*International Journal of Computer Applications (0975 - 8887)*
*Volume 186 - No.1, January 2024*

# Online Advertising and Fraud Click in Online Advertisement: A Survey

Ranjeet Vishwakarma
Department of Computer Engineering
Shri G.S. Institute of Technology and Science
Indore, Madhya Pradesh, India

Rajesh Dhakad
Associate Professor
Department of Computer Engineering
Shri G.S. Institute of Technology and Science
Indore, Madhya Pradesh, India

## ABSTRACT

The advertising industry is rapidly growing, as compared to the past advertising medium has evolved significantly. Previously, the advertisement medium was primarily print-based. With the growth of the internet, it has shifted to internet-based online advertisement. Nowadays, each and every company advertises on the Internet, considering its presence in people's daily lives, due to this the advertising industry has become a multi-billion-dollar industry. One widely used revenue model in online advertising is Pay-Per-Click (PPC). However, PPC also brings about challenges such as click fraud, where advertising agencies generate fake clicks, resulting in a rise in advertising costs and reduced Return on Investment (ROI). Click fraud, including activities like click farms, automated bots, and manual clicking. It is a significant issue that can significantly impact on business's financial performance. To address this problem in the past, various approaches have been proposed and implemented. By detecting and preventing click fraud, advertisers can ensure the effectiveness of their advertisements and only pay for legitimate clicks. Fraud click detection can lead to substantial cost savings. This paper presents a survey that aims to provide insights into click fraud detection and the domains actively involved in countering this fraudulent behaviour.

### Keywords

Fraud Click, advertisement, Pay-Per-Click, click farms, automated bots, detection method

## 1. INTRODUCTION

Online advertising is rapidly growing in today's world. The internet has presented us with a new opportunity everywhere, due to the use of rapid increase in internet access in the last few years. At the same time it has become easy to commit fraud and fraudulent clicks in advertising are a serious issue. In the year 2020, over 4 billion individuals engaged with the internet every day [8]. According to "Statista - Digital ad spend worldwide 2026" [34], worldwide spending in 2021 on digital advertising is $522.5 billion and according to the data analysis from previous years, the projected spending by 2026 is $836 billion.

There are various revenue models used to calculate payments for displaying online advertisements. Typically, the payment is determined by three primary methods: cost-per-click (CPC), cost-per-thousand-impressions (CPM), or cost-per-action (CPA) [9]. One of the commonly utilized revenue model is CPC which is also known as Pay-Per-Click (PPC) [21]. According to the PPC model payment is made based on per user click on the advertisement. As the payment mode is user based so fraud clicks can be done to generate revenue which is known as click fraud. Click fraud is a big problem in online advertising that can affect a business's profits. It is a practice where individuals or automated programs or bots generate fake clicks on an online advertisement to increase the number of clicks. This fraudulent click can be done by crowdsourcing, incentivized traffic, click farms, hit inflation attacks, bots, impression fraud, or even by competitors [11]. The main goal of click fraud is to drain the advertiser's advertising budget and decrease the effectiveness of online advertising.

To find click fraud, advertisers use different methods like analyzing the click patterns, traffic analysis, data mining, machine learning algorithms, and honeypot, in which the information contained when the click happened, where it came from, and how the user behaved [11]. From this information, specific data is generated to detect the click as fraudulent or legitimate.

By finding and stopping fraudulent clicks, advertisers can make their online ads more effective and make sure they only pay for real customers. They do this by blocking suspicious clicks that don't match the usual patterns of normal clicks. It saves money and makes advertisers' ads work better overall.

Some important technical terms which are used in the literature to click fraud and online advertisement [11].

**Ad Click** - Clicking or Interacting with an advertisement by clicking on it.

**Ad Fraud** - The act of intentionally misrepresenting or providing false information in online advertising

**Ad Network** - An intermediary platform that serves as a central hub, connecting advertisers and publishers together.

**Advertiser** - An advertiser is someone who promotes a product or service on an advertising platform.

**Ad Injection** - The Unauthorized placement of additional advertisement without consent, disrupting the intended advertising experience.

**Blacklisting** - A compilation of identified and excluded problematic IP addresses and domains within an ad network.

**Bots** - Automated bots also known as robots, mimic human actions and perform tasks on the internet. They engage in activities such as watching ads, viewing videos, and clicking on ads. These actions can be intentional or unintentional, but they often lead to fake engagement or impressions.

**Click Farm** - This form of ad fraud requires the involvement of extensive networks of a group of human employees or workers, typically situated at one or many geographical locations, who are compensated to view, additional click-on advertisements on behalf of someone else.

**Impression** - The primary currency for online advertisements is usually measured and recorded by ad servers in order to bill advertisers or agencies representing them.

**Human Impression** - A legitimate click or ad interaction generated by an actual human user.

**Invalid Impression** (Fraud Click) - An impression that is not genuinely delivered to an actual human user is considered fraudulent activity and can occur due to various factors.

**IP Address** - An internet protocol (IP) address, is a distinct numeric label assigned to an Internet-connected device or network. It serves as a means of identifying and communicating with these devices on the Internet.

**Pay-Per-Click** - Whenever a user interacts with a link, the advertiser commits to paying a fixed fee to the ad network.

## 2. OVERVIEW

This section provides a concise summary of the evolution of advertising, various online advertising formats and revenue models.

### 2.1 Evolution of Advertisement

The advertising industry has experienced significant changes over the years, adapting to advancements in technology, shifts in consumer behaviour and the rise of new media platforms. Here's an overview of the major stages in the evolution of advertising.

*2.1.1 Print Era (Late 1800s).* According to E. Applegate[2], advertising emerged in the form of printed materials including newspapers, pamphlets and posters. These materials were designed to promote products and services using a combination of text, images, and illustrations.

*2.1.2 Broadcast Era (1920s - 1950s).* The arrival of radio and television brought advertising to mass audiences. Businesses started using catchy jingles, memorable slogans and eye-catching visuals to engage consumers. Television advertisement became a popular medium for advertising campaigns[2].

*2.1.3 Mass Media Era (1950s - 1980s).* From the mid-20th century to the late 20th century, the Mass Media Era had a significant impact as different communication platforms influenced people. Advertisers extensively utilized print and broadcast methods to reach wide audiences and create brand identities. This era played a major role in shaping how we received information and entertainment[2, 17].

*2.1.4 Digital Era (1990s - Present).* As per Lee et al., [3] the rise of the internet brought about a significant shift in advertising. This era opened up new opportunities for targeted marketing and interactive advertising. Banner ads, pop-ups and email marketing became common strategies to reach online audiences, the same is considered as the digital era.

*2.1.5 Mobile Era (2000s - Present).* Billore et al.,[5] discussed that due to the widespread adoption of smartphones and mobile applications transformed advertising once again. Mobile advertising includes various formats such as in-app ads, push notifications, mobile video ads and sponsored content optimized for mobile devices. Location-based advertising also became possible, targeting users based on their geographical proximity.

*2.1.6 Social Media Era (2000s - Present).* From the mid-2000s, a new era of advertising came in with the success of social media sites like LinkedIn(2003), Facebook(2004), YouTube(2005), Twitter(2006), Instagram(2010), Snapchat(2011) etc. Brands began utilizing these platforms to establish their online presence, engage with customers, and run targeted advertising campaigns based on user demographics and interests of the users [16, 26].

*2.1.7 Native Advertising Era (2010s - Present).* Native advertising focuses on seamlessly integrating advertisements within the content to provide a non-disruptive user experience. It is more engaging and less intrusive since it follows the form and functionality of the platform on which it appears [6].

*2.1.8 Influencer Marketing Era (2010s - Present).* According to Kim et al., social media influencers gained popularity by social media platforms and brands started collaborating with them to promote their products or services. It involves collaborating with a popular public figure who has a significant following to promote products through their content [15].

*2.1.9 Personalization and Data-Driven Advertising (2010s - Present).* Advances in data analytics and technology have allowed for highly targeted and personalized advertising. Advertisers can now gather and analyze consumer data to deliver personalized messages, offers, and recommendations based on individual preferences and behaviours [17].

*2.1.10 Augmented Reality and Virtual Reality (2010s - Present).* Jayawardena et al., discussed that Augmented Reality (AR) and Virtual Reality (VR) technologies introduce immersive advertising experiences. Customers can interact with brands' products and services in a more immersive and compelling way by creating virtual and interactive experiences for viewers. [29].

*The Future of Advertising.* The advertising industry is a dynamic field, constantly evolving with new technologies and consumer trends. The future of advertising may bring further advancements, with artificial intelligence, voice activation and other innovative forms of promotion.

### 2.2 Online Advertising Formats

Businesses and marketers utilize a variety of Internet advertising formats to market their goals, services, or brands. Shaari et al., [30] present some example of popular online advertising:

*2.2.1 Display-based advertising.* Display-based advertising is pictorial or visible commercials that show on websites, mobile applications or social media platforms. They could be made up of images, text, logos, videos, animations or other graphical

interactive elements. Display ads are typically shown to users based on their browsing behaviour, demographics, or interests.

*2.2.2 Search engine marketing (SEM).* Search engine marketing is embedding advertisements on search engine result pages. These ads are typically text-based and appear alongside organic search results. When a user enters certain keywords into a search engine, the advertiser's advertising is displayed.

*2.2.3 Mobile advertising.* Mobile advertising targets users on their smartphones, tablets or other smart devices. It can include various formats such as in-app ads, mobile banners, interstitial ads, SMS or MMS. Mobile advertising often utilizes location data for Geo-targeting or delivering location-based advertisements.

*2.2.4 Social Media advertising.* Advertising on social media platforms such as Facebook, YouTube, Instagram, Snapchat, LinkedIn or Twitter is known as social media advertising. Based on demographics, interests, age, behaviours, language or connections, advertisers can produce tailored adverts.

*2.2.5 E-mail advertising.* Email marketing involves sending promotional messages or advertisements directly to a recipient's email inbox. This can include newsletters, product announcements, special offers, or personalized recommendations. Email marketing campaigns can be highly targeted and customized based on user preferences or behaviour.

### 2.3 Revenue Models

Revenue from online advertising campaigns is typically displayed and accounted for in one of three primary categories[9]:

*2.3.1 Cost per mille (CPM).* It refers to the amount that an advertiser must spend for every one thousand impressions of their advertisement. It doesn't matter if the user engages with the advertisement or not, it counts as an impression when it is displayed. Ad revenue in CPM is determined by the quantity of provided impressions.

*2.3.2 Cost per click (CPC).* It represents the amount that an advertiser pays for each click on their ad. The amount of clicks receive determines how much advertising revenue is generated in CPC. Regardless of the number of impressions, advertisers are only charged when a user clicks on their advertisement.

*2.3.3 Cost per action (CPA).* It reflects the amount that an advertiser pays for a particular conversion, like a purchase, sign-up or download brought on by the advertisement. The completion of a specific action by the user is tied to ad revenue in CPA. Advertisers are charged only when a desired action is completed.

## 3. TYPES OF ONLINE ADVERTISEMENT FRAUD

The growth of online advertising has led to a rise in ad fraud incidents. Ad fraud involves falsely recording views, clicks, actions or data events to dishonestly obtain revenue or mislead users. Revenue-generating ad fraud is common. As shown in Fig. 1 advertising fraud primarily comes in three categories. One is placement fraud, the other is traffic fraud and the third is fraud related to actions. Each type involves reporting fraudulent visitors as genuine, whether robotic, human or a combination of the two.

### 3.1 Action Fraud

Action Fraud misleads users into performing crucial business activities such as filling out forms, making transactions or utilizing user behaviour for targeted ads. It directly affects the cost-per-action model, ad pricing and campaign planning within the advertising ecosystem. Advertisers rely on authentic user actions to make informed decisions, but action fraud disrupts this process and compromises the integrity of the advertising ecosystem. It commonly includes Affiliate fraud, Re-targeting fraud and Conversion fraud[42].

*3.1.1 Affiliate Fraud.* It involves falsely claiming commission in affiliate marketing is paid to affiliates for bringing visitors to a website. Fraudulent affiliates use tactics including cookie stuffing, forced clicks, fake referrals, ad stacking and sub-ID fraud to target potential customers who are ready to make a purchase. To maintain trust and mitigate its effects, it is important to monitor and uncover such fraud. Affiliate fraud commonly occurs through three primary methods[31]:

*3.1.1.1 Malware and Adware.* Without affiliate assistance or referrals, visitors to a branding company's website will not be able to claim the commission. The user may be redirected to the affiliate's marketing link and the referral might be wrongly credited to the affiliate if the visitors' system has affiliate-powered spyware. Affiliate fraud occurs when the user makes a transaction after that.

*3.1.1.2 URL Hijacking.* URL hijacking or typosquatting happens when users make typing errors in website addresses (e.g. typing "www.advertisement.com" instead of "www.advertisements.com"), leading to redirection and commission claims. In October 2021, the Facebook business changed its name to Meta since then, there are more than 5000+ similar newly registered domains discovered on the internet [7].

*3.1.1.3 Cookie Stuffing.* Affiliates attract audiences to a website and then implant cookies on their computers. The affiliate receives payment if the audience makes a purchase from the advertiser's website during the cookie's lifespan. They accomplish this by creating web pages that attract potential customers with false promises of coupons or discounts.

*3.1.2 Re-targeting Fraud.* Re-targeting or re-marketing fraud is an efficient kind of online advertising that identifies potential buyers based on their past online activities, such as their buying history or web browsing behaviour. Advertisers use cookies or pixels to track user activities and serve relevant ads. In platforms like Google AdWords and Bing, this is called re-marketing or re-messaging [19]. In re-targeting fraud, fraudsters aim to mimic genuine customer behaviours using computer-generated agents, deceiving advertisers that bots are worthwhile potential customers and driving up the price of ad auctions or bot-generated impressions [25].

*3.1.3 Conversion Fraud.* A conversion in the context of an Ad network refers to a significant user action, such as purchasing something or completing the form. Conversion fraud involves manipulating or falsely generating conversions for financial gain. Clicks and conversions are tracked separately and fraudulent activities often target lead-generation conversions that require a minimal financial commitment. Typically, the user's cookie information is compared to identify users whose click result in conversions. The method used for the conversion of fraud is split into two categories:

*3.1.3.1 Lead Bots.* A lead bot is a piece of software created to automatically fill out lead forms with either randomly generated or

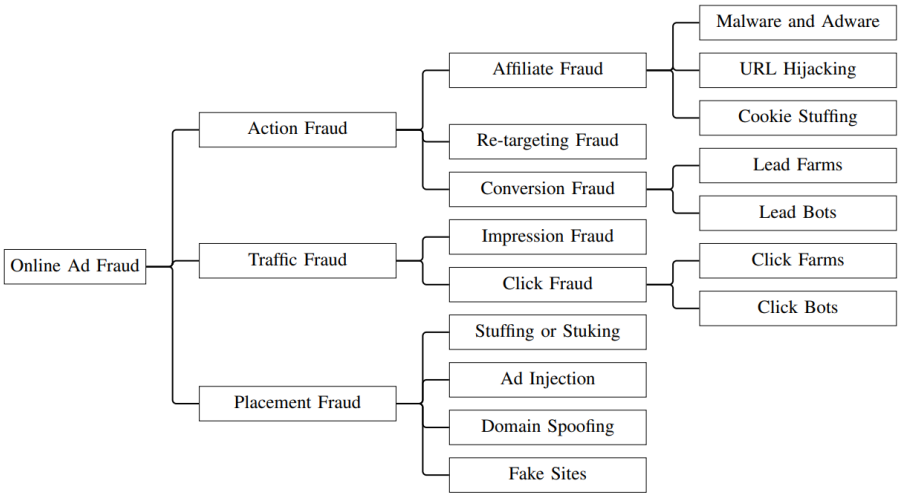


Fig. 1. Types of Online Advertisement Fraud

partially accurate data. These bots can also perform simple actions like clicking download links.

*3.1.3.2 Lead Farms.* Since bots cannot process all conversions and advertisers prefer sites with higher conversion rates, fraudsters turn to hire individuals from underdeveloped countries with less expensive labour to generate conversions. This leads to conversion fraud through the use of lead farms, where real human workers are used to complete lead forms or perform other necessary tasks to convert clicks into conversions.

## 3.2 Traffic Fraud

Boosting revenue for publishers often involves increasing website traffic, making it a target for traffic fraud. This type of fraud focuses on artificially inflating the number of impressions produced from specific websites or placements. In Addition, only when consumers click on displayed advertising do cost-per-click campaigns produce income.

*3.2.1 Impression Fraud.* Impression fraud is a form of fraud that attempts to boost website traffic and produce more impressions. This kind of fraud greatly affects CPM campaigns, as inflated impressions offer advertisers barely any benefits. It also affects CPC and CPA campaigns by lowering the click-through rate (CTR). Impression fraud can be carried out through various methods, including employing people to manually view advertisements using bots to generate impressions or sending visitors from lapsed domains to external pages [32]. Some hybrid approaches combine human actions with automated bot functions to increase website traffic. Mitigation strategies to address impression fraud include filtering zero-sized viewports, restricting traffic from particular networks and blacklisting publishers involved in fraudulent activities

*3.2.2 Click Fraud.* A click on an advertisement indicates potential interest from a viewer and is utilized to measure effectiveness, such as click-through rate (CTR). Click fraud is a widespread form of fraud in the advertising ecosystem, especially in cost-per-click (CPC) campaigns. Fraudsters employ manual or automated methods to generate fake clicks on advertisements, resulting in advertising losses for the brand. These fraudulent clicks do not lead to meaningful business actions, diminishing their value. Click fraud is commonly carried out through two primary types [28]:

*3.2.2.1 Click Farms.* A click farm is composed of numerous hired individuals who manually click on advertisements. While human viewers genuinely perform these clicks, they have little to no intention of becoming actual customers, making their clicks malicious or fraudulent in nature.

*3.2.2.2 Click Bots.* A click bot is a computer program or system, that automatically clicks on links or part of a botnet, designed to simulate human behaviour by repeatedly accessing URLs linked to advertisements and generating mouse click events. In click fraud botnets like ZeroAccess, infected hosts can be managed by a master bot to retrieve online ads and click on them without the knowledge of the host user.

## 3.3 Placement Fraud

A typical ad placement consists of an iframe with text, images or videos as part of the ad content. These placements can be positioned in various areas on a webpage, including sides, top, bottom or integrated with other content. Floating or fixed positions are also possible. Google AdSense provides [12] the following recommendations for ad placement:

1. Keep the user's viewpoint in mind and make the site simple to use.
2. What action do they take while visiting a specific page?
3. Place advertisements adjacent to user-interested content.
4. Check that advertisements are easily distinguished and don't mislead users.
5. Avoid overloading the page with an excessive amount of advertisements.

Placement fraud is the practice of manipulating publishers' websites or changing the pages that users see. These dishonest practices use a variety of tricks, such as keyword stuffing and falsely representing where ads will appear. Malvertising is one of the more malicious practices that involves installing advertising

viruses to divert people to dangerous websites, inflating impression counts.

*3.3.1 Stuffing or Stuking.* Stuffing is a technique for showing information that is impossible to view by the human eye. Ad keywords are stuffed in HTML tags that are difficult to see in the background or identical to the background colour, rendering them invisible or when the placement has an appropriate dimension but the visibility is set to "none", making it impossible to observe. Although hidden keywords can't be visible to the human eye, ad network agents can see them when they crawl web page content to find the relevant pages associated with particular ads. In reality, these strategies are frequently used in search engine optimization or search engine cloaking to raise.

*3.3.2 Ad Injection.* Ad injection refers to a manipulative technique where unauthorized advertisements are injected into web pages without the consent of website owners or advertisers. This practice typically involves malicious software or browser extensions that modify the content of a webpage to display additional ads or replace existing ones. Ad injection can disrupt the user experience, mislead visitors, and generate illegitimate revenue for the perpetrators. It is often used as a form of advertising fraud, as the injected ads may come from unauthorized sources and generate revenue for the fraudsters instead of legitimate publishers or advertisers. Ad networks and security measures actively counter ad injection as it is considered a malicious practice.

*3.3.3 Domain Spoofing.* Web spoofing is a fraudulent practice where fraudsters create websites that mimic legitimate ones to engage in malicious activities, such as identity theft or stealing login information. In the context of ad networks, advertisers maintain whitelists of reputable publishers and blacklists of fraudulent ones. Advertisers prefer to display their ads on high-quality sites and avoid blacklisted sites. To avoid detection, fraudsters engage in domain spoofing, manipulating their domains to appear as if their traffic originates from trusted whitelisted publishers. This enables them to bypass blacklists and trick advertisers into displaying their ads on fraudulent sites.

*3.3.4 Fake Sites.* Fake site fraud occurs by creating sites with legitimate domain names dedicated to displaying ads, fraudsters can generate substantial revenue through participation in ad networks. These sites often contain stolen or irrelevant content and are filled with numerous embedded advertisements. Additionally, fraudsters trick visitors by duplicating content from reputable websites or registering domain names that are similar in nature. These fake sites may redirect users to fraudulent campaigns or trick them into downloading malware.

## 4. TECHNIQUES FOR CLICK FRAUD DETECTION

The present study explores various approaches to identify and mitigate click fraud. These methods can be broadly classified into traditional and recent techniques [11]:

### 4.1 Traditional techniques

*4.1.1 Rule Based Detection.* Using specified criteria and thresholds rule-based detection looks for suspicious click activities by analyzing parameters like click frequency, click time, IP address anomalies and conversion rates.

*4.1.2 IP Address Analysis.* IP address analysis is performed to detect anomalies in click behaviour, such as multiple clicks coming from identical IP addresses or IP ranges that are infamous for fraudulent activities.

*4.1.3 Traffic Source Analysis.* Traffic analysis is a technique used to gather and analyze transaction data in order to derive insights from communication patterns. In the context of advertising, transaction data within an ad network is collected and analyzed to gain a deeper understanding of traffic behaviour. As more data are observed, captured and processed, the capacity to infer traffic patterns and trends increases.

### 4.2 Recent techniques

*4.2.1 Machine Learning Based Approaches.* It is an advanced system capable of analyzing data patterns, making predictions and adapting to new situations even without explicit programming. It improves its performance over time through experience, making it highly effective for fraud detection tasks. By learning from past data, machine learning algorithms can identify fraudulent patterns and anomalies for enhancing their ability to detect and prevent fraudulent activities.

*4.2.2 Anomaly Detection.* Anomaly detection is a technique used to identify patterns or instances that deviate significantly from the norm or expected behaviour. It involves analyzing data to detect outliers or unusual events indicating fraudulent or suspicious activities. By identifying anomalies, it helps in flagging potential fraud and enabling proactive measures to mitigate risks.

*4.2.3 Bot Detection Technique.* Bot detection techniques involve employing various methods to identify and block automated click bots. The method includes CAPTCHA challenges, which verify the user's humanity, browser fingerprinting to analyze unique browser characteristics and JavaScript challenges to assess user interactions. These techniques help in distinguishing between human users and bots, enabling effective bot detection and prevention.

*4.2.4 Collaborative Filtering.* Collaborative filtering involves sharing click fraud data with industry partners, ad networks and fraud detection platforms to identify patterns. It utilizes collective intelligence to enhance click fraud detection. Collaborating and sharing knowledge improves accuracy and effectiveness in preventing click fraud.

## 5. LITERATURE REVIEW

Advertisers face the challenge of online click fraud in different ways, as the prevalence of such fraudulent activities has risen due to increased internet usage. Researchers have proposed multiple approaches to address this problem. The following literature provides a comprehensive overview of current computational methodologies developed to address click fraud challenges.

This literature outlines research papers' titles along with their publication years, as well as methodologies, datasets used, achieved accuracy, limitations, and findings.

Table 1 comprises a comprehensive literature review encompassing various papers, detailing all the pertinent fields. Table 2, the technique mentioned is not required with any dataset type. It collects data at run time and uses it for analysis. In Table 3 displays a literature review that omits accuracy due to the absence of any references to it in the papers examined. Table 4 outlines commonly employed terminologies found within literature reviews.

Table 1. Literature Review

| Title | Year | Techniques | Dataset | Limitation | Result | Accuracy |
|---|---|---|---|---|---|---|
| "Deep Learning-based Model to Fight Against Ad Click Fraud" [36] | 2019 | Auto Encoder, Semi-supervised GAN | Kaggle: TalkingData AdTracking Fraud Detection Challenge | Anomalies require a prior classification to detect, and bots can increase error rates. | Supervised GANs achieve high accuracy levels | 89.7% |
| "FCFraud: Fighting Click-Fraud from the User Side" [13] | 2016 | HTTP GET Function, JavaScript, Mouse Event Test, Blacklist | Self-Generated, Records: 165 426 | Detecting complex JavaScript used in click fraud schemes presents a significant challenge | Enhances server-based detection techniques by offering additional capabilities | 85.6% |
| "Light GBM Machine Learning Algorithm to Online Click Fraud Detection" [22] | 2019 | Feature parallelism, Data parallelism, Voting parallelism | Kaggle, Records: 203 694 359 | Limited resources available for training | Gradually enhances the detection performance | 98% |
| "A Multi-time-scale Time Series Analysis for Click Fraud Forecasting using Binary Labeled Imbalanced Dataset" [37] | 2019 | Learning-based Probabilistic Model and Estimator, Auto Regressive model, Data Pre-processing and smoothing | Kaggle: TalkingData AdTracking Fraud Detection Challenge | Identifying Botnets that simulate genuine user behaviour to click on ads poses a challenge in detection | The probability model has proven more effective than the learning-based probabilistic estimator model, exhibiting superior performance and accuracy | 96% |
| "Data Analysis Algorithm for Click Fraud Recognition" [10] | 2018 | KNN, Clustering, Classification, Data Collection | Self-Generated, Records: 30 000 | Obtaining algorithms for user classification of organic versus non-organic interactions can be challenging | The application of KNN classifier for the clustering algorithm shows excellent performance. | 97.11% |
| "A New Approach for Advertise CTR Prediction Based on Deep Neural Network" [39] | 2018 | Sparse data prediction methods, Logless | Frappe, SIGKDD Cup2012 track2 | Efficiency was reduced due to the additional overhead introduced during the pre-training phase. | The model examines feature relationships, leading to an enhanced Click-Through Rate | 79.81% |
| "Crowdsourcing for click fraud detection" [23] | 2019 | Android Application | Self-Generated, Records: 500 000 | Incorporates an additional move to combine the library, enhancing its functionality | Click Fraud Crowd-sourcing yields favourable outcomes | 93% |
| "Prediction of Click Frauds in Mobile Advertising" [35] | 2015 | SMOTE, Wrapper feature selection | BuzzCity | Validated for mobile advertisement only | Validation use for detect true users and identify instances of click fraud | 64.07% |
| "Real-Time Ad Click Fraud Detection" [33] | 2020 | Multilevel Perceptron, Heuristics-Sessions, NB, GBoost, KNN | Kaggle: TalkingData AdTracking Fraud Detection Challenge | In real-time scenarios, the classification utility cost is very limited, but availability of datasets for detecting click fraud is sparse | Neural Network based machine learning methods demonstrate the highest precision when evaluated based on precision metrics | MLP=95% NB=90% GB=71% KNN=69% |
| "Click Stream Data Analysis for Online Fraud Detection in E-Commerce" [4] | 2016 | Clustering, SVC, DT, NB, MARS | Records: Train- 3 173 834, Validate- 2 689 005, Test- 2 598 815 | Limited access to user click information, and containing more attributes result in lower accuracy | Used for pre-processing purposes when online processing is necessary | 59.38% |

Table 2. Literature Review

| Title | Year | Techniques | Limitation | Result |
|---|---|---|---|---|
| "Identification of Click Fraud and Review of Existing Detection Algorithms" [14] | 2019 | Behavioural Analysis on OS, Browser, IP | Analyse only behavioural patterns | The algorithm tested to evaluate performance and effectiveness and compare with existing methods is conducted to enhance its capabilities |
| "Click Fraud Detection and Prevention System for Ad Network" [1] | 2019 | Offline analysis and Online analysis | Undetected: Low-frequency attacks, Lack of public datasets | Click fraud detection using a combination of online and offline rules yields effective results |
| "Behavioural Verification: Preventing Report Fraud in Decentralized Advert Distribution Systems" [20] | 2017 | Traffic Analysis, Cost-Per-Click, Cost-Per-Impression, Cost-Per-Action | User privacy decreases and digital signatures are not completely detected | Filtering ad reports based on honest and dishonest users |
| "Method for performing real-time click fraud detection, prevention and reporting for online advertising" [18] | 2016 | Client and Server Side Tracking code, HTTP GET Function, Mouse Event Test, Browser Functionality Test | It is susceptible to additional risks, including the presence of parasite | Assists in distinguishing between valid and invalid clicks to prevent fraudulent billing |

Table 3. Literature Review

| Title | Year | Techniques | Dataset | Limitation | Result |
|---|---|---|---|---|---|
| "A Click Fraud Detection Scheme based on Cost-sensitive BPNN and ABC in Mobile Advertising" [41] | 2018 | ABC, BPNN, SMOTE | BuzzCity Mobile Advertisement Dataset | A cost ratio of 7 does not lead to optimal outcomes | Click Fraud detection using BPNN and ABC algorithm has been found to yield good results in this particular approach |
| "Advertisement Click-Through Rate Prediction Based on the Weighted-ELM and Adaboost Algorithm" [40] | 2017 | WELM Adaboost, Data Preprocessing, AUC, LR, SVM | Private | An imbalance in the distribution of advertising data significantly affects the prediction accuracy, leading to lower accuracy | WELM-Adaboost algorithm outperforms ELM and WELM methods, showing enhanced performance and effectiveness |
| "Clicktok: click fraud detection using traffic analysis" [24] | 2019 | Traffic matrix construction and partitioning, Pooling | Self-Generated, Records: 217 334 190 | can be rendered ineffective by reducing the network loads | Efficient search methods detect repetitive patterns in clickstreams to identify organic click fraud attacks |
| "User click fraud detection method based on Top-Rank-k frequent pattern mining" [27] | 2019 | Frequency of clicks, Time spent on Ad | Not Mentioned | Not effectively optimizing the Top-Rank-k in the frequent pattern mining algorithm may lead to insufficient handling of dynamic clicks. | The employed method is highly efficient in terms of accuracy, as it utilizes graph-based pattern analysis to verify and identify patterns |
| "Detection of Advertisement Click Fraud Using Machine Learning" [38] | 2020 | XGBoost, EDA, Data pre-processing, Data prediction | Not Mentioned | The cost associated with prediction utility for classification is minimal | Detecting and mitigating malware that exploits click fraud to generate revenue |

Table 4. Terminology used in Literature Review

| | |
|---|---|
| **ABC** | Artificial Bee Colony |
| **AUC** | Area Under Curve |
| **BPNN** | Back Propagation Neural Network |
| **DT** | Decision Tree |
| **EDA** | Exploratory data analysis |
| **GAN** | Generative Adversarial Networks |
| **KNN** | K-Nearest Neighbour |
| **LR** | Logistic Regression |
| **MARS** | Multivariate Adaptive Regression Splines |
| **NB** | Naive Bayes |
| **SMOTE** | Synthetic Minority Over-Sampling Technique |
| **WELM** | Weighted Extreme Learning Machine |
| **XGBoost** | Extreme Gradient Boosting |
| **SVC** | Support Vector Classifier |

## 6. CONCLUSION

The paper provides a brief overview of some technical terms, revenue models, and types of online advertisements, also the evolution of advertising and types of click fraud. Now come to some common techniques for click fraud detection and finally a brief overview of the literature survey with all the necessary details. Various techniques mentioned have their advantages and disadvantages. NB, LR, Traffic Analysis, DT, JavaScript, and SVM are some commonly used approaches that can be summarized and presented in tabular form. JavaScript-enabled systems show 90% accuracy in detecting click fraud. LR and offline rules based on the user interface system are effective in detecting fraud. Regression models perform well with more attributes, while offline rules minimize redundant processing. User interface systems monitor activity and validate clicks based on time spent before clicking.

## 7. REFERENCES

[1] Paulo S. Almeida and João J. C. Gondim. Click fraud detection and prevention system for ad networks. *Journal of Information Security and Cryptography*, 5:27, January 2019.

[2] Edd Applegate. Advertising in the US: past, present, future. *Journalism Studies*, 1(2):285–302, January 2000.

[3] Stuart Barnes. Wireless digital advertising: Nature and implications. *International Journal of Advertising*, 21:399–420, January 2004.

[4] Ladislav Beránek, Václav Nỳdl, and Radim Remeš. Click stream data analysis for online fraud detection in e-commerce. In *INPROFORUM 2016*, 2017.

[5] Aditya Billore and Ashish Sadh. Mobile advertising: A review of the literature. *The Marketing Review*, 15(2):161–183, August 2015.

[6] Colin Campbell and Lawrence J. Marks. Good native advertising isn't a secret. *Business Horizons*, 58(6):599–606, November 2015.

[7] CircleID. A Look into New Cybersquatting and Phishing Domains Targeting Facebook, Instagram, and WhatsApp.

[8] Anshuman Dash and Satyajit Pal. Auto-detection of click-frauds using machine learning. *International Journal of Engineering Science and Computing*, 10:27227–27235, 2020.

[9] Neil Daswani, Chris Mysen, Vinay Rao, Stephen Weis, and Kouros Gharachorloo. Online advertising fraud. *Crimeware: understanding new attacks and defenses*, 40:1–28, 2008.

[10] Marcin Gabryel. Data Analysis Algorithm for Click Fraud Recognition. In *Information and Software Technologies*, volume 920, pages 437–446. Springer International Publishing, Cham, 2018.

[11] Nayanaba Gohil and Arvind D Meniya. A survey on online advertising and click fraud detection. In *2nd national conference on research trends in information and communication technology*, 2020.

[12] Google Inc. Best practices for ad placement - Google AdSense Help.

[13] Md. Shahrear Iqbal, Md. Zulkernine, Fehmi Jaafar, and Yuan Gu. FCFraud: Fighting Click-Fraud from the User Side. In *17th International Symposium on High Assurance Systems Engineering*, pages 157–164, Orlando, FL, USA, January 2016.

[14] Shubhangi Jain, Falguni Jindal, Anmolika Goyal, and Savy Mudgal. Identification of Click Fraud and Review of Existing Detection Algorithms. In *ICSSIT*, pages 894–899, Tirunelveli, India, November 2019. IEEE.

[15] Do Yuon Kim and Hye-Young Kim. Influencer advertising on social media: The multiple inference model on influencer-product congruence and sponsorship disclosure. *Journal of Business Research*, 130:405–415, June 2021.

[16] Johannes Knoll. Advertising in social media: a review of empirical evidence. *International Journal of Advertising*, 35(2):266–300, March 2016.

[17] Heejun Lee and Chang-Hoan Cho. Digital advertising: present and future prospects. *International Journal of Advertising*, 39(3):332–341, April 2020.

[18] John Linden and Tobias Teeter. Method for performing real-time click fraud detection, prevention and reporting for online advertising, jun 2016. US Patent 9,367,857.

[19] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. AdReveal: improving transparency into online targeted advertising. In *Proceedings of the 12th ACM Workshop on Hot Topics in Networks*, pages 1–7, College Park Maryland, November 2013.

[20] Stylianos Mamais and George Theodorakopoulos. Behavioural Verification: Preventing Report Fraud in Decentralized Advert Distribution Systems. *Future Internet*, 9(4):88, November 2017.

[21] Bhargavi Mikkili and Suhasini Sodagudi. Advertisement Click Fraud Detection Using Machine Learning Algorithms. In Vikrant Bhateja, Suresh Chandra Satapathy, Carlos M. Travieso-Gonzalez, and T. Adilakshmi, editors, *Smart Intelligent Computing and Applications*, volume 282, pages 353–362. Springer Nature Singapore, 2022.

[22] Elena-Adriana Minastireanu and Gabriela Mesnita. Light GBM Machine Learning Algorithm to Online Click Fraud Detection. *Journal of Information Assurance & Cybersecurity*, pages 1–12, April 2019.

[23] Riwa Mouawi, Imad H. Elhajj, Ali Chehab, and Ayman Kayssi. Crowdsourcing for click fraud detection. *EURASIP Journal on Information Security*, 2019(1):11, December 2019.

[24] Shishir Nagaraja and Ryan Shah. Clicktok: click fraud detection using traffic analysis. In *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, pages 105–116, Miami, Florida, May 2019. ACM.

[25] Peter Nowak. Deceptibots: when machines go bad. *New Scientist*, 214(2870):45–47, June 2012.

[26] Esteban Ortiz-Ospina. The rise of social media. *Our World in Data*, 2019. https://ourworldindata.org/rise-of-social-media.

[27] Lijiao Pan, Shibiao Mu, and Yingyan Wang. User click fraud detection method based on Top-Rank- $k$ frequent pattern mining. *International Journal of Modern Physics B*, 33(15):1950150, June 2019.

[28] Paul Pearce, Vacha Dave, Chris Grier, Kirill Levchenko, Saikat Guha, Damon McCoy, Vern Paxson, Stefan Savage, and Geoffrey M. Voelker. Characterizing Large-Scale Click Fraud in ZeroAccess. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 141–152, Scottsdale Arizona USA, November 2014.

[29] Nirma Sadamali Jayawardena, Park Thaichon, Sara Quach, Ali Razzaq, and Abhishek Behl. The persuasion effects of virtual reality (VR) and augmented reality (AR) video advertisements: A conceptual review. *Journal of Business Research*, 160:113–739, May 2023.

[30] Hala Shaari and Nuredin Ahmed. An extensive study on online and mobile ad fraud. *The Third Conference for Engineering Sciences and Technology*, 2020.

[31] Peter Snyder and Chris Kanich. No please, after you: Detecting fraud in affiliate marketing networks. In *WEIS*, 2015.

[32] Kevin Springborn and Paul Barford. Impression fraud in on-line advertising via pay-per-view networks. In *USENIX Security Symposium*, pages 211–226, 2013.

[33] Apoorva Srivastava. *REAL-TIME AD CLICK FRAUD DETECTION*. Master of Science, San Jose State University, CA, USA, May 2020.

[34] Statista. Digital ad spend worldwide 2026.

[35] Mayank Taneja, Kavyanshi Garg, Archana Purwar, and Samarth Sharma. Prediction of click frauds in mobile advertising. In *2015 8th International Conference on Contemporary Computing*, pages 162–166, Noida, India, August 2015. IEEE.

[36] G. S. Thejas, Kianoosh G. Boroojeni, Kshitij Chandna, Isha Bhatia, S. S. Iyengar, and N. R. Sunitha. Deep Learning-based Model to Fight Against Ad Click Fraud. In *Proceedings of the 2019 ACM Southeast Conference*, pages 176–181, Kennesaw GA USA, April 2019.

[37] G. S. Thejas, Jayesh Soni, Kianoosh G. Boroojeni, S.S. Iyengar, Kanishk Srivastava, Prajwal Badrinath, N.R. Sunitha, Nagarajan Prabakar, and Himanshu Upadhyay. A Multi-time-scale Time Series Analysis for Click Fraud Forecasting using Binary Labeled Imbalanced Dataset. In *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution*, pages 1–8, Bengaluru, India, December 2019. IEEE.

[38] B Viruthika, Suman Sangeeta Das, E Manish Kumar, D Prabhu, and Srm Ist Chennai. Detection of Advertisement Click Fraud Using Machine Learning. *International Journal of Advanced Science and Technology*, 2020.

[39] Qianqian Wang, Fang'ai Liu, Shuning Xing, and Xiaohui Zhao. A New Approach for Advertising CTR Prediction Based on Deep Neural Network via Attention Mechanism. *Computational and Mathematical Methods in Medicine*, 2018:1–11, September 2018.

[40] Sen Zhang, Qiang Fu, and Wendong Xiao. Advertisement Click-Through Rate Prediction Based on the Weighted-ELM and Adaboost Algorithm. *Scientific Programming*, 2017:1–8, 2017.

[41] Xin Zhang, Xuejun Liu, and Han Guo. A Click Fraud Detection Scheme based on Cost sensitive BPNN and ABC in Mobile Advertising. In *2018 IEEE 4th International Conference on Computer and Communications*, pages 1360–1365, Chengdu, China, December 2018.

[42] Xingquan Zhu, Haicheng Tao, Zhiang Wu, Jie Cao, Kristopher Kalish, and Jeremy Kayne. Ad Fraud Categorization and Detection Methods. In *Fraud Prevention in Online Digital Advertising*, pages 25–38. Springer International Publishing, 2017.