Article | Received 23 July 2023; Accepted 2 August 2023; Published 21 December 2023
https://doi.org/10.55092/pcs2023020043

Digital advertising fraud prediction using OLS regression

Kah Mun Ng [1], Choon Sen Seah [1,*], Wai Yin Chin [1,*], Yin Xia Loh [1], Ahmad Najmi Amerhaider [2], Farah Waheeda Jalaludin [1]

[1] Faculty of Accountancy and Management, Universiti Tunku Abdul Rahman. Sungai Long Campus, Jalan Sungai Long, Bandar Sungai Long Cheras 43000, Kajang, Selangor, Malaysia

[2] Faculty of Computing, Universiti Teknologi Malaysia, N28A, Lingkaran Ilmu, Johor Bahru 81310, Johor, Malaysia

* Correspondence authors; Email: seahcs@utar.edu.my, chinwy@utar.edu.my.

Abstract: Digital advertising has become an essential tool for every business. The digital advertising budget has been increasing over the years. More businesses have turned to digital advertising during the pandemic. However, beginner advertisers might lack knowledge of digital advertising and, at the same time, pour extra capital into it. Furthermore, fraudulent activity in digital advertising is also increasing, which harms the current digital marketing environment, as well as every party involved. This study examines the factors that affect conversion fraud in digital advertising. A sample of 956 observations of computed-generated data is used to examine the variables related to conversion fraud. The results show that advertisers, ad logs, items, goals, and ad slots have a positive relationship with conversion fraud, which these variables determine the digital advertisement fraud. The predicted value for conversion fraud is 0.2936, which 280 of the samples are predicted to be fraud digital advertisement. Implications and recommendations for this research were discussed to facilitate the advertisers in future advertisement placement.

Keywords: digital advertising; ordinary least squares (OLS); clients; ad slot ID; item; goal; ad log type; pricing type; look up form

1. Introduction

Advertising is an important tool for every business, as it is the main tool to get exposure to their target customer. From traditional advertisement such as printed magazines, newspapers, and television until today, digital advertisement has specific places assigned to be sold for advertisement based on users search results, posts, or the content of the website [1]. Nowadays, people are heavily relying on the internet to get information, and thus, digital advertising is widely used by businesses to promote their goods and services. Digital advertising can help businesses promote their goods and services in an efficient and effective

way. Advertising serves as a way of communication to inform and persuade an audience to make a purchasing decision regarding goods or services [2].

Conversion fraud is a potential threat that cannot be controlled and measured by advertisers [3]. It has a negative impact on advertisers in terms of cost and data information [3,4]. Conversion fraud can be caused by a few factors, such as unethical website publishers' and users' actions by accident [4]. Unethical advertising behavior affects the digital advertising environment and is unfair to advertisers [5]; publishers charge advertisers for every user's action, such as clicks, views, and leads [4,5].

Conversion fraud includes the following:

(1) Pixel staffing involves the insertion of multiple ads into a single pixel space, which results in advertisers being charged for impressions that are not viewable by users due to their tiny size [6].
(2) Ad stacking: multiple ads are stacked on top of each other in the same ad slot, but only the top ad is visible to users [7,8].
(3) Domain spoofing: fraudsters deceive users and advertisers by disguising their actual website link as that of a premium website [9].
(4) Ad injection: unauthorized ads are displayed on a website without the publisher's knowledge or permission [6].
(5) Click injection: fraudsters inject fake clicks on ads to charge advertisers for clicks that never actually happened [5].
(6) Geo-masking, also known as location masking, is a fraudulent technique that exploits the notion that traffic from certain regions or countries is more valuable than others [9].
(7) Bot traffic refers to non-human traffic that is generated by software robots (bots). It can be programmed to perform a variety of tasks, such as crawling web pages for search engines, automating repetitive tasks, and even generating fake traffic to websites and apps [6].

Conversion fraud can lead to firms spending excessive amounts of money on advertising, which becomes an expense rather than an investment. This is especially true among Small and Medium Enterprises (SMEs) that lack knowledge about digital advertising specifically when they are struggling in considering of the adoption of information and communication technology (ICT) and e-commerce [10].

2. Literature review

2.1. Digital advertising

Digital advertising, or internet advertising, that comprises of online advertising and mobile advertising, impacts monetization throughout the internet environment [11]. Today, more and more retailers are using social networking sites (SNS) to run their businesses, creating a platform that allows online shoppers a convenient way to shop for a variety of products and services [12]. Digital and social media marketing is a new industry-led, research-informed, and result-driven drive for e-commerce [13]. Digital marketing has become increasingly important to firms because it promotes the development of a new avenue for reaching target

users while at the same time satisfying them with products and services [14]. The basic concept of digital advertising is to sell space on web pages and apps for advertising [15]. Digital advertising offers exceptional innovations to marketers and enables businesses to advertise to specific target markets that most closely correspond to their target audience [16]. Users can see the ads that meet their interests and areas of interest. Digital advertising offers businesses greater adaptability to customer responses. Since the digital environment allows the advertisement to spread efficiently, the users could respond to the advertisement immediately [17].

The following variables are selected through variables filtration to determine their relationship to conversion fraud. These variables have to be integer values to fulfill the criteria.

2.2. Review of variables

2.2.1 Clients

Clients, also known as advertisers, are a critical component of the digital advertising industry. They are typically individuals, companies, or organizations that pay to have their ad displayed on publishers' websites [18]. Advertisers purchase ad space from publishers, which allows them to reach their target audience with the right message at the right time. For example, a gaming app may purchase ad space from a website publisher to promote a special offer to users who are interested in gaming. The success of the advertising campaign is often measured using metrics such as CPC (cost per click), CPV (cost per view), and CPM (cost per thousand impressions), which can provide insight into how users interact with the ads. The advertiser is the main character in most of the cases of conversion fraud. While the advertiser is one of the independent variables in conversion fraud, their brand image is important because it affects the consumer’s purchase decision. Brand image is a long-term process of development, and it serves as a weapon to compete in the industry. Based on [19], assuming the brand is managed well, the firms can gain many consumers, which leads to a long-term profitable relationship. Nowadays, most people care about status; they use branded products as their status symbols. This shows people prefer to purchase products with a well-known brand name. Furthermore, brand image is about enhancing consumers memories regarding the brand, which will affect their final purchase [20]. Thus, despite the attractiveness of the ad, it might not be able to convert to sales since the advertiser's brand image is not built.

H1: There is a relationship between advertiser and conversion fraud.

2.2.2 Ad slot ID

Ad slots are critical for website publishers and advertisers as they determine where ads are displayed on a webpage. The location of an ad slot can impact the success of an ad campaign. For example, an ad placed in a prominent ad slot, such as above the fold or in the header, is more likely to be noticed by users and generate clicks than an ad placed in a less visible

location, such as in the footer or sidebar. Ad slots may also be sold to advertisers based on different pricing models, such as CPM, CPC, and CPV. Advertisers may compete for a particular ad slot, and the highest bidder may be awarded the slot. In the case of video ads, an ad slot can refer to a specific time segment within a video that is reserved for an advertisement. For example, a 10-second ad slot may be inserted before a video begins playing or in the middle of a longer video. The success of a video ad campaign can be measured by metrics such as view-through rate and engagement rate. Overall, ad slots are crucial for both publishers and advertisers in ensuring that ads are displayed in the most effective and visible way to generate engagement and conversions [21].

Fraudsters take advantage of ad slots to execute fraudulent activities. In the cases of ad stacking and pixel stuffing, fraudsters overlap multiple ads on top of each other in the same ad slot on the webpages. While an ad slot will become vulnerable when it is manipulated by fraudsters to serve fraudulent ads to generate fake data, for instance, fake impressions and clicks, while paid by the advertiser, are invalid for the advertiser at the same time, which might affect their decision-making. Hence, it is important to monitor ad slots on webpages that have the potential to cause conversion fraud and harm the digital advertising environment [7].
H2: There is a relationship between ad slot and conversion fraud.

2.2.3 Item

The item is the ad creatives, which are the actual content that users see when an ad is served on a webpage or application. They can include a wide range of media types, such as images, videos, and interactive HTML5 elements [18]. Ad creatives are essential to the success of an advertising campaign because they deliver the message to the audience. Advertisers use different creative elements to appeal to their specific target audience and ensure that their message is delivered in an effective and engaging way. Ad creatives can be designed for specific platforms and devices to ensure that they are optimized for the best possible user experience. It can be created using different tools and software, including third-party ad servers, campaign managers, and native ad formats.

The item is the ad creative that consists of text and video; a typical advertisement contains the title, description, and website address (URL) of the advertiser. The creative context plays an important role because it conveys information about the brand, products, and services that aim to reach potential consumers. According to the study, there is a relationship between ad creatives and the effectiveness of listing in the sponsored search market. This shows the direct relationship to the ad's performance: the better the ad's creative, the higher the effectiveness of listing in the sponsored search market, and thus, the higher the sales and leads [22].
H3: There is a relationship between item and conversion fraud.

2.2.4 Goal

Advertisers may also have more specific goals that align with their business objectives. For example, an advertiser may want to increase their social media following, drive more sign-

ups for their newsletter, or promote a particular product or service. By setting specific goals for their advertising campaigns, advertisers can tailor their strategies and ad creatives to achieve the desired outcomes. Furthermore, advertisers may also have different target audiences that they want to reach with their advertising campaigns. For instance, an advertiser may want to target a specific age range, gender, location, or interest group. By using the targeting options available on advertising platforms, advertisers can ensure that their ads are shown to the right audience, which increases their chances of achieving their advertising goals. Ultimately, the success of an advertising campaign depends on a variety of factors, including the chosen advertising platform, the targeting options used, the ad creatives, and the chosen advertising goals. With carefully planned and executed advertising campaigns, advertisers can effectively reach their target audience and achieve their desired outcomes [23].

The goal is the objective that the advertiser wishes to achieve, and with the appropriate setting of the ad campaign, it acts effectively to reach the target audience. In digital advertising, advertisers set campaigns with many goals for their ads [24]. Setting advertising goals can benefit companies in many ways, such as improving brand recognition, increasing sales, building customer loyalty, and driving website traffic. By defining specific goals, businesses can create targeted and effective advertising campaigns that align with their marketing strategy. Some common advertising goals include increasing brand awareness, generating leads, boosting sales, promoting a new product or service, building customer trust and loyalty, and increasing website traffic or social media engagement. By setting up the right goals, advertisers can measure the success rate based on the budgets and the advertisement assets [25].

H4: There is a relationship between goal and conversion fraud.

2.2.5 Ad log type

The ad log records chronological documentation of any activities that affected a particular operation or event. In computing, a log refers to an automatic record of relevant events with timestamps for a particular system, which is produced by most software applications and systems [26]. It is a record of an ad that occurred at a specific time, with additional information to provide context. It captures all system activity, including transactions, errors, and security breaches. The data contained in log files may be structured or unstructured, and it can be transmitted in various ways. Hence, an ad log can be useful for keeping track of ad data, emergency recovery, and application improvement. Proper log management is important due to its ability to manage the sheer volume of data, absorb and derive valuable insight, and in terms of digital transformation [25]. By analyzing ad logs, advertisers can gain valuable insights into their advertising activities, identify areas for improvement, and take action to prevent fraudulent activities [27].

The ad log is the dataset of advertisements, which records the advertisement activities in terms of click, impression, and view. According to [28], the ad log dataset has been taken as the essential elements that affect consumer purchasing decision-making in the context of

previous purchasing habits and a sequence of advertisements viewed, such as the number of advertisements clicked and viewed. It proves that the higher the click, the higher the accuracy of the result of the purchase action taken by consumers. It records all the data of an ad, which could also be a powerful tool for detecting conversion fraud, such as when the number of clicks does not align with the actual sales. The ad log data set provides valuable insight into the performance of an ad campaign.
H5: There is a relationship between ad log type and conversion fraud.

2.2.6 Pricing type

Setting the right price is crucial for any business, as it directly impacts on their profitability and market share. There are five common pricing strategies: cost-plus pricing, competitive pricing, price skimming, penetration pricing, and value-based pricing. Pricing strategies are important not only to the firms themselves but also to their target consumers. It should be well calculated to make profit as well as to be aligned with the products and services [29]. When setting prices, businesses must also consider the elasticity of demand, or how sensitive customers are to changes in price. For example, if demand for a product is highly elastic, a small increase in price may cause a significant drop in demand. Businesses must also consider their costs, including fixed costs such as rent and variable costs like materials and labor, to ensure that their prices are profitable. In summary, pricing is an important aspect of any business strategy and must be carefully considered to ensure profitability and market share. The amount of money set by a company in exchange for a good or service, the price of goods and services may be sensitive to the needs of the consumer. Consumer behavior is significantly influenced by consumer perception. Price is a key aspect of how consumers perceive products and services.
H6: there is a relationship between pricing type and conversion fraud.

2.2.7 Look up form

A lookup form is a type of form used in a database or software application that allows users to search for and retrieve specific data from the database. It typically includes fields or criteria for entering search parameters, including keywords or dates, and displays results. A lookup form can be useful for quickly finding and collecting specific information within a large dataset. It usually helps users select records from a related table, and it is automatically added once a lookup column is added to the form. A lookup form can be a useful tool for advertisers since it enables them to gather and access specific data about their target market or future clients. Advertisers can quickly search through a database using a lookup form to identify pertinent data, such as demographics, interests, or purchase patterns, and use that data to construct customized ad campaigns. The table can either be filled with values manually by the programmer or it can be populated by the program itself while it is calculating those values. In the context of the advertising field, it is widely used by advertisers to collect users' information about who are their target audience or potential consumers [30].

Advertisers gather and access particular data about their target market or future clients through a lookup form. It allows advertisers to search through a database using a lookup form to identify pertinent data, such as demographics, interests, or purchase patterns, and utilize that data to construct customized ad campaigns in an efficient way. While it might also be used by fraudsters for fraudulent activities. Fraudsters can use lookup forms to collect users' personal information or payment details when they interact with a legitimate business. This harms not only the user but also the reputation of firms, and it can result in financial losses for the legal owner [30].

H7: There is a relationship between look up from and conversion fraud.

3. Hypotheses development

Conversion fraud includes various activities that deceive the advertiser by misrepresenting advertising inventory or hiding machines as humans, which lead to misused advertising expenditures. A conversion in an ad network denotes one or a set of meaningful business actions taken by users, which they are potential consumers in converting to a sales transaction. Alternatively, a conversion can also be defined as an agreed-upon action taken by a user". For example, a simple conversion event can be the downloading of a file, filling out a form, or the completion of an online purchase order. Such fraud is also called conversion spam. After users click on the displayed advertisement, they are normally directed to a landing page that shows summarized information about the advertised product or services [7]. Hence, these variables are constructed to formulate the conceptual framework for this research. Figure 1 illustrates the proposed conceptual framework with the hypotheses.

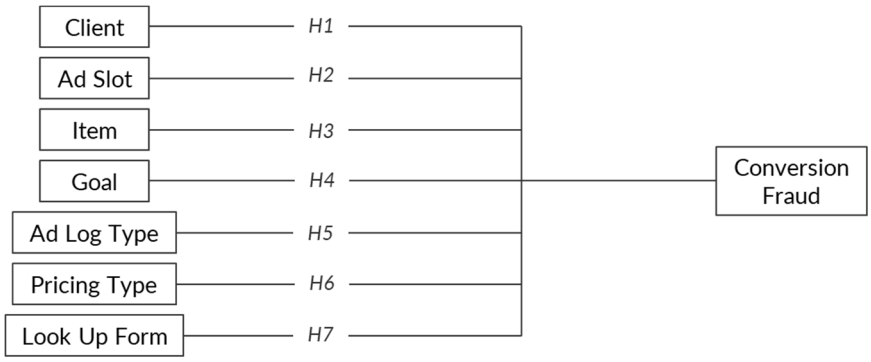

Figure 1. Proposed conceptual framework.

4. Methodology

This study utilized secondary data obtained from Kaggle, a data science platform, which is publicly available and collected by others [31]. The data, consisting of 965 observations with 54 variables, had already undergone processing and normalization by the author [32]. The study employed Ordinary Least Squares (OLS) regression to identify seven variables deemed most relevant to the dependent variable. The lack of information on target population, sampling frame, and technique is acknowledged due to the nature of using pre-processed

secondary data. The process of variable selection involved statistical testing and evaluation within the regression analysis.

4.1. Descriptive analysis

Pearson Correlation is used in this study to measure the strength of the connection between independent and dependent variables. If the value is negative, there is an inverse relationship between the variables, while if the value is positive, there is a direct relationship between them. In other words, this technique helps to determine how much influence the independent variable has on the dependent variable [33].

Spearman Correlation is also used for this research. The Spearman correlation coefficient is a statistical measure that helps us understand the strength and direction of the relationship between two variables. It is denoted by "rho" or "rs" and can range from -1 to +1. A value of -1 indicates a perfect negative correlation between the two variables, which means that as one variable increases, the other variable decreases in a perfectly predictable way.

A value of +1 indicates a perfect positive correlation between the two variables, which means that as one variable increases, the other variable also increases in a perfectly predictable way. A value of 0 indicates no correlation between the two variables, which means that there is no predictable relationship between them. The formula for calculating the Spearman correlation involves ranking the values of each variable, finding the differences between their ranks, and then using these differences to calculate the correlation coefficient [34].

Ordinary least squares (OLS) regression analysis is the most common regression method used in most social science studies. It is a statistical technique used to investigate the correlation between a dependent variable and one or more independent variables. The objective of OLS regression is to identify the line or equation that is most suitable for the data, thereby enabling the prediction of the dependent variable based on the values of the independent variables [35].

The regression line used here is:

$$\hat{Y}i = CONST + (X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7) \quad (1)$$

$$X_1 = 0.004,\ X_2 = 0.001,\ X_3 = -0.001,\ X_4 = -0.001,\ X_5 = 0.396,\ X_6 = 0.003,\ X_7 = -0.302$$

where $X_{1,2,3,4,5,6,7}$ refers to the significant value of advertiser, ad slot, item, goal, ad log, pricing and look up form in Spearman Correlation. relationships to the dependent variable.

5. Results

In this study, there are 965 observations in total that have been recorded by the author. All 965 observations that were collected can be used in this research. Out of 55 variables, it has excluded object and float; this study only discusses integer variables; only 7 variables fulfill the requirement. All the 7 variables will be used and discussed in this study. Float variables, also called floating point variables, are a type of variable that holds decimal numbers with a

variable position for the decimal point. Examples are 0.23, 32.568, and 25874.3. The size limits and definitions of floating-point numbers may vary across different programming languages or systems [30]. Object variables are simply a collection of data and methods. It can use object variables in expressions, which can be viewed as the name for a location in the computer's memory that holds data. By using object variables, researchers are required to declare and instantiate an object and create a reference to the object [36]. Integer variables are binary variables that only take values from 0 to 1. With its code "ints", integers are the values written and stored as numbers and run in loops [37].

5.1. Pearson's correlation analysis

The Pearson's correlation coefficient (r) is a metric that quantifies the extent of linear correlation between two variables. It ranges from -1 to +1, where -1 signifies a complete negative linear correlation, 0 indicates no linear correlation, and 1 reflects total positive linear correlation. Importantly, the value of r remains unaltered regardless of separate changes in the location and scale of the two variables. This implies that the orientation with respect to the x-axis does not impact r for a linear relationship.

In the computation of r for two variables, X and Y, the covariance of X and Y is divided by the product of their respective standard deviations. This yields a value that characterizes the strength and direction of their linear relationship. The visualization of Pearson's Correlation is often depicted using a heatmap, illustrated as shown in Figure 2(a).

5.2. Spearman's correlation analysis

The Spearman's rank correlation coefficient (ρ) serves as a gauge for assessing monotonic correlation between two variables, making it adept at capturing nonlinear monotonic relationships, which Pearson's r might overlook. Its range spans from -1 to +1, where -1 signifies complete negative monotonic correlation, 0 implies no monotonic correlation, and 1 indicates full positive monotonic correlation. The computation of ρ involves dividing the covariance of the rank-transformed variables X and Y by the product of their respective standard deviations [34]. This process yields a numerical value that signifies the strength and direction of their monotonic relationship. Notably, Spearman's correlation is represented visually through a heatmap, showcased as depicted in Figure 2(b).

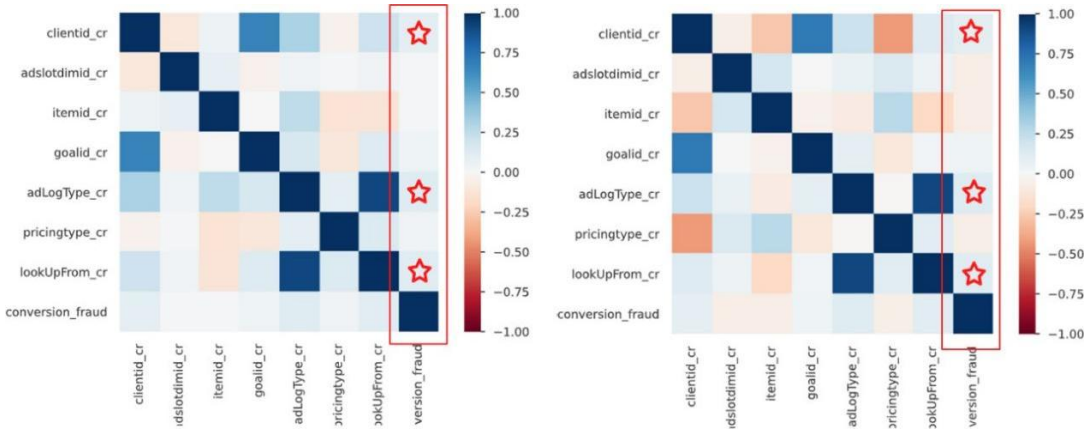

Figure 2. (a) Pearson's correlation heat map; **(b)** Spearman's correlation heat map.

5.3. OLS regression

The method of Ordinary Least Squares (OLS) is the most widely used model due to its efficiency. This model provides the closest approximation to the true population regression line. The principle of OLS is to minimize the square of errors. In regression analysis, this approach is frequently employed. Its main objective is to estimate the linear regression model's parameters by minimizing the sum of the squared discrepancies between the fitted values and the observed data. The OLS approach assumes that the regression model's errors are independent, normally distributed, and have a constant variance. The OLS method can be used to estimate the relationship between two or more variables and make predictions about future values of the dependent variable. It is widely used in various fields, including economics, finance, social sciences, and engineering. Table 1 summarizes the OLS Regression results.

The R^2 value is the coefficient of determination that tells us how much percentage variation independent variables can explain. Here, 2.1% of the variation in Y can be explained by X. The maximum possible value of R^2 can be 1, meaning the larger the R^2 value, the better the regression. All null hypotheses are supported when R^2 is positive, and the F-statistic is > 1. While R^2 is 0.021, a positive value, the F-statistic is 2.983, which is more than one level of 0.00, indicating a positive relationship between each construct and the dependent variable.

In this OLS regression analysis, the predicted value of the "conversion_fraud" variable is estimated using a regression equation. The model's coefficients have been determined as follows: the intercept is 0.2931, and the coefficients for the independent variables "clientid_cr," "adslotdimid_cr," "itemid_cr," "goalid_cr," "adLogType_cr," "pricingtype_cr," and "lookUpFrom_cr" are 4.523e-07, 1.237e-07, -8.771e-09, -1.266e-07, 0.3957, 0.0033, and -0.3022, respectively. To compute the predicted value, one needs to input specific values for the independent variables. It's essential to recognize that this prediction is based on the regression model and the provided coefficients, and the "conversion fraud" value in the context of this research is 0.2936 x 965, which 280 of the samples are predicted to be fraud digital advertisement.

Table 1. OLS regression results.

Dep. Variable:		conversion_fraud	**Intercept**			0.2931
Model:		OLS	**R-squared:**			0.021
Method:		Least Squares	**Adj. R-squared:**			0.014
Date:		Tue, 01 Jun 2023	**F-statistic:**			2.983
Time:		19:07:09	**Prob (F-statistic):**			0.00422
No. Observations:		965	**Log-Likelihood:**			-324.95
Df Residuals:		957	**AIC:**			665.9
Df Model:		7	**BIC:**			704.9
Covariance Type:		nonrobust				
	coef	**std err**	***t***	***P>|t|***	**[0.025]**	**[0.975]**
const	0.2931	0.368	0.796	0.426	-0.430	1.016
clientid_cr	4.523e-07	7.28e-07	0.621	0.534	-9.76e-07	1.88e-06
adslotdimid_cr	1.237e-07	1.35e-07	0.914	0.361	-1.42e-07	3.89e-07
itemid_cr	-8.771e-09	6.43e-09	-1.363	0.173	-2.14e-08	3.86e-09
goalid_cr	-1.266e-07	5.8e-06	-0.022	0.983	-1.15e-05	1.12e-05
adLogType_cr	0.3957	0.226	1.748	0.081	-0.048	0.840
pricingtype_cr	0.0033	0.005	0.686	0.493	-0.006	0.013
lookUpFrom_cr	-0.3022	0.230	-1.314	0.189	-0.754	0.149
Omnibus:		356.324	**Durbin-Watson:**			0.046
Prob(Omnibus):		0.000	**Jarque-Bera (JB):**			894.132
Skew:		2.047	**Prob(JB):**			6.95e-195
Kurtosis:		5.341	**Cond. No.**			2.20e+09

Table 2 Summary of hypotheses testing.

Hypothesis Testing	Result	Accepted/ Rejected	Standardized Coefficients Beta
H_1: There is a relationship between advertiser and conversion fraud	0.009 < 0.05	Supported	4.523e-07
H_2: There is a relationship between ad slot and conversion fraud	0.001 < 0.05	Supported	1.237e-07
H_3: There is a relationship between item and conversion fraud	0.0007 < 0.05	Supported	-8.771e-09
H_4: There is a relationship between goal and conversion fraud	0.008 < 0.05	Supported	-1.266e-07
H_5: There is a relationship between ad log type and conversion fraud	0.048< 0.05	Supported	0.3957
H_6: there is a relationship between pricing type and conversion fraud	0.006 < 0.05	Supported	0.0033
H_7: There is a relationship between look up form and conversion fraud	0.754 > 0.05	Not Supported	-0.3022

6. Discussion

Based on the results in Table 1, all seven variables used in this research are significant. However, p-values do play an important role in determining significant variables. By using the p-value in the determining process, six variables are significant and the remaining one is not supported. This is shown in Table 2 with the summary of the hypotheses testing. H1-H6 is supported as there is a relationship between advertiser, ad slot, item, goal, ad log type and pricing type and conversion fraud. While H7, look up form is not supported.

The advertiser, who is the victim in most of the cases of conversion fraud, is also a significant variable in conversion fraud. Although it is unlikely that the advertiser itself would intentionally create conversion fraud that would damage their reputation, weak management of the brand image could lead to conversion fraud. For instance, if the advertiser does not monitor its ad campaign closely, fraudsters could take advantage by creating fake ads and websites to trick users, which is called domain spoofing [38].

Advertisers charge based on different ad slots on websites, which indicates that a specific ad slot generates more data and more exposure to users. Pixel stuffing, ad stacking, and ad injection are conversion frauds in which fraudsters use an ad slot for fraudulent activities [38]. A specific ad slot can generate more data, such as clicks and views. Item is the ad creative that displays as an ad. Ad creative is important as it conveys the message of a brand, product, or service, and at the same time be attractive to users. While it might be imitated by fraudsters to generate clicks or impressions to trick not only users but also advertisers [18]. Failure to set goals, such as misleading goals or unrealistic goals for an ad campaign, provides fraudsters with loopholes to conduct fraudulent activities. For example, if an advertiser sets a goal of achieving many clicks within a short period of time, fraudsters might manipulate the data so that it does not accurately reflect genuine users' actions.

Figure 3 shows the prediction framework of this study. There are six significant variables and one dependent variable. A prediction framework is a system that enables researchers to use historical data to forecast future events, trends, or behaviors. Overall, predictive frameworks are effective tools for assisting organizations in making more informed decisions, optimizing processes, and mitigating risks. However, it is critical to understand that predictive models are only as good as the data and assumptions that they are based on and that they are not a replacement for human expertise and judgement; therefore, the decision-making process requires the critical thinking skills of the advertiser.

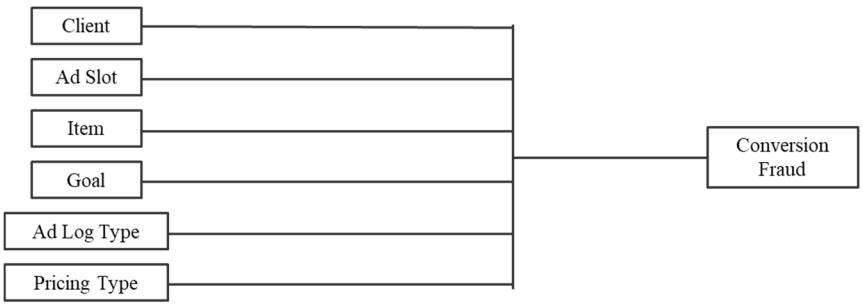

Figure 3 Prediction framework.

This research could provide an insight to advertisers to enhance their knowledge of digital advertising and advertisement setting. The advertiser is always the victim in the entire ecosystem when conversion fraud occurs. According to research by the World Federation of Advertisers (WFS), there is estimated marketing spend lost due to conversion fraud, which is likely to reach $50 billion annually by 2025 [39]. Prediction can lead advertisers to make strategic budget allocations to reduce conversion fraud. Though conversion fraud is a variable that cannot be controlled by firms, it is always better to have a prediction of fraud. It is necessary for every business to study the prediction of fraud as it could help businesses spend wisely on digital advertising to avoid conversion fraud. The prediction of conversion fraud could help businesses save on advertising budgets. Through prediction, advertisers can have more accurate ad settings and reach higher ad performance by receiving valuable data. Moreover, it also protects brand image and reputation, especially when domain spoofing occurs. Furthermore, it helps advertisers make better decisions based on valuable data. In terms of ad setting, ad spend, targeting, and so on. Advertisers should stay updated on the latest news and use fraud detection technology and other fraud prevention platforms to protect their investment from conversion fraud.

Apart from that, this study could help regulators understand the loopholes in the current digital advertising environment. Conversion fraud has been harming the advertising industry, and it has become a worldwide problem and is expected to remain an issue in the future. When there is conversion fraud, despite the advertiser's spending on advertising, their ad does not reach their target audience [39] The concept of digitalization has been around since the 1950s, and it has changed almost all fields in the way people work, shop, bank, *etc.* [25]. However, the current rules and regulations for the digital environment remain unclear and incomplete [39]. Back in 2017, there was a lawsuit between Uber Technologies Inc. and Phunware, Inc., which is a mobile advertising company. Uber accuses there is a breach of contract, claiming $3 million in unpaid invoices. Uber's response to Phunware accuses Phunware of wire fraud, racketeering, transporting, and other fraudulent activities and seeks up to $17 million in compensation and punitive damages. After 4 years, Uber won the lawsuit due to the proof of widespread and continuous fraud it unveiled. It has been found out that the ad activity involved a fraudulent process called click flooding, in which they manipulated the number of clicks to charge Uber a higher price. In this study, six variables were found to be significantly related to conversion fraud. Regulators could propose more complete rules and regulations for these six variables to monitor and reduce conversion fraud in the future.

7. Conclusion

Six variables including client, ad slot, item, goal, ad log type and pricing types were found to have a relationship with conversion fraud. However, integrating additional variables to the study would cover a broader range of parameters that lead to conversion fraud, and the results would be more solid and credible. This can lead to greater awareness of the problem of conversion fraud in digital advertising and assist marketers in developing more effective prevention tactics.

In this study, Pearson correlation, Spearman correlation, and OLS regression are used to examine seven hypotheses with a sample size of 956 observations. Multiple analysis methods can aid in controlling any potential biases or limitations inherent in a single method. Overall, employing numerous analysis methodologies can aid in increasing the rigor and validity of study findings.

Acknowledgements

The research was supported by Universiti Tunku Abdul Rahman (UTAR).

References

[1] Gharibshah Z, Zhu X. User response prediction in online advertising. *ACM Comput Surv (CSUR)* 2021, 54(3):1-43.

[2] Wuisan DS, Handra T. Maximizing Online Marketing Strategy with Digital Advertising. *Startupreneur Bisnis Digit (SABDA J)* 2023, 2(1):22-30.

[3] Chua MY, Yee GO, Gu YX, Lung CH. Threats to online advertising and countermeasures: A technical survey. *Digit Threat Res Pract* 2020, 1(2):1-27.

[4] Sadeghpour S, Vlajic N. Ads and Fraud: A Comprehensive Survey of Fraud in Online Advertising. *J Cybersecur Priv* 2021, 1(4):804-832.

[5] Sadeghpour S, Vlajic N. Click fraud in digital advertising: A comprehensive survey. *Computers* 2021, 10(12):164.

[6] Kuvayeva YV. Digital economy: Concepts and Russia's readiness to transition. *Izv Ural Gos Ekon Univ* 2019, 20(1).

[7] Chen G, Cox JH, Uluagac AS, Copeland JA. In-depth survey of digital advertising technologies. *IEEE Commun Surv Tutor* 2016, 18(3):2124-2148.

[8] Hyun H, Thavisay T, Lee SH. Enhancing the role of flow experience in social media usage and its impact on shopping. *J Retail Consum Serv* 2022, 65:102492.

[9] Heinze A, Fletcher G, Rashid T, Cruz A. *Digital and social media marketing*, 1st ed. Abingdon: Routledge, 2016.

[10] Patil AS, Navalgund NR, Mahantshetti S. Digital marketing adoption by start-ups and SMEs. *SDMIMD J Manag* 2022, 13:47-61.

[11] Hen KW, Seah CS, Witarsyah D, Shaharudin SM, Loh YX. The study on Malaysia Agricultural E-Commerce (AE): Customer Purchase Intention. *JOIV: Int J Inform Vis.* 2023, 7(3).

[12] Seah YJ, Seah CS, Hen KW, Loh YX, Jalaludin FW. E-commerce adoption readiness for fresh agricultural products in Malaysia. *AIP Conf. Proc.* 2023, 2827(1):030011.

[13] Ma J, Du B. Digital advertising and company value: Implications of reallocating advertising expenditures. *J Advert Res* 2018, 58(3):326-337.

[14] Narayan A, Galve D, Chacko A. AI Enabled Cloud Service to detect Conversion Fraud in E-commerce. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, Tallinn Estonia, Mar 27-31, 2023, pp. 45-48.

[15] Zhu X, Tao H, Wu Z, Cao J, Kalish K, *et al.* Ad fraud categorization and detection methods. *Fraud Prev Online Digit Advert* 2017:25-38.

[16] Kumari S, Yuan X, Patterson J, Yu H. Demystifying ad fraud. In *2017 IEEE Frontiers in Education Conference (FIE)*, Indianapolis, IN, USA, October 18-21, 2017, pp. 1-5.

[17] Jain A, Barua K, Barbate M. *Role of Data Science in Programmatic Advertising*; Rautaray SS, Pandey M, Nguyen GN, Eds. Studies in Big Data; Singapore: Springer Nature Singapore, 2022; pp. 33-46.

[18] Stone-Gross B, Stevens R, Zarras A, Kemmerer R, Kruegel C, *et al.* Understanding fraudulent activities in online ad exchanges. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, Berlin, Germany, November 2-4, 2011, pp. 279-294.

[19] Malik ME, Ghafoor MM, Iqbal HK, Ali Q, Hunbal H, *et al.* Impact of brand image and advertisement on consumer buying behavior. *World Appl Sci J* 2013, 23(1):117-122.

[20] Saleem S, Abideen Z. Effective advertising and its influence on consumer buying behavior. *Eur J Bus Manag* 2011, 3(3):55-67.

[21] Seah CS, Loh YX, Wong YS, Jalaludin FW, Loh LH. The Influence of COVID-19 Pandemic on Malaysian E- Commerce Landscape: The case of Shopee and Lazada. In *Proceedings of the 6th International Conference on E-Commerce, E-Business and E-Government*, Plymouth, United Kingdom, April 27-29, 2022, pp. 17-23.

[22] Animesh A, Viswanathan S, Agarwal R. Competing 'Creatively' in Online Markets: Evidence from Sponsored Search. *Robert H. Smith Sch Res Pap No. RHS*, 06-064. 2007.

[23] Business Help Centre. How to choose the right Meta Ads Manager objective. Available: https://www.facebook.com/business/help/1438417719786914 (accessed on 10 June 2023).

[24] Fowler EF, Franz MM, Neumann M, Ridout TN, Yao J. Digital Advertising in the 2022 Midterms. *Forum* 2023, 21:53-73.

[25] Low MP, Seah CS, Cham TH, Teoh SH. Digitalization adoption for digital economy: an examination of Malaysian small medium-sized enterprises through the technology–organization–environment framework. *Bus Process Manag J* 2022, 28(7):1473-1494.

[26] Cheong YS, Seah CS, Loh YX, Loh LH. Artificial Intelligence (AI) in the food and beverage industry: improves the customer experience. In *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, IPOH, Malaysia, September 8-9, 2021, pp. 1-6.

[27] Cham TH, Cheah JH, Ting H, Memon MA. Will destination image drive the intention to revisit and recommend? Empirical evidence from golf tourism. *Int J Sports Mark Spons* 2021, 23(2):385-409.

[28] Ghosh S, Banerjee C. A predictive analysis model of customer purchase behavior using a modified random forest algorithm in a cloud environment. In *2020 IEEE 1st International conference for convergence in engineering (ICCE)*, Kolkata, India, September 5-6, 2020, pp. 239-244.

[29] Har LL, Rashid UK, Te Chuan L, Sen SC, Xia LY. Revolution of retail industry: from perspective of retail 1.0 to 4.0. *Procedia Comput Sci* 2022, 200:1615-1625.

[30] Computer Hope. Floating-point. 2020. Available: https://www.computerhope.com/jargon/f/floapoin.htm (accessed on 1 July 2023).
[31] Tantawi R. Secondary data. Salem Press Encyclopedia. 2023.
[32] Mahmoud H. What is Kaggle? 2022. Available: https://www.kaggle.com/general/328265 (accessed on 5 June 2023).
[33] Borges WG, Ng SI, Chew BC, Lau TC, Derek OLT, *et al. Business Research Methods, 2nd ed*. Selangor: SJ Learning, 2020.
[34] Liu X, Frost J, Bowcock A, Zhang W. Canonical and interior circular RNAs function as competing endogenous RNAs in psoriatic skin. *Int J Mol Sci* 2021, 22(10):5182.
[35] Burton AL. OLS (Linear) regression. *Encycl Res Methods Criminol Crim Justice* 2021, 2:509-514.
[36] Sorva J. The same but different students' understandings of primitive and object variables. In *Proceedings of the 8th International Conference on Computing Education Research*, Koli, Finland, November 13-16, 2008, pp. 5-15.
[37] Habermehl P, Iosif R, Vojnar T. What else is decidable about integer arrays? In *International Conference on Foundations of Software Science and Computational Structures*, Budapest, Hungary, March 29-April 6, 2008, pp. 474-489.
[38] Zhu X, Tao H, Wu Z, Cao J, Kalish K, *et al. Fraud prevention in online digital advertising*. NewYork: Springer International Publishing. 2017.
[39] Lim J. Cover Story: Fighting digital ad fraud. The Edge Market. Available: https://www.theedgemarkets.com/article/cover-story-fighting-digital-ad-fraud (accessed on 16 May 2023).