Design of Phishing Simulation Dashboard Using Analytic Data Concepts

R Septiana¹, R K Julian²
¹,²Dept. of Computer Engineering, Universitas Diponegoro, Semarang, Indonesia
Email: rismaseptiana@live.undip.ac.id

Abstract. Security awareness is the main concern of phishing attack prevention. Every internet user will deal with a serious issue. Phishing is the falsification of messages that aims to retrieve important information from the recipient of the message. In a company, phishing becomes a serious problem because it can come to all employees suddenly. Employees who do not understand the dangers of phishing are the most dangerous threats because they are likely to be the main supporters of the success of phishing attacks. Understanding of phishing attack can be seen from the way the worker responds to incoming messages. A company makes a phishing simulation to find out the handling of the employee on a fake message came. One of the phishing simulation tools is web forgery. In this research, phishing simulations attack every employee in a company. The employee will receive a message that contains a field. The result will show a percentage of employee that successfully attacked. The results will be displayed on a dashboard after being processed using an analytic data concept. The processes of data analytic involve business intelligent software called QlikView. The software gives a percentage of the employee who opens the fake message and fills some data in the fields is about 13.9%.

1. Introduction
In this globalization era, the rate of data exchange increases very quickly, both raw data and metadata that are already in the form of information. In the field of data communication, the development of information technology that is constantly changing and security is an important issue, neither securities of the physical equipment nor data and applications. The issue of secret data security is one important aspect that needs to be kept such as user names, passwords, Personal Identification Number (PIN), and others that are the personal secret[1]. Nowadays, there is still a lot of awareness of users who are still laid to maintain the secrecy of the information[2]. Sometimes they are trapped by unauthorized parties who intentionally send an illegal e-mails using a domain name from a legal organization to get important information from the target. The fraud incident is called phishing. Phishing often attacks employees in various companies. The aim of the attacker is to get some important secret information from the personal employee. Therefore, it is important for companies to provide phishing simulation, so the company can monitor the way employees deal with phishing problem[3]. From the monitoring result, the company will take decisions that increase the awareness of employee in a company in dealing with phishing that enters their e-mails. Monitoring data of the phishing simulation result can be easily done using a dashboard. The processing of valid monitoring data uses an analytic data method. The data entered through the phishing simulation process is processed and transformed into an information chart. The dashboard displays the information concluded from the chart using percentage and ratio of the employee caught in the phishing. In this context, a presentation of data about phishing simulations that are easy to
Phishing simulation activities and data analytic processes are two separate research sections. This research tries to combine both sections, so a company can use the result of phishing simulation activities as raw data to make information used to make a decision related to security awareness.

A number of previous research about Phishing explain various phishing technique. Based on information from the AWPG (Anti-Phishing Work Group) are increasingly showing. Phishing techniques can be divided into two types of Technical Subterfuge and Social Engineering. Assault on a company that is often used is a social engineering technique. The complicated issue of phishing attack causes the problem to wipe out incompletely[5]. There are two solutions to overcome the problem; first, developing the reliable anti-phishing technique and the second, educating people about phishing awareness. Previous research about data analytics discusses various tools used to make valid information[4]. QlikView is one of the excellent tools to process the data analytics method and display the result on the dashboard[6]. QlikView one of the data visualization tools implemented in business intelligence and used to analyze business performance. The result of business analysis could use to forecast the next decision in the business process[7].

This research use QlikView as a tool to analysis phishing simulation result using a dashboard display. The company can analyze the result easily through the chart in the dashboard display.

2. Research Methodology

2.1. Phishing Simulation

Phishing is a fraudulent attempt to obtain sensitive information such as user names, passwords, and Personal Identification Number (PIN) details, often for malicious reasons, by posing as trustworthy entities in electronic communications. According to the 2013 Microsoft Computing Safety Index, released in February 2014, the annual global impact of phishing could reach US $ 5 billion. Phishing can be done in many ways with the aim of directing users to enter personal information on faked websites, the look and feel of a legitimate site. Communications that claim to come from social websites, auction sites, banks, online payment processors, or IT administrators are often used to lure victims. Phishing emails can contain links to websites that distribute malware. From various techniques to do phishing, a company uses Website Forgery techniques to simulate phishing of its employees. Through Website Forgery techniques, users as victims who visit phishing pages cannot know for sure, whether the page is genuine or faked, because the page will be made in the same way as the original website[8].

2.2 Business Intelligeent Implementation

Business Intelligence (BI) is a collection of techniques and tools to transform raw data into useful and meaningful information for business analysis purposes. BI technology can handle large amounts of unstructured data to help identify, develop, and otherwise create new business strategy opportunities. The purpose of BI is to facilitate the interpretation of such large amounts of data, the information provided by BI can certainly facilitate decision making because the information presented in BI was more effective and understand easily. Qlikview is Business Intelligence (BI) software founded by the company QlikTech. QlikTech has focused on simplifying decision making for business users throughout the organization. In general, existing BI tools still use traditional concepts, where the concept has existed since the era of the 80s. QlikView, as BI software, has a very different concept from traditional BI. QlikTech is pioneering a new approach to accessing, managing and interacting with data. The QlikView business is a platform invention that is recognized as an innovative solution. Combined with a relentless focus on customer and community success, it's no wonder that there are more than 26,000 companies in more than 100 countries using Qlikview, with a leading industry satisfaction rate of 96% [9][10]. QlikView Features:

a. Consolidation
The patented QlikView in-memory technology makes QlikView able to quickly retrieve data from various sources. QlikView can easily handle data from Oracle, SAP, Salesforce.com, Business Objects, Cognos, Hyperion, SQL Server, MySQL, Postgre SQL, AS/400 or Excel.

b. Search
   When words or letters are typed into the search function in QlikView, QlikView will immediately get the results and see how the relationships between each data are interconnected from various sources.

c. Appearance
   We are quickly able to retrieve and display data in a chart on QlikView. Make a pivot table, bar chart, speedometer, and make adjustments to make it look more attractive. With dynamic, you can directly navigate on the chart.

2.3 Data Analytics
Data analytics is a data processing method that is able to unite the substance of each data collected from various sources to be transformed into a big picture of critical information. Data analytics technology and techniques are widely used in the commercial industry to enable organizations to make more informed business decisions.
As a term, data analytics refers to a variety of applications, from basic Business Intelligence (BI), Reporting and Online Analytical Processing (OLAP) to various forms of advanced analytics.
Based on the results, analytics data is divided into four types, namely[11]:
   a. Descriptive Analytics
      Descriptive Analytics is a process of data analytics to get a general picture of the data that has been collected.
   b. Diagnostic Analytics
      At this stage, historical data can be measured with other data to answer the question of why something happened. Thanks to Diagnostic analytics, it is possible to browse, find out dependencies and identify patterns.
   c. Predictive Analytics
      Predictive Analytics is data analytics that provides predictive results about something to come.
   d. Prescriptive Analytics
      Prescriptive Analytics is an analytics process that generates answers to questions about why something happened and provides suggestions for conditions that might occur in the future.

2.4 Data Visualization
Data visualization is seen in many fields of science as modern visual communication. Data visualization is not under any field, but rather an interpretation among many fields (for example, it is sometimes seen as a modern branch of descriptive statistics by some people, but also as a basis for developing tools by others). Data visualization involves the creation and study of visual representations of data, meaning "information that has been abstracted in a schematic form, including the attributes or variables of the information unit". The main purpose of data visualization is to communicate information clearly and efficiently to users through selected information charts. Effective visualization helps users in analyzing and reasoning about data and evidence, which then makes complex data accessible, understood and useful. Users can do certain analytical work, such as doing comparisons or understanding causality[12].
3. Result and Discussion

This research implements analytics data process and phishing simulation dashboard in the PT Pertamina (Persero). Phishing Simulation Dashboard is an instrument board that functions to display data about phishing simulation in the form of bar charts experienced by employees of PT. Pertamina (Persero). This instrument board is used to make it easier for officials and employees of PT. Pertamina (Persero) in knowing the level of awareness of phishing actions by electronic mail disguised as authorized persons or organizations sent to their personal e-mails and seeing patterns or interrelations in data for one or more variables formed in this phishing simulation data. The result displayed in the dashboard is the result of analytics data process using QlikView.

3.1. The Main Component

QVD data is the main component of the phishing simulation dashboard using the QlikView analytical data tool. The function of the data is to process of loading data on QlikView more quickly and reduce the memory usage during the data loading process. QVD consists of two tables, SOURCE_DATA from a data source whose name begins with 2018 and TOTAL_REPORT from a data source named ReportedTicket. The process of preparing QVD data is appeared in the Figure 1.

![Figure 1. Preparing QVD Data](image)

Figure 1 shows that in SOURCE_DATA.QVD contains basic information about employees who are targeted in phishing simulations such as email, ID, name, position, and status. The TOTAL_REPORT.QVD contains the date, month, and year, and the total number of employees who reported being trapped by this phishing simulation.

3.2. Simulation Progress Report

Simulation Progress Report is the first step to get the chart result. The process of creating a chart, previously started by entering the QVD data into the script on the sheet as well as creating a new table called Master Calendar which will be used in QlikView to create a chart.

![Figure 2. Table View Displayed](image)
Figure 2 explains three contents of table view. The Ms_Calendar table is created from the Date column in the SOURCE_DATA and TOTAL_REPORT tables. This Ms_Calendar table was created as a date reference to make it easier to create charts.

![Simulation Progress Chart](image)

**Figure 3.** Progress Simulation Chart

Figure 3 displays the amount of each data in the status column of SOURCE_DATA.QVD such as Email Clicked, Email Opened, Clicked Link and Submitted Data and the sum of the Report Amount column from TOTAL_REPORT.QVD. In the Simulation Progress chart, there are dates, months, and years on the X-axis and on the Y-axis the total data of employees affected in this phishing simulation is displayed. In the blue bar, it represents the total number of e-mails sent to the target. The red bar represents the total number of employees who opened the phishing simulation e-mail. The green bar represents the total number of employees who opened the link containing phishing. The yellow bar represents the total number of employees filling the data on the link that contains phishing, and on The Tosca bar represents the total number of employee e-mails reported due to getting caught in a phishing simulation.

### 3.3 Click and Report Rate Result

This chart, displays data about Simulation Clicked and Simulation Reported that will be calculated further in the Expression window of the status column in SOURCE_DATA.QVD such as Email Clicked, Email Opened, Clicked Link, and Submitted Data.

![Click and Report Rates Chart](image)

**Figure 4.** Click and Report Rate Result

In Figure 4, there are dates, months, and years on the X-axis. On the Y-axis the percentage of the comparison of the number of employees who clicked on a phishing simulation but did not fill in the data presented on the blue bar. The number of employees who clicked against phishing simulations as well as filling in the data presented on the red bar.
Figure 5 shows the Clicked Simulation object is the result of an average calculation every six months from the status column in the SOURCE_DATA table, calculated from the last month available in the table.

3.4 Report To Click Ratio Result

This chart is the simplest chart, because this chart only shows data about the Report-to-Click ratio which will be calculated further in the Expression window of the Clicked Link data and Submitted data from the status column on SOURCE_DATA.QVD and the Number of Reports data in TOTAL_REPORT AMOUNT.QVD.

Figure 6 displays the Report-to-Click Ratio chart. In the Report-to-Click Ratio chart, there are dates, months, and years on the X-axis and on the Y-axis the percentage of comparison results or the ratio between the number of employees who clicked on a phishing simulation but did not fill in the data and the number of employees who clicked on phishing simulation and data filling.

Figure 7 shows Text Object Report-to-Click Ratio. This Report-to-Click Ratio object is the result of the average calculation for each of the last six months from the Number of Report columns as well as
Click Link and Submitted Data from the status column in the SOURCE_DATA table, calculated from the last month available in the table.

4. Conclusion
Phishing Simulation Dashboard is used to see and increase the awareness of the employees of a company in the face of phishing entering their e-mails. The phishing simulation instrument board is included in the data analytics with a descriptive analytics type because the instrument board contains a general description of the data that has been collected. Through the method of data analytics and Business Intelligence (BI) technology, the data is presented in a very informative and effective way by connecting various variables. Data visualization in the form of bar charts helps communicate information clearly and efficiently and helps users analyze and reason at the same time about data and evidence.

5. References
[1] J. Mao, “Phishing page detection via learning classifiers from page layout feature,” 2019.
[2] J. Milletary, “Technical Trends in Phishing Attacks,” pp. 1–17.
[3] S. Nisha and A. N. Madheswari, “Secured Authentication For Internet Voting In Corporate Companies To Prevent Phishing Attacks,” vol. 22, no. 1, pp. 45–49, 2016.
[4] S. Baadel, “Data Analytics : intelligent anti-phishing techniques based on Machine Learning 12,” no. April, 2019.
[5] B. Sharma and P. Singh, “PhishAlert : An Efficient Phishing URL Detection via Hybrid Methodology,” no. 12, pp. 3936–3941, 2019.
[6] A. Shukla and S. Dhir, “Tools for Data Visualization in Business Intelligence : Case Study Using the Tool Qlikview.”
[7] R. J. Podeschi, “Experiential Learning using QlikView Business Intelligence Software,” vol. 13, no. July, pp. 71–80, 2015.
[8] I. W. Jr, “Website Forgery : Understanding Phishing Attacks & Nontechnical Countermeasures,” 2015.
[9] Feby Artwodini Muqtadiroh, R. Tyasnurita, and A. F. Sholeh, “Analisis Perbandingan Pengembangan Dashboard Sistem Dengan Menggunakan Fusion Chart, Qlikview Dan Pentaho,” in Seminar Nasional Sistem Informasi Indonesia, 2013, pp. 510–515.
[10] S. Redmond, QlikView for Developers Cookbook. Birmingham: Packt, 2013.
[11] Alex Bekker, “4 Types of Data Analytics to Improve Decision-Making,” 2019. [Online]. Available: https://www.scnsoft.com/blog/4-types-of-data-analytics. [Accessed: 05-Dec-2019].
[12] B. U. Dur, “Data Visualization and Infographics In Visual Communication Design Education at The Age of Information,” no. June, 2016.