Improving Operational Efficiency of Government using Artificial Intelligence

Saif Salam Ibrahim Al-witwit¹, Abdullahi Abdu Ibrahim²

¹Electrical and Computer Engineering, Altinbas University
²Electrical and Computer Engineering, Altinbas University

E-mail: saifsalam459@gmail.com

Abstract

In this paper we developed a technique for Improving the operational efficiency of the government using Artificial Intelligence, we discussed the concept of insurance and how the concept of insurance will change and Artificial Intelligence (AI) is already disrupting the state of this industry. Insurers worldwide are using AI to automatize processes and tasks, such as fraud detection, underwriting, and claims processing. Additionally, there has been a rise of new competitors in the market, such as InsurTechs, that are bringing innovative solutions for insurance using Artificial neural network (ANN), responding to the new trends in customers’ lifestyles and behaviors, that are more demanding for services directed for their needs. This study aims to understand how personalization of insurance policies, created with Artificial Intelligence and how its efficiency can be improved, and how it will disrupt this industry in the future and what will be the impact on the government's operational efficiency. We have chosen worldwide Governance indicator dataset which is publically available the personalization of an insurance policy with AI would encompass the definition of the coverages and premiums more appropriate for an individual customer and do the risk evaluation, in a market of one strategy. This innovation would take advantage of the accrual of Big Data from customers for the optimization, as people are each time more connected and information about them is constantly being shared, allowing companies to use it to know consumers better and for the training, testing and validation a well-known MATLAB R2019a software was used for this purpose. We achieved an accuracy of 95.25% using 9-Fold cross-validation.

Keywords: Artificial Intelligence, ANN, Big Data, optimization.

1. Introduction

Industries all over the world are being disrupted by technological advances that are transforming their existence, and the way companies react and adapt to these disruptors will have massive impacts on how industries will look like in the future.

An example is the Insurance industry that has been under the disruption of new trends, such as Artificial Intelligence (AI), driven by the quick advances in the technological field. This transformation includes several other disruptive technologies, such as telematics, Internet of Things, block chain and digital platforms [1] and the accumulation of Big Data has been feeding these advances. Technology based innovations have several benefits for insurers, for example, for controlling risks, cost efficiency, better engagement with its customers and tailoring insurance offers [3]. With the evolution of technologies, insurance will change its current state from a “detect and repair” to a “predict and prevent” approach [4]. Some of the main issues to tackle in insurance are technology advancements, regulatory implementations, product development, mergers and acquisitions, privacy issues and tax reforms, which insurers must take into consideration in order to improve operational efficiency, increase productivity, lower costs and customize their policies [5]. Customers’ expectations towards companies and the desire for personalized and flexible insurance policies is emerging and insurers must assess their ability to respond to it. As a result, an alternative for insurers is to join forces with InsurTechs to implement new approaches, platforms and policy plans [9]. The purpose of this study is to hypothesize the future of the Insurance Industry in Turkey, by identifying the major disruptors in this industry, with focus on Artificial Intelligence. More specifically, it intends to assess how AI powered solutions can be used to personalize insurance policies, in a market of one strategy. Personalization would encompass creating a policy suitable to a customer’s individual needs, involving the definition of the most appropriate coverage extent, capital available, premium and evaluation of the risks.Every day, millions of data about customers’ lifestyle and behaviors are generated and accumulated as Big Data. Artificial Intelligence has been a solution to transform and analyze the high amounts
of information that can be further used to create personalized policies based on client’s exact need, for different categories. The creation of this innovation requires high amounts of data relative to customers, which can originate privacy constraints, in which individuals may be withdrawn to share. Furthermore, as the insurance industry is highly regulated, Turkey insurers must consider the regulation applied for their activity and regarding discrimination and data protection. In Turkey, an example is the General Data Protection Regulation (GDPR), created by the Turkey Union in 2018 (EY, 2019) [10].

Personalization of insurance policies using Artificial Intelligence may be an opportunity to increase the number of policyholders by increasing the intention to acquire an insurance, as many people still do not want to acquire it, whereas because they consider themselves to be healthy – in the case of health Insurances, or due to its high prices or unnecessary coverages. In addition, it can lead to increases in revenues and accelerate industry growth. From the managerial relevance of this study, it is pertinent to highlight the Insurance industry and the AI developments. Insurers contribute to the economic growth, financial stability and development of Turkey and companies search each time more for solutions that embed AI benefits in their strategies, especially for reducing costs and increasing efficiency. The academic relevance of this study is to present a recent and in high demand topic to the academic community. In addition, it intends to deliver insights on how to define strategies to overcome or leverage disruption in industries. The problem this dissertation attempts to understand is how a market of one strategy, with personalized insurance policies by Artificial Intelligence, will affect the future of the Insurance industry in Turkey. The Insurance industry is composed by two main groups of policy contracts: life and non-life (general) insurance. The insurer pays a compensation to the policyholder in case of loss, damage or death, in return of a regular payment of the premium. Insurance policies contain coverage for losses arising from uncontrollable factors and allow risk avoidance, due to wealth constraints, to transfer risk to an insurer that takes a risk neutral role [7].

Life Insurance is a contract in which the policyholder acquires an insurance that will be reimbursed upon death or maturity, being considered a form of investment. It is a complex service, often associated with uncertainty of its unsure future benefits, and the way customers choose it depend on their proposal, agent, image of the insurer company and it also varies according to the culture of the country [8]. General insurance consists in all types of insurance of non-life policies (e.g. health, work, property, vehicle), and consist in reimbursements for uncertain events. Non-Life insurance products are standardized and prices follow actuarial principles. The price definition involves risk evaluation of the insured person or object, that is private to the firms [2]. Insurers hold a portfolio of insurance policies and an investment portfolio, and the income of insurers come from two sources. The first one is from the portfolio of customers insured, originated from the underwriting process, that is the difference between the premiums paid from customers and the payments made to the policyholder for incurred losses and expenses [6]. There exists a concept of mutualism, based on a principle of pooling risks, in which policyholders who do not have losses subsidies those who do [8]. The second source is from insurers investing the premiums earned in common equities and fixed-income securities [9]

2. Methodology

In this research work, Collection of raw data is very hard to do. Moreover, working with raw data is tough as it contains many repetitive rows and anomalous values which do not reflect the true situation. Therefore, raw data has to be filtered and moulded into a dataset that can be utilized according to the researcher’s purpose. Similarly, our data is filtered to remove rows that are exactly same. Synthetic values are introduced in order to reflect better results on the sets. The purpose of this paper is to developed a technique for Improving operational efficiency of government using Artificial Intelligence, how personalization of Insurance policies, created by Artificial Intelligence, will affect the Insurance industry status-quo. The research questions that the study aims to address are: RQ1. What is the impact that AI will have in the Insurance industry on government operational efficiency? RQ2. How is this industry going to look like in the future, once this disruptor has taken full effect? Therefore, for the purpose of this study, primary and secondary data were used. The secondary data was collected from existing literature, in specific journals, academic articles and companies’ publications, in order to get knowledge about Insurance, Artificial Intelligence and the usage of this technology applied to the industry. The quantitative data was gathered based on industry indicators in Turkey, such as market shares and revenues, as well as information related to the main players in the Turkey market. From the Artificial Intelligence point of view, information regarding the investments made by enterprises in these innovations and the adoption of companies was collected through websites, studies and surveys previously done by other entities. In addition, a survey was held with Turkey residents, to assess respondents’ opinion regarding the insurance industry, the personalization of Insurance policies based on AI methods and their perception of it.

2.1 Online Survey
The online survey was an instrument to collect qualitative data to better understand Turkey respondents’ perception about Insurance, their motivation for acquiring innovative insurance products and their opinion regarding the usage of AI for a personalized policy. The survey was expected to have a total of 100 respondents, residents in Turkey countries, with ages from 18 to 65+ years old and different occupations. It was shared through Social Media channels. The survey was analyzed with excel, using descriptive statistics. The quantitative data from the study was collected through online searches on studies done for collecting information from the Insurance industry and the use of Artificial Intelligence. The information was extracted from the Insurance Turkey website, the Narrative Sciences survey in 2018 and the International Data Corporation.

2.2 Short to medium term impact of Artificial Intelligence in Insurance on Turkey Government

Worldwide insurance premiums have been increasing (Insurance Turkey, 2019) but the state of the Turkey market registered a slow growth and weak performance during the last 5 years, with decreases in the annual gross written premiums between 2012 and 2017 of -0.5% for life insurance and -1.2% for non-life insurance (EY, 2019). In 2018, the Turkey market corresponded to 31.6% of the total global insurance premiums, total premiums increased 6.2% and total claims and benefits paid grew 3.1% in comparison to 2017 (Insurance Turkey, 2019).

![Fig 1: Source: Insurance Turkey 2019](image1)

Life Insurance was the segment with the highest premiums and claims paid. In Turkey, on average €2.170 was spent per capita in insurance, from which 1264€ was in Life insurance, 238€ in Motor, 232€ in Health and 174€ in Property (Insurance Turkey, 2019).

![Fig 2: Turkey Premiums, Claims and Benefits paid in 2018](image2)

The Turkey countries with highest total premiums written in 2018 were the United Kingdom, France, Germany and Italy (Statista, 2019). Economic conditions had a strong effect on the performance of non-life insurances over the last years, and insurers must adapt to customer’s expectations and innovate in operations by making use of Big Data, analytics and AI (EY, 2019).
2.3 Artificial Neural Network Approach

Artificial Neural Network algorithm is also a classification algorithm more commonly known as ANN. The ANN algorithm classifies the data into different groups which maximizes the utility for any researcher. Based on the different groups, we can separate the data belonging to one class from data belonging to another class. ANN has been applied to places such as time series prediction, face recognition to biological data processing. The ANN algorithm draws a hyper plane which is said to separate the two classes.

The hyper plane is just not any ordinary line, but a line which tends to separate the two classes in such a way, such that the distance from the line to the closest point of both the classes remain as large as possible. In other words, it maximizes the distance from the hyper plane to each of the nearest data point of both the classes. The following equations describe the process on how this hyper plane is drawn and two different classes are identified.

\[ w\times x - b = 0 \]  \hspace{1cm} (1)

The equation is of the hyperplane which separates the two classes equally. Here \( w \) is the normal vector to the hyperplane and \( x \) can vary for each set of data, \( b \) is the distance between the nearest point to the hyperplane and the hyperplane itself.

\[ w\times x - b \geq 1 \] \hspace{1cm} (2)

If, however the result to this equation is positive and equal to one, then we get the hyperplane of the positive class, and any value greater than 1 would give data points of the positive class.

\[ w\times x - b \leq -1 \] \hspace{1cm} (3)
Similarly, if the result is equal to -1, then it gives the hyperplane of the negative class and any value less than -1 would be for data points belonging to the negative class.

2.4 Training with Artificial Neural Network

For our dataset, we first found the value of ‘k’ that best suited our case. We initially classified with ANN. However, this had no real logic behind it. As a result, we decided to find the optimal value that will give better and correct accuracy. It works by training our model on a range parameters and then finds out the best result based on comparison. In our case, we provided a range of values of k neighbors from 1 to 25 and checked which value of k neighbor provided the most accurate and reliable output. It is seen that best suited our data set. After determining the value, we load the data set. Duplicate rows with same values for all the four features are dropped to avoid training the same data over and over again. It also prevents testing and training with same dataset, avoiding accuracy levels that are unrealistic. Next, we split the data into train and test. 80 percent of the sample is train set and 20 percent of the data is test set. The 80 percent data is trained with ANN classifier with k initialized as 2. We then test the 20 percent data. It is done by calculating the Euclidean distance between instances of the clusters or classes formed. The distances are then sorted in ascending order. The top N rows are then selected and the most frequent class of these rows are then returned. Finally, we calculated the accuracy and it was very high.

3. RESULTS

The results are presented in this section. The finest algorithm for operational efficiency and classification on this dataset is ANN. First of all, its accuracy is very high compared to other algorithms. Moreover, the execution time is also very low. As the complexity of the kernel decreases, the algorithm begins to get faster. The other algorithms are more or less same in terms of their execution time. Random Forest has the highest execution time due to the fact that it builds numerous tree before coming to a conclusion or decision. In this section, we have considered two parameters while analyzing the machine learning algorithm (ANN) on the dataset. They are accuracy in terms of percentage and time in terms of seconds. The experiment is carried using Matlab R2019a Windows environment on 32GB RAM and 2.8 GHz Intel Core i9. After carrying out several experiments using the artificial neural network, we have come to present all our findings here. We carried out artificial neural network algorithms on our data-set. We have done the test procedure with 20 percent data which is equal to 13583 data. All tests were run on a dual 12 core AMD Magny Cours equipped with a NVIDIA Tesla C2050 GPU with 14 Streaming Processors (32 cores each). The artificial neural network was created and trained on all 115 features. Default Keras optimization hyper-parameters were used. Training was limited to 100 epochs with additional condition of early stopping, using functionality provided by Keras. This ensures that the model is trained only until the score on validation set is not getting worse – this in turn helps to avoid overfitting the training set.

3.1 Number of training samples

The effect of varying the number of training samples. There is a clear trend of smaller test error as increase in the number of training samples. At 689 training samples, the error is only 1.32%. Increasing the number of training samples had the biggest impact in reducing the error when compared to other factors.

(a) Varying number of training samples.
3.2 Number of Features

The effect of varying the number of features. Across most of our experiments, I observed that using 2000 features outperforms 3000 features, which outperforms 1000 features. However, the difference in error is small. This suggests that if I sampled 1000, 2000, and 3000 features again, I may observe a different trend. Overall, none of the alternative architectures achieved a higher maximum classification accuracy on the five selected classes than the base multi-class design for budget balancing.

3.3 System Modeling

The model trained in total for 24 epochs, taking about 27 seconds for each epoch, totaling in 648 seconds of training time. Figure 6 demonstrates the training curve. Loss is defined as MSE between original and predicted values. Training set had 15310 samples and optimization (validation) set 15311 samples.

Additionally, neural network models were trained on reduced number of features for the purposes of their comparison. Some modifications in the training procedure were introduced, such as the maximum number of epochs was limited to 50.
We make a comparison between more traditional ML methods for operational efficiency and the artificial neural network implemented in this work.

### 3.4 Personalization of an Insurance policy

Different suggestions associated with the personalization of an insurance policy were presented in the survey (Figure 9), and the option with highest interest for Turkey respondents was “I can activate/deactivate it whenever I need/want”, considered as Very Interesting, followed by the proposals of “Personalize the extent of services coverage” and “Define the amount I need for my coverage”. The option considered less interesting for customers was “Insurers using Artificial Intelligence to know me better”, considered Moderately Interesting (2.95).

However, when respondents were asked about their level of agreement with the usage of Artificial Intelligence techniques to create personalized insurance policies, the majority of respondents agreed (35%), followed by a neutral opinion (28%). When asked to consider that a personalized policy made by Artificial Intelligence techniques would be able to define the services coverage based on their lifestyle, having what they need and at a price based on their purchasing power, but it
required information about themselves (e.g. information shared online), the majority responded that they still agree (33%) or are neutral (26%) about it. However, negative responses for disagree (20%) and strongly disagree (14%) were also relevant, and it is relevant to highlight that the neutral opinions might be more likely to become a negative position towards this innovation.

3.4.1 Information allowed to be used for personalization of an insurance policy

When respondents were asked about if they would allow or not allow insurers to use certain information about themselves, the most ones that respondents would allow to be used is their physical activity, driving behavior, eating patterns and shopping behavior. The information that they would definitely not allow insurers to use would be photos of themselves and social media activity (Table 1). This information could be gathered mainly though their smartphones, devices for health monitoring, car sensors and fitness trackers (Table 1).

Table 1: Customer's information to be used in insurance

| Information that Turkey respondents would allow to be used for personalization of their insurance (%) of respondents | Would Not Allow |
|---------------------------------------------------------------|-----------------|
| Physical activity (67%)                                         | Photos of themselves (87%) |
| Driving behavior (65%)                                         | Social media activity (75%) |
| Eating patterns (54%)                                          | Social life habits (67%) |
| Shopping behavior (41%)                                        | Other behaviors (e.g. Smoking) (51%) |

Table 2: Devices for data collection

| From the following devices, which ones would you allow to share information for AI processes? (%) of respondents |
|---------------------------------------------------------------------------------------------------------------|
| Smartphone                                                                                                    | 55% |
| Devices for health monitoring (e.g. blood pressure wristbands)                                               | 52% |
| Car sensors                                                                                                   | 35% |
| Fitness trackers                                                                                                | 34% |
| Computer                                                                                                      | 29% |
| Payment methods                                                                                                | 24% |
| Smartwatch                                                                                                     | 22% |
| Clothing with sensors                                                                                          | 14% |

3.5 Evaluation using Artificial Neural Network

The confusion matrix for ANN. Out of the total data, 80 percent is used to train the data. The remaining data is used to test the data. From the figure, we can see that 8703 data are True positive. These are the data rows whose labels are labelled as normal or 1 and are also identified as normal or 1. Similarly 3868 data are true negative. These are data that are identified as abnormal or 0 and are actually labelled as 0 or abnormal. The 861 data which are actually abnormal or 0 are predicted to be normal or 1. Similarly, 421 data that are actually normal or 1 are predicted to be abnormal or 0. This gives ANN a higher accuracy rate compared to the other algorithms, which is about 95.2 percent. The reason for this high accuracy is due to the fact that the hyperplane that is drawn can separate the two classes accurately. We have used a sigmoid hyperplane to separate these data into their according classes for which the sigmoid line performs very well at separating the abnormal and normal data. Other kernels of Linear lines will perform well because our data is very continuous rather than discrete. Since it is spread out everywhere randomly, we preferred to choose a sigmoid curve.
Fig 10: The confusion matrix for ANN.

Artificial neural network works better with multiple features as input. Since our data set has 4 classes of features, the algorithm works well. The accuracy also depends on how many classes the artificial neural network algorithm has to convert it to. In our case, the logistic function has to map between two classes. In binary artificial neural network implementation, the value set as the threshold value for the boundary between two classes can be high which means that it can separate between different sets of data more accurately. On the other hand, there would have been multiple thresholds for each set of output classes if there were more than 1 group of outputs. Thus, each threshold would have been lesser in value and the difference between one threshold value for one class and another threshold value of another class would also be lower. This can eventually lead into more errors. Thus, for binary artificial neural network, the chances of making mistakes is lower. Therefore, artificial neural network in our case has a high value of accuracy.

4. DISCUSSION

Most Turkey residents own at least one type of insurance (84%) and the survey results shown that the categories with more policyholders are Health, Auto, House/Property and Life insurance. The main reason for not owning any insurance is because it is considered as too expensive (Figure 11).

Figure 11 Reasons for not owning any insurance

The most important insurance categories for Turkey respondents were Health (4.58 out of 5), Auto (4.34) and House/Property (4.26). The categories Life (3.82), Personal Accident (3.62), Pets (2.93), and Technological Items (2.14) had lower scores. Furthermore, the insurance industry was considered by Turkey respondents as traditional, useful, apathetic, expensive, relevant, complex and more untrustworthy than trustworthy (Table 3).

Table 3: Characterization of the insurance industry

| Classification | Innovative | Traditional |
|----------------|------------|-------------|
| How would you describe the Insurance industry role in government efficiency? | | |
Artificial neural network was also used and it works by using multi-variable analysis as it predicts a single outcome from multiple variable features. The model is trained by using the neural function which sets a threshold value in probability that finally determines the two classes. This algorithm shows an accuracy of 95.2 percent in 2.17s. Artificial neural network is implemented on the network layer and input features that were trained for improving efficiency and allocating budget equally.

5. CONCLUSION

In this novel research work, we have developed an artificial intelligence based system for improving efficiency of government, we have taken insurance into consideration. Insurance is a highly regulated industry and is considered as expensive, traditional and with low engagement by its customers. The response to the Research Question 1 is that the impact of AI policies in a Turkey level will be made more difficult. Turkey respondents would be willing to have this type of personalization and it could be a way to tackle the respondents who stated that they do not have an insurance because it is too expensive or unnecessary for them. However, due to the hurdle of the recent GDPR law, regulation will be the limitation that will be more difficult to mitigate. In addition, to create a policy using Artificial Intelligence, it is necessary to have a data set with information from customers that would be used to design the most appropriate policy, based on the customer’s exact needs. For this, several strategies can be taken, such as doing questionnaires with customers or with the help of brokers. For Research Question 2, on a worldwide scale, the disruption of AI policies would be faster to be implemented. Traditional insurers will take longer to adapt this innovation due to the higher risk aversion related to regulation and the portfolio of existing customers, as the impact in revenues of innovations will be higher. However, partnering with Start-Ups and InsurTechs can be a strategy to improve and find new solutions to keep aligned with the demand for higher customization of policies. Additionally, it will be necessary to create protocols and data ecosystems to guarantee that data can be used and shared. Artificial Intelligence has still some limitations and for insurers working in English, it may be faster to implement a market of one strategy, due to the higher amount of noted data and because it is an easier language.

In the future, insurers must be able to rethink their mutualism models and strategies to stay competitive with the new entrants in the market. For this, it is essential to develop services and ecosystems that improve the engagement with customers and that will be the source to get access to customer’s data. Telematics is already being used in car insurance, with a concept of “pay as you drive”, in which premiums are based on the driving behavior. It is a good indicator that other categories can move towards the same direction and use sensors to monitor customer’s behaviors. InsurTechs also offer a concept of Usage-Based Insurance and micro coverage insurance for technological items that be activated/deactivated, for instance, being one of the possibilities that Turkey respondents value the most. A market of one in insurance policies can harm some clients due to their data, and it can be an obstacle. However, in the future, insurers could develop to a more
sophisticated model that would encompass the definition of the whole policy, but traditional insurers would need to move towards the same strategy and regulation would have to adapt to the changes that the advances in technologies and customer’s expectations would bring. Moreover, ANN’s run time and accuracy are very high in terms of improving the efficiency of the government. Hence, it can be included while improving operational efficiency. Furthermore, for the network layer, an artificial neural network has a very decent accuracy with the least execution time. Although we have achieved very good accuracy with the artificial neural network algorithms, there are other data sets with more features and different complexities where these algorithms might not detect intrusions with great precision. On those data sets, other machine learning algorithms or intrusion detection techniques might not work better. Lastly, the intrusion detection and classification is done on the end devices. We can carry out this detection and classification on the cloud or server in the near future with the help of deep learning as recommended.

References

[1] Ashish Arora, S. B. (2017). The decline of science in corporate R&D. Strategic Management Journal. Chang, T., Lee, C., & Chang, C. (2014, 12 27). Does insurance activity promote economic growth? Further evidence based on bootstrap panel Granger causality test. Turkey Journal of Finance, 20(12), 1187-1210.
[2] Deloitte. (2019). 2019 Insurance Outlook - Growing economy bolsters insurers, but longer-term trends may require transformation.
[3] Desyllas, P., & Sako, M. (2013). Profiting from business model innovation: Evidence from Pay-As-You-Drive auto insurance. Research Policy, 42(1), 101-116.
[4] Duan, Y., Edwards, J., & Dwivedi, Y. (2019, 10 1). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management, 48, 63-71.
[5] Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018, 9 3). Using artificial intelligence to create value in insurance. International Journal of Bank Marketing, 36(6), 1145-1168.
[6] Salathé, M., Wiegand, T., & Wenzel, M. (2018, 9 13). Focus Group on Artificial Intelligence for Health.
[7] Serpa, J., & Krishnan, H. (2017, 2 1). The strategic role of business insurance. Management Science, 63(2), 384-404.
[8] Statista. (2019). Leading countries for total premiums written on the insurance market in Turkey in 2018. Retrieved from https://www.statista.com/statistics/820843/insurance-Turkey-total-premiums-written/
[9] Xie, M. (2019). Development of Artificial Intelligence and Effects on Financial System. Journal of Physics: Conference Series.
[10] Duan, Y.; Edwards, J.S.; Dwivedi, Y.K. Artificial intelligence for decision making in the era of Big Data—Evolution, challenges and research agenda. Int. J. Inf. Manag. 2019, 48, 63–71.