The Use of Internet of Things and Big Data to Improve Customer Data in Insurance Company

Gunawan Wang[1], Natanael Alamas[2], and Marcelina Anggraeni[3]
Information Systems Management Department, BINUS Graduate Program - Master of Information Systems
Management, Bina Nusantara University, Jakarta, Indonesia 11480
[1]gwang@binus.edu, [2]natanael.alamas@binus.ac.id, [3]marcelina.anggraeni@binus.ac.id

ABSTRACT

Recent technologies create a connectivity between business and their customers. Business can collect real time customer data by using internet of things and then processing those data using big data management to resolve problems, or to create new products. Insurance is a business industry that uses historical customer data to decide certain business aspects, such as premium amount for customer. Insurance company can combine internet of things and big data to collect customer's data in real time and then use it to give fair and accurate premium calculation for the customer. Customer data is collected using various sensors in their smart watch and saved by an application. The application send the data to insurance company's system through the internet. The customer data are stored and processed in Hadoop as the big data management system. Customer data are going through an analytical process to create result about specific customer's risk which later can be used to calculate premium.

Key word*s* **:** analytics, big data, customer data, internet of things

1. INTRODUCTION

In digital era, lots of data are collected from people every day. People's data can be retrieved from various sources, such as social media, web surfing, shopping history, and even daily activities. These people's data can also be collected very quickly, in real time. Since it happens very quickly, the volume of data also grows larger very quickly too. This kind of data is called as big data. Big data is characterized by three aspects: the data volume is huge, data consists of structured and unstructured data, and the data are generated and collected very quickly [1]. Big data are valuable resources for business. If big data is utilized correctly, it can give competitive advantage. Business can learn about their customer's habit and preferences to have more understanding about their customer. By understand about their customer then business can create new business opportunities, such as new services offered to the customer or new operations model [2].

There are many ways to collect data from customer. For example when customer buys some products through website, they submit data about which pages that they view,

which products that they have interest in, their customer profile, and their transaction information. In conclusion, all activities that customer does can become data. However, there are some valuable data that may come from customer's activity that is not done directly with the company. As example, data about how customer uses their car may be valuable for a car company. This kind of needs bring new technology which is called Internet of Things (IoT). IoT is an approach to collecting data from different kinds of things that can be connected to the internet [3]. Using this technology, company can collect data from any devices and objects owned by the customer.

Insurance industry is a business industry which relies on data to determine certain aspects in their business. Insurance company have actuarial department, which practice mathematical methods to calculate the pricing for insurance contracts such as the premium rate, or valuation of insurance contract [4]. The actuarial team assembles and analyzes data to assess the cost and possibility of certain unfortunate events like sickness, accident, and death [5]. However, this traditional calculation method produces same premium rate for certain gender and age. So it is possible that customer living healthy lifestyle must pay same premium amount as customer that lives bad, unhealthy lifestyle. If somehow, insurance company can obtain real time data from customer, they will be able to calculate more accurate and fair premium for each customer. Obtaining real time customer's data and analyze it are possible by utilizing IoT and big data. Therefore this study will explore about how to utilize IoT and big data to obtain and process customer's data, using insurance company as a case study.

2. LITERATURE REVIEW

2.1 Internet of Things

IoT is created to enable information and data exchange between objects. The objects are usually embedded with sensors technology to capture data from its surrounding [6]. Embedding devices with technology make it more intelligent, convenient, secure, safe and efficient [7]. IoT has been widely used for various cases. As example, Korea implements IoT to provide payment services, security services, mobile operators, heath care, and smart farm [8]. In bigger scale, IoT can be used to transform a city into a smart city. IoT can help to detect real time location and

utilization of public transportation. IoT can detect occupancy of parking spaces. IoT can detect traffic jam and violation of traffic rules. IoT can also detect some other conditions, such as weather conditions, air pollution, water contamination, and energy consumption [9].

IoT also have been specifically used to collect data from people's activity and health condition. Article [10] proposes a model to connect doctors and patients using web platform and IoT called EcoHealth. It utilize sensors in smart watch to capture data about heartbeat rate, body temperature, and blood pressure of the patients. Article [11] reviews smart watch functionality to collect data from the wearer. Smart watch can collect data about step counts, posture, heartbeat rate, and temperature using different sensors, such as accelerometer, gyroscopes heart rate sensor, and contact sensor. Article [12] mentions that IoT can bring benefits such as interconnectivity, faster communication and data exchange, especially in organization that need to quickly response to their customer, like in healthcare. Article [13] proposes an idea to monitor elderly person or patients that currently in medical care. The data from those people are collected to be analyzed so that they can get treatment based on each of their personal needs. Article [14] proposes architecture to monitor patient's condition through IoT. Patient's data are collected using sensors which are transmitted through internet and stored in the server. Article [15] proposes the use of IoT to track blood pressure using microcontroller in wearable device, which is connected to mobile phone. When blood pressure is higher than normal, it will trigger mobile phone to send SMS.

2.2 Big Data

Big data is almost ubiquitous and can also be applied in various businesses. A biotechnology company uses sensor data to analyze and evaluate about plantation result. It gathers information like temperature, soil, and water to determine the ideal ecological circumstance for create the plantation [16]. Big data can also be used in new product development. Customer's expectation and response of the new product are gathered in big data and analyzed to evaluate whether new product is sufficient for customer's needs [17]. Another example is big data can be used in education. It brings some benefits such as improved instruction, matching students to program, matching students to employment, and efficient system administration [18]. Hadoop is the most popular big data management technology that being used by organizations. Article [19] mentions that HDFS, YARN and MapReduce are some of the famous and cost-efficient technologies for big data. Article [20] discusses about the big data ecosystem to analyze tweets data. Big data management that is used is Hadoop, which has some components: HDFS, MapReduce, Pig, Sqoop and Zookeeper. Article [21] proposes big data architecture for intensive care system in a hospital. They use Hadoop as big data management to handle all data that come from different data sources.

In insurance and financial industry, big data can be used to support various business needs. Article [22] proposes the use of big data for fraud detection. They use Hadoop for big

data management, which have components such as HDFS and MapReduce. Article [23] proposes the use of big data analytics to leverage fraud detection. They use Hadoop as big data framework. Their analytical processes consists of data pre-processing, attribute selection, factor analysis and decision tree. Article [24] proposes an idea to transform insurance company from "cure" to "care" approach. By using IoT and big data, insurance company can monitor customer's health and then promote healthy lifestyle to the customer. Article [25] proposes the use of big data to help risk assessment in insurance company. Analytical processes consists of data pre-processing, visual analytics, dimensionality reduction, and decision tree. Article [26] proposes a meta-model of big data management by using MapReduce, Zookeeper, Hive, and Pig.

3. METHODOLOGY

The approach of this study is using Systematic Literature Review. It is performed by identifying, evaluating, and combining information from several other studies which relevant with this study's research questions. The sources of articles are ScienceDirect, ResearchGate, and other internet sources. Following steps are performed to those sources:

1. Explore articles about definition and benefits of IoT and big data.
2. Explore articles about IoT and big data implementation in various industries.
3. Explore articles about more specific implementation of IoT and big data to capture and process customer's data, especially the one related with insurance industry
4. Extract the method from those articles and develop it to propose a solution for the insurance company case study.

4. RESULTS

The proposed solution of this study combines IoT and big data to help insurance company to understand about their customer's risk. IoT is used to collect real-time data about customer's activities and health condition. Those entire customer's data are joined together with existing customer data from insurance company internal system into the big data management. All of those data will be processed and analyzed to give the meaningful information for actuarial team so they can calculate specific and personalized premium amount for that particular customer.

4.1 IoT Architecture

Collecting customer's activities and health condition in real time is very important to understand about their risk. To collect customer data, we utilize smart watch because it is already widely used by people and have the capabilities to monitor people activities such as steps taken, burned calories, heart rate, temperature, stress level, and sleeping habits. People can also take note of certain information by themselves, such as foods and water that they consumed. Due to popularity of smart watch, there are various fitness applications in the market which utilize smart watch to encourage people to do exercises and start doing healthy

lifestyle. Usually, these fitness applications requires the person using it to input physiological details such as their weight, height, gender to help the application make better sense of the data that it collects. Insurance company can follow this trend by develop an application that needs to be installed in the customer's smart watch. The customer must login to the application and input their profile so insurance company can recognize them. The application will utilize some sensors in smart watch to collect data. Following are the sensors that will be used by the application:

- Accelerometer: Accelerometer is used to continuously sense movements of the body. By using orientation and acceleration force, it can track movement in every direction, determine whether the device is horizontal or vertical, and whether it is actually moving or not. The common used of accelerometer is to detect and measure the exercises done by the person, for example, like whether the person are just walking or running, how fast the walking or running speed, and so on.
- Gyroscope: Gyroscope is used to detect orientation and rotation for various purposes. When its functions is combined with accelerometer, they can give more accurate representation of exercise's movement.
- Altimeter: Altimeter helps to capture information about person's elevation when moving in incline steps and climbing. Altimeter is commonly used by mountain climbers because it can measures the altitude while mountain climbing. It can be used to determine the elevation of people when they walking or running.
- Heart rate sensor: Heart rate sensor detect heartbeat rate by using LED light and optical sensor. It will determine the heart rate by rhythm of blood color-changes. By detecting a person's heartbeat rate, we can check if there irregularity on their body or mental state.
- Temperature sensor: Temperature sensor detects body temperature when it touches the skin. By detecting changes of body temperature, we can know whether the person is healthy or not.

The application will save customer's data from these sensors any time they detect and collect data from the customer. When the smart watch connects to internet, the application transmits customer's data to insurance company's system as shown in Figure 1. Insurance company's system to collect customer's data will have three functionalities:

- Data manipulation: Data manipulation provide functionality to translating the data collected from the smart watch in a specific context so it can be understood by the system.
- Data storage: Data storage provide functionality to store and maintain all of collected customer data such as the exercises (walking/running) and body vital signs (heart rate/temperature). Data storage can be database or file system.
- Common services: Common services provide functionality such as infrastructure security and user authentication. Customer data security and privacy is always the biggest concern and top priority in insurance industry.
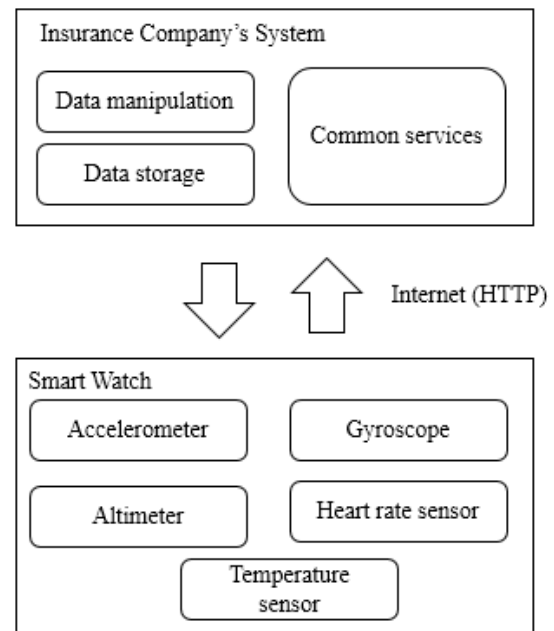

Figure 1: IoT Architecture [10]

4.2 Big Data Architecture

All of customer's data stored in insurance company's system later are sent into big data management to be used for analytics purposes. We are going to use Hadoop as big data management system because it is the most commonly used software for big data. Hadoop offers some important capabilities: it can store any kind of data from various sources in a large scale and with inexpensive costs. Hadoop can also do very sophisticated way of data analysis quickly and easily. Following are the main components of Hadoop that will be used:

- Zookeeper: Zookeeper's function is to maintain the configuration of all Hadoop's ecosystem. In this solution, Zookeeper is used to manage and connect between all Hadoop's components to make sure everything is running properly.
- Hadoop Distributed File System (HDFS): HDFS is a distributed file system which able to operate on commodity hardware. In this solution, customer's data stored in insurance company's system will be imported into HDFS.
- Yet Another Resource Negotiator (YARN): YARN is the resource management and job scheduling technology in Hadoop. YARN is the one which allocate system resources to various applications running in a Hadoop cluster and scheduling tasks to be running on different cluster nodes.
- MapReduce: MapReduce is software framework for programming applications which process vast amounts of data in parallel on large clusters of commodity hardware. Analytical process will use MapReduce because it will help to process the huge amount of customer's data.
- Sqoop: Sqoop is used to transfer data between Hadoop and database, which in this case is the database of insurance company's system. Sqoop will import data

from insurance company's system and store the data in HDFS.

- Pig: Pig's function to analyze the big data. Pig is built on top of MapReduce to utilize MapReduce for data processing. Script in Pig language will be used for big data analytics process of customer data.

Figure 2 shows the connection between insurance company's system and Hadoop to transfer and processing the customer's data. Sqoop is the component that import the customer data from insurance company's system and stored it into HDFS. When business user execute Pig script to perform analytical process, the process will run in parallel at the same time because of MapReduce. The parallel process of MapReduce is supported by YARN, which is the one who one allocate system resources. The analytical process make use of customer data which already stored in HDFS.
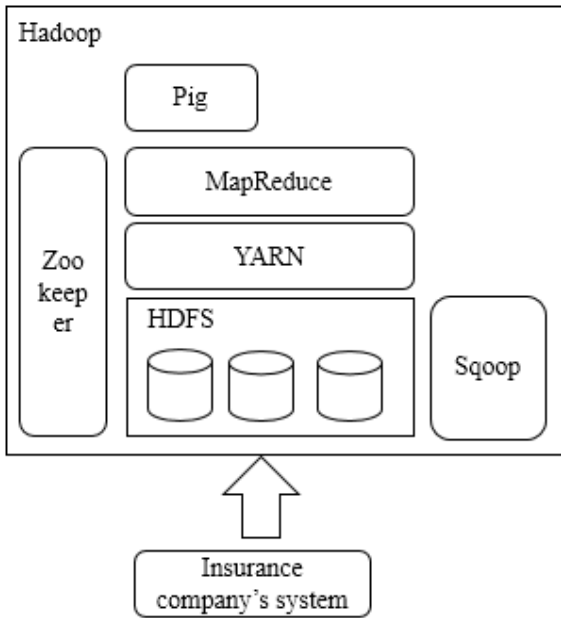

Figure 2: Big data Architecture [20]

4.3 Big Data Analytics

Analytical process is performed to transform the various customer data into a meaningful result that can be used by business to make a decision, which in this case are to understand customer's risk based on their activities and health condition. There are several steps to be performed in analytical process:

1. Data pre-processing: Data pre-processing is an act of cleaning data, which means that we want to remove unwanted data from the targeted data set. This step is challenging because of the presence of heterogeneity on the data and the occurrence of missing data and inconsistencies.

2. Attribute selection: This step aims to determine which data that may become the factors that have significant impact to the response variable. The selected data can also be visualized using charts or graphs to show the distribution of data set to have better knowledge of which prediction models that will be more suitable for the data set. For premium calculation case, we can use

attributes such as customer's age, height, weight, exercise duration, exercise frequency, heart rate, and so on.

3. Dimensionality reduction: This steps aims to reduce the number of variables in order to have efficient modelling. Dimensionality reduction can be divided into feature selection and feature extraction. Feature selection is a process to select the prominent variables. While feature extraction is a process to derive new variables from the existing variables, to increase the accuracy by eliminating redundant and irrelevant variables.

4. Decision tree: After we obtain the targeted data set, then we can build the predictive model. There are several techniques, one that is commonly used in decision tree. Decision tree is machine learning technique to find target value and check the other possibility of trends with the different branches. A decision tree consists of root node, branches, and leaf nodes to represent data in tree-like graph. The internal node represents performed test and branches represents the test's outcome. Figure 3 shows an example of decision tree graph that can be used to predict customer' risk based on their profile, activities, and medical information. Based on customer's risk we can determine premium calculation which is suitable for that particular customer.

The analytics result can be used by related business team based on their needs. In premium calculation case, actuary department can use the analytics result to understand the actual risk of the customer based on their activities and health condition. Active customer with good lifestyle and good health condition can get a cheaper premium than customer which is not active and has worse health condition.
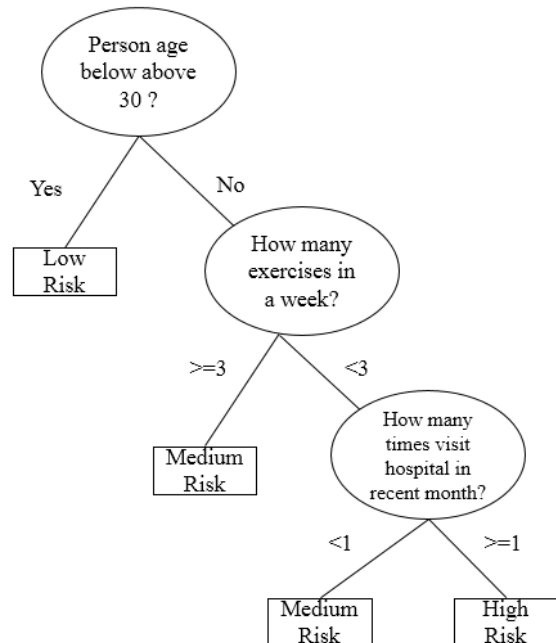

Figure 3: Example of decision tree to analyze customer's risk (Source: Author)

5. CONCLUSION

Current advance in technology era gives opportunities for business to evolve their practices. Business can utilize technology to create innovation in form of new product or new service. Long-established business industries, such as insurance industry, are no exception. Insurance industry is a business that heavily relies on historical data of customer to determine certain important aspects to run their business such as for the premium amount. With technology, insurance company can collect customer's data accurately in real time. IoT can be used by insurance company to collect actual customer data in real time. Customer data which processed in big data can help insurance company to analyze customer's risk. By knowing customer's risk, insurance company then can give more fair and accurate premium to the customer. Customers with healthy lifestyle and good health condition can get discounted premium, so they can pay less than other customers with worse lifestyle and health condition. The benefit will surely make customer happy and can be more engaged with the insurance company. Our proposed solution can be used a basis for insurance company to implement IoT and big data in their organization. We recommend this study to be expanded in the future. Such as, adding new technology in the architecture, or use it to solve other business problems. Further study to implement and evaluate the architecture is also much recommended.

REFERENCES

[1] N. Khan, I. Yaqoob, I.A.T. Hashem, Z. Inayat, W.K.M. Ali, M. Alam, M. Shiraz, and A. Gani, **Big Data: Survey, Technologies, Opportunities, and Challenges**, *Scientific World Journal*, 2014. https://doi.org/10.1155/2014/712826

[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A.H. Byers, **Big Data: The next frontier for innovation, competition, and productivity** available at http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation, 2011.

[3] Ling-yuan Zeng, **A Security Framework for Internet of Things Based on 4G Communication**, *Computer Science and Network Technology*, pp. 1715-1718, 2012. https://doi.org/10.2478/jamsi-2014-0008

[4] I.D. Luptakova, and M. Bilikova, **Acturial Modeling of Life Insurance Using Decrement Models,** *Journal of Applied Mathematics, Statistics, and Informatics*, vol. 10, no. 1, pp. 81-91, 2014.

[5] Joshua S. Adeyele, **Knowledge and Perceptions of Actuarial Science Among Students and Academics: Evidence from JABU**, *JABU International Journal of Social and Management Sciences*, vol. 3, pp. 11-20.

[6] M.U. Farooq, M.Waseem, S. Mazhar,A. Khairi, and T. Kamal. **A Review on Internet of Things (IoT)**, *International Journal of Computer Applications*, vol. 113, no. 1, 2015.

[7] S. Madakam. **Internet of Things: Smart Things**, *International Journal of Future Computer and Communication*, vol. 4, no. 4, pp. 250-253, 2015.

[8] Young-Mo Kam, Mi-Ran Han, Kyeong-Seok Han, and Jong-Bae Kim. **A Study of Internet of Things (IoT) Applications**, *International Journal of Software Engineering and its Applications*, vol. 9, no. 9, pp. 117-126, 2015.

[9] S. Talari, M. Shafie-khah, P. Siano, V. Loia, A. Tommasetti, and Joao P.S. Catalão. **A Review of Smart Cities Based on Internet of Things Concept**, *Energies*, vol. 10, no.4, 2017. https://doi.org/10.3390/en10040421

[10] P. Maia, T. Batista, E. Cavalcante, A. Baffa, F.C. Delicato, P.F. Pires, and A. Zomaya. **A Web Platform for interconnecting body sensors and improving health care**, *Fourth International Conference on Selected Topics in Mobile & Wireless Networking*, 2014.

[11] B. Reeder, PHD, and A. David, BS. **Health at hand: A systematic review of smart watch uses for health and wellness**, *Journal of Biomedical Informatics*, vol. 63, pp. 269-276, 2016.

[12] C.E. Turcu, and C.O. Turcu. **Internet of Things as Key Enabler for Sustainable Healthcare Delivery**, *The 2nd International Conference on Integrated Information*, 2013.

[13] S.V. Zanjal, and G.R. Talmale. **Medicine Reminder and Monitoring System for Secure Health Using IOT**, *International Conference on Information Security & Privacy*, 2015. https://doi.org/10.1016/j.procs.2016.02.090

[14] J. Gomez, B. Oviedo, and E. Zhuma. **Patient Monitoring System based on Internet of Things**, *The 7th International Conference on Ambient Systems, Networks, and Technologies*, 2016. https://doi.org/10.1016/j.procs.2016.04.103

[15] E.B. Panganiban. **Microcontroller-based Wearable Blood Pressure Monitoring Device with GPS and SMS Feature through Mobile App**, *International Journal of Emerging Trends in Engineering Research*, vol 7, no. 6, pp. 32-35, 2019. https://doi.org/10.30534/ijeter/2019/02762019

[16] S. Soomro, A.W. Jumani, and M.A. Kartio. **Big Data Analytics and Its Applications**, *Annals of Emerging Technologies in Computing*, vol. 1, no. 1, pp. 45-54, 2017.

[17] Y. Zhan, K.H. Tan, Y. Li, and Y.K. Tse. **Unlocking the Power of Big Data in New Product Development**, *Ann Oper Res*, 2016.

[18] F. Almeida. **Benefits, Challenges and Tools of Big Data Management**, *Journal of Systems Integration*, vol. 8, pp. 12-20, 2017.

[19] D.A. Koutsomitropoulos, and A.K. Kalou. **A standards-based ontology and support for Big Data Analytics in the insurance industry**, *ICT Express 3*, vol. 3, pp.57-61, 2017.

[20] C. Uzankaya, T. Ensari, and Y. Kavurucu, **Hadoop Ecosystem and Its Analysis on Tweets**, *World Conference of Technology, Innovation and Entrepreneurship*, 2015 https://doi.org/10.1016/j.sbspro.2015.06.429

[21] A, Gonçalves, F. Portela, M.F. Santos, and F.Rua. **Towards of a Real-time Big Data Architecture to Intensive Care**, *International Workshop on*

Healthcare Interoperability and Pervasive Intelligent Systems, 2017

[22] Ana-Ramona Bologa, R. Bologa, and A. Florea. **Big Data and Specific Analysis Methods for Insurance Fraud Detection**, *Database Systems Journal*, vol. 1, no.1, pp. 30-39, 2010.

[23] P. Dora, and Dr. G.H. Sekharan. **Healthcare Insurance Fraud Detection Leveraging Big Data Analytics**, *International Journal of Science and Research*, vol. 4, no. 4, pp. 2073-2076, 2015.

[24] A. Silvello. **IoT and Connected Insurance Reshaping the Health Insurance Industry. A Customer-centric "From Cure to Care" Approach**,

EAI Endorsed Transactions on Ambient Systems, vol. 4, no. 15, 2017.

[25] N. Boodhun, and M. Jayabalan. **Risk Prediction in Life Insurance Industry using Supervised Learning Algorithm**, *Complex & Intelligent Systems*, vol. 4, no. 2, , pp. 145-154, 2018.
https://doi.org/10.1007/s40747-018-0072-1

[26] A. Erraissi, and A. Belangour. **Meta-modeling of Big Data management layer**, *International Journal of Emerging Trends in Engineering Research*, vol 7, no. 7, pp. 36-43, 2019.
https://doi.org/10.30534/ijeter/2019/01772019