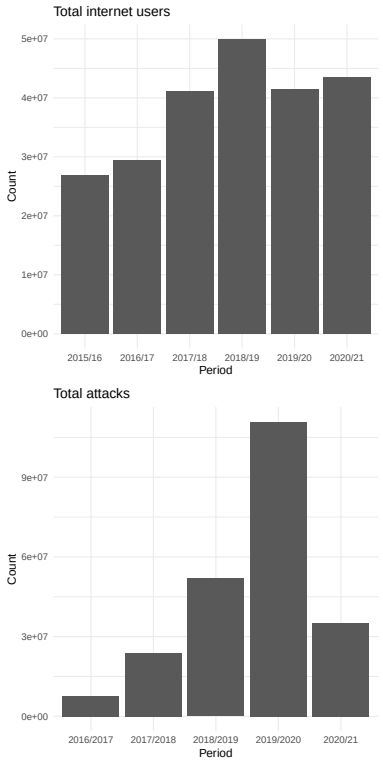
- Malware attack: This is a common cyber-attack where malicious software executes unauthorized actions on the victim's system.
- DDOS/Botnet: A Denial of Service attack involves a single machine used to flood a targeted resource wih packets, requests or queries. The DDoS attacks are commonly executed by botnets *(a collection of compromised computers)*.
- Web Application Attacks: These refers to conditions when malicious individuals/software exploits vulnerabilities in programs, websites, and softwares to gain access to a server or databases. Common web application attacks include: Cross-site scripting (XSS Attacks), SQL Injection attacks, and Broken Access Control Attacks.
- System Misconfiguration:[16]: Conditions in systems such as incorrect or sub-optimal configurations, that could be exploited by threat actors to gain unauthorized access to a system's functions and data.

[16] For this analysis, figures for system vulnerabilities are captured by this variable

- Online Abuse[17]: This is any form of abuse that happens over the internet, such as: Sexting, Cyberbullying, trolling etc.

- Online Impersonation[18]: This refers to cases where a threat actor is using someone's online identity for malicious reasons, such as: financial gain, ruining brand or reputation etc.

[17] For this analysis, i used "Online Abuse" as a combination of *child online abuse* and *general online abuse*



[18] Figures for *Online Fraud* are also incorporated into this

METHODS

Insurance pricing

In this section, this study aims to analyze the data on: Web application attacks, Online abuse and Impersonation in Kenya, with an aim of estimating the probabiities of the occurence of those events. Using the estimated probabilities, we then use the same for coming up with a fair premium to charge using the equivalence principle shown below:

$$Premium = I_R * C_C + E + (O + P) + U_L$$

where:

I_R: The incidence rate of probability. This is simply th probability that an internet user in Kenya experiences the above e.g. Online abuse, Impersonation etc.

C_C: The cost of claiming, which we will keep at a maximum value of 100,000.

E: The expenses incurred. This study uses a 10% Insurer's administrative costs.

$O + P$: The commissions plus profit margin. In this study, we use a fixed 17.5% as profit margin.

U_L: The insurer's uncertainty loading. In this study, we set the uncertainty loading equal to the model's standard deviation.

Beta-Binomial

This model takes a Bayesian approach in estimating the probabilities of occurence of the three risk events using a Beta-Binomial model for the same. We choose to use a Bayesian approach as opposed to maximum likelihood estimation, since under the presence of very few data points, the MLE approach leads to unstable estimates due to over-fitting.

We assume that the number of people who will experience either of the three events of interest(*Online abuse, impersonation and web application attacks*) follows a Binomial distribution as shown below:

$$X_t \sim Binomial(N_t, \theta)$$

The probability mass function is given by:

$$P(X_t = x) = \binom{N_t}{x} \theta^x (1-\theta)^{N_t - x}$$

Further:

$$\theta \sim Beta(\alpha, \beta)$$

The probability density function for the parameter θ becomes:

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Where:

X_t: is a random variable representing the number of people who experience any of the chosen events e.g. Online abuse.

N_t: is the population at risk of the events, i.e. the total population of internet users in Kenya. In forecasting the population at risk, we use a one-period arithmetic growth rate moedl given by the following:

$$N_i = N_{i-1}(1+ri)$$

Where:

$$r = \frac{N_t - N_o}{tN_o}$$

θ: The probability that an individual in the risk set(*Internet users in Kenya*) experiences any or one of the events of interest

The model hyper-parameters α and β are assumed known.

Under the Bayesian analysis, the posterior distribution $\pi(\theta|x)$ is then given by:

$$\pi(\theta|x) = \frac{g(x,\theta)}{m(x)}$$

With the joint distribution being:

$$g(x,\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} * \prod_{i=1}^{n}\binom{N_t}{x_i}\theta^{\alpha+\sum x_i}(1-\theta)^{\beta+N_t-\sum x_i}$$

The marginal distribution becomes:

$$m(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} * \frac{\Gamma(\alpha+\sum x_i)\Gamma(\beta+N_t-\sum x_i)}{\Gamma(\alpha+\beta+N_t)} * \prod_{i=1}^{n}\binom{N_t}{x_i}$$

The posterior becomes a *Beta* distribution[19] with the following parameters:

[19] The Beta prior is a conjugate prior to the Binomial distribution

$$\pi(\theta|x) \sim Beta(\alpha + \sum x_i, \beta + N_t - \sum x_i)$$

With the following properties:

$$Mean = \frac{\alpha + \sum x_i}{\alpha + \beta + N_t}$$

$$Mode = \frac{\alpha + \sum x_i - 1}{\alpha + \beta + N_t - 2}$$

$$Variance = \frac{(\alpha + \sum x_i)(\beta + N_t - \sum x_i)}{(\alpha + \beta + N_t)(\alpha + \beta + N_t + 1)}$$

The martingale model

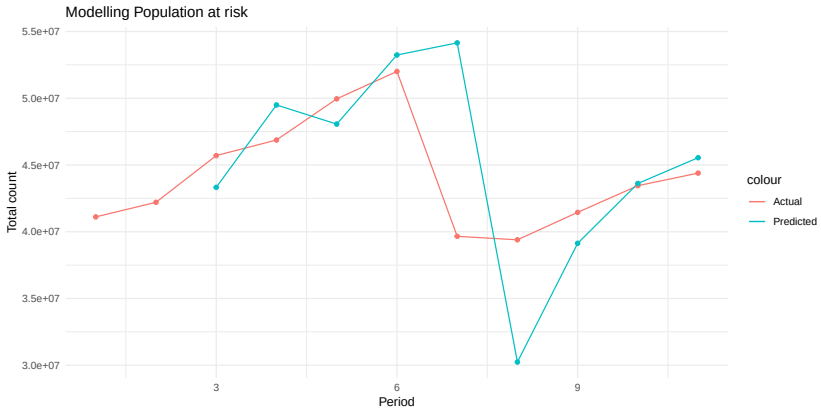
We fit a martingale model, as a baseline model to the beta-binomial model. THis model assumes that the expected drift in the proportion of individuals experiencing the event of interest is 0, and thus, the expected probability of experiencing the event of interest in the next quarter(Q_{i+1}) is simply this quarter's(Q_i) proportion of people who experienced the event of interest.[20]

[20] This is simply a MA(1) model.

ANALYSIS

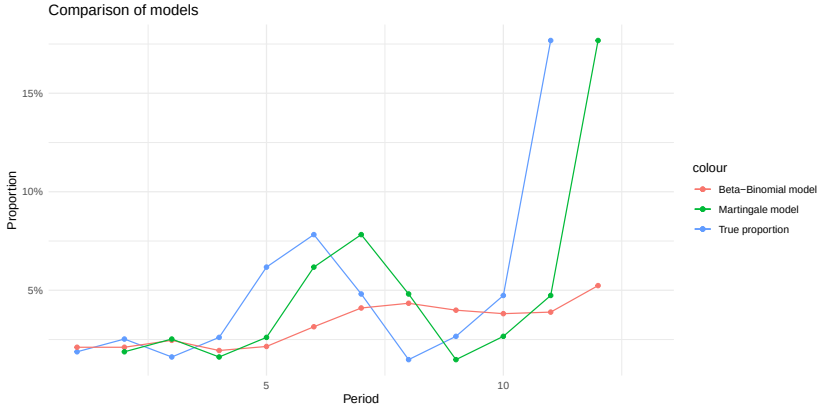
This section covers the modelling of the probabilities for the events of interest. In estimating the probabilities of interest from their posterior distributions, we takes on a sequential approach, where the posterior distribution of the probability of occurrence is composed of the previous prior and the data upto one time step back $(t-1)$. Then, once we compute the posterior distribution for the current time step (t), it then becomes the prior distribution for the future time step $(t+1)$.

The projected population model is shown in comparison with the actual population:



Web application attacks:

We select a prior beta distribution: $Beta(\alpha = 22.614, \beta = 1005.22)$ for web application attacks. The prior distribution is selected based on information from the first two quarters in the data. These two quarters will be omitted in subsequent analysis. The graph below shows the comparison of the true proportion of web application attacks and the estimates from the beta-binomial model, and the martingale model.



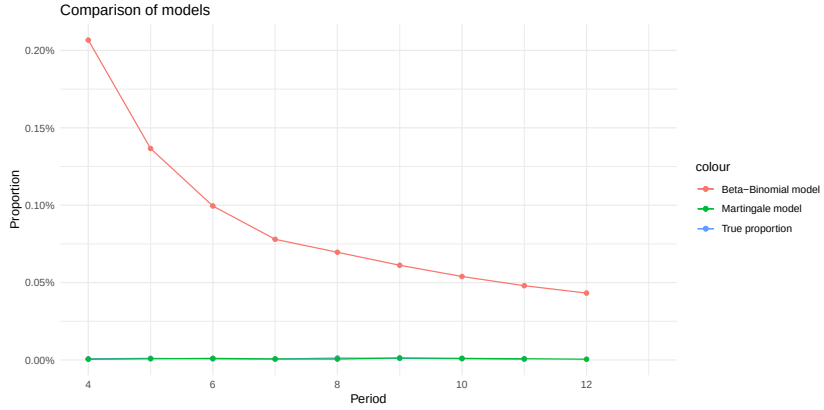
From the above chart, it is evident that the Martingale model per-

forms better that the Beta-Binomial model in modelling proportion of internet users affected by web application attacks. It is although important to note that the period April-September 2019 experienced a sharp increase in total population of internet users as well as those affected by the numerous cyber attack vectors. This 'outlier' period greatly affects the model performance. The Mean Squared error analysis for the models indicates that the MSE for the benchmark model(martingale model) is slightly lower than that for the Beta-Binomial model.[21]

[21] For the benchmark model, the MSE is: 0.002110968, while for the Beta-Binomial model it is: 0.00240827

Online abuse and Impersonation

In this analysis section, we prefer to combine the numbers of online abuse and online impersonation into one, due to the similarity of the nature of the two. THe prior distribution for modelling the proportion of internet users in Kenya who experience online abuse and impersonation, was chosen to be: $Beta(191958, 191955)$. The choice of the prior was informed by the proprortion of the first two quarters.



From the chart above, the Beta-Binomial model fails to capture the proportion of online abuse and impersonations as accurately as the martingale model. The MSE analysis indicates that the best model is the benchmarking model.[22]

[22] The martingale model ha an MSE of: 1.741546e-11, while the Beta-Binomial model has an MSE of: 0.02500211

Product pricing

For the purpose of insurance pricing, we use the martingale model to estimate the proportion of population at risk experiencing: Web application attacks or Online abuse and Impersonation. The insurer's uncertainty loading is set to be proportional to the variance of the martingale model. The analysis is shown below:

	Web application attacks	Online abuse and Impersonation
Quarterly incidence rate	0.176778409555525	5.0009585170491E-06
Insurer's admin cost	0.1	0.1
Insurer profit margin	0.175	0.175
uncertainty loading	0.100730236584941	5.65812299445922E-06
Average cash benefit	100000	100000
Quarterly risk premium	17677.8409555525	0.50009585170491
Quarterly office premium	21285.2742755477	0.537018942522563
Monthly office premium	5321.31856888692	0.134254735630641

Figure 2: The summary of the constructed insurance pricing model.

From the above table, the monthly office premium for the Web application attacks is about 5,300, while for the Online abuse and impersonation, the premium is less than a shilling due to its low incidence rate.