A FUZZY RULE-BASED FRAMEWORK FOR DETECTING FAKE ADVERTISEMENTS AND MIS-MARKETING IN SOCIAL MEDIA PLATFORMS

PURUSOTHAMAN J.

International Baccalaureate Diploma Programme:
Creativity, Activity, Service Coordinator and HOD Mathematics, Udon Thani international school, Udon Thani, Thailand

Abstract

Recently, the development of social media has ignited in a proliferation of fake ads and mis-marketing where it is the popularity that rules, not gravity- defying pricing; misleading refund claims; doing or attempting to off-sell on platform transactions, obtuse payment flows. It is difficult to detect such deceptive advertising material because fraudulent behaviour is by nature uncertain and partial, which does not work well with the constraints brought by deterministic binary classification. An interpretable fuzzy rule-base model for evaluating advertising risk in social media is presented in this paper. Five characteristic features, including price deviation, communication preference, response behaviour, refund claim type, and currency clarity are represented with the help of Gaussian membership functions and combined in a Mamdani-type fuzzy inference system to classify advertisements into five risk levels: Certified Authentic, Real, Counterfeit, Fake and Highly Fake. Simulation with real data-based scam patterns and typical situations shows that the proposed method can serve to detect deceptive and borderline mis-marketing ads effectively, with interpretable risk scores and confidence levels for transparent and ethical decision support applications.

Keywords: *Fuzzy rule-based System, Fake Advertisement Detection, Mis-Marketing, Social Media Fraud, Gaussian Membership Functions, Uncertainty Modeling, Risk Assessment*

Introduction

Social Media Advertising Growth and New Threats

The rising popularity of social media hasdramatically changed advertising on the Internet, as low-cost, high-volume campaigns promote products and services to online consumers. Companies are starting to depend more on such platforms to target audience’s withouthaving traditional marketing channels. But such accessibility has also encouragedthe widespread tendency to mis-advertise and mis-market. It has been observed that fake advertisementswith unreasonably low prices, obscure refunding clauses and shadowy payment methods have proliferated, which may cause potential hazards to end-users as well as damage trustworthiness of digital worlds [1], [2].

Attributes of Fraudulent Advertisements and Mis-Marketing

Even the fake social media ads have recognisable but subliminalpatterns. Products of high value listed for very low prices, alongside scarce information about sellersand no real online presence. Add to this the fact that these ads only offer you some direct messaging contacts instead of official sites or secure paymentgateways and it also seems to encourage the users to go offline. Follow-on interactions could focus on spreads of refund lottery promises, misleading currency names or unauthorized data usage. These techniques are in the grey area between aggressive marketing and willful fraud, and they are hard to detect [3].

Limitations of Current Detection Methods

The majority of the current detection methods of the online fraud and fake advertisement are based on binary classification, keyword filtering, or supervisedmachine learning models. Although they are successful forthe case of fraud that is explicit, their performance deteriorates when dealing with partial deception and linguistic ambiguity. Notions like “very lowprice”, “suspicious communication behaviour” or “unclear refund policy” are by no means capable of being represented in all detail using crisp logic. As such, rigid detection systems tend to either generate false positives or miss advanced mis-marketing tactics that are holiday invariants and/or may fall into gray areasof interpretations [4].

Applicability of Fuzzy Logic in Deception Analysis

Fuzzy logic offers an effective mathematical tool for the treatment of vagueness, uncertainty and imprecision implicit in deceptive advertising problems. Contrary to classical logic, fuzzy systems admit variables with membership degrees, which allow us to model linguistic concepts like “very suspicious”, “fairly insecure”, or “rather authentic”. This feature makes fuzzy logic especially well-suited for the examination of social media ads, since disinformation on such channels is seldom black or white butinstead stems from a set of weak but jointly suspicious cues [1], [5]. In addition, fuzzy logic-based systems are transparent and interpretable in nature, favoring transparency/explainability of automated decision-making.

Literature Review

Online Fraud and Fake Advertisement Detection Studies

Online scam and dishonest ads are two important issuesto be studied because of their growing economic cost and social hazard. Previous works tried to discover some explicit fraud patterns of e-commerce transaction from the growing data through rule-based systems (Nie et al., 2015) and statistical methods. These heuristics mainly based on detecting transaction anomalies, abnormal pricing and user complaints (about sellers) toidentify the fraudulent behavior [7]. Though suitable for clear fraud scenarios, such solutions had the shortcomings that they were incapable of handling fraudulent marketing techniques which did not involve instant money exchanging or a direct law broken.

As social media advertising rose in popularity, scholars studied the trends of platform-specific fraud related to fake product promotion, impersonation account, and deceptive advertisement. Some other studies indicatedthat social media fraud is distinct from e-commerce fraud in many aspects such as informal operations, no standardized seller checking procedures and with some degree of user interaction outside the platform [8]. Those features make automatic detection far from trivial, they also ask for more versatile analysis models.

Learning-based Methods for Deceptive Content Classification

Fake ads and online scams detection have been studied recently using more advanced machine learning and deep learning methods. Supervised learning algorithms including support vector machines, random forests and neural networks have been used to classify ads using text features, image context and user engagement signal [9]. Deep learning architectures, including convolutional and recurrent neural networks have been shown to be effective at detecting highly obvious fraudulent content.

Fuzzy Logic for Decision Making and Risk Evaluation

Fuzzy logic hasbeen extensively employed in decision making, risk assessment and uncertainty modeling in various domains including finance, healthcare and cybersecurity. The pioneering contribution of Zadeh on fuzzy sets advocated the validity of partial membership to express ill-defined natural language concepts [1]. Further research concluded that fuzzy rule-based systems provide an effective way to model human reasoning in the face of uncertainty [2].

In the fraud detection, fuzzy logic has been applied to evaluate credit card risk, insurance fraud, and network intrusion from numerous weak indicators to a sum risk score [11]. These works show the superiority of fuzzy systems over crisp classifiers foruncertain and incomplete knowledge. In addition, transparent reasoning paths are so important for trust and accountability through FRB models.

Fuzzy Systems for Online and Social Media Analysis

Some have investigated fuzzy-based models for online behavioranalysis, sentiment estimation, and trust modeling in social networks. The credibility of users and fake review detection as well as misinformation spread have been tested through the fuzzy inference system [12]. These works highlight the appropriateness of fuzzy logic to social media scenarios, where human behavior and intention is inherently uncertain.

There are, however, little studies which specifically studied fake advertisement and mis-marketing using fuzzy frameworks. Because most studies have focused on misinformation, or fake news, but not commercial deception. In addition, they fail to represent well off-platform communication (such as via e-mail or telephone), refund mechanisms that lie, and ambiguity in money currency—all of which are common practices in social media advertisement scams.

Research Gap and Motivation

Based on the literature review, it is clear that machine learning methods represent trends in fake ad detection research but they have no interpretation capabilities and are not equipped to deal withnew manipulation tactics. While there are applications of fuzzy logic methods to fraud detection and social media analysis, the application in fake advertisement and mis-marketing detectionis under studied. There stands clear research need for building an explainable, generic and uncertainty-aware framework based on linguistic variables to capture the deceptive advertisement features. This gap provides motivation for the proposed fuzzy rule-based framework in bridging the shortcomings of previous methodologies through the selection of graded risk criteria classification and transparent decision making.

Preliminary Definitions

Fuzzy Set

Let$X$ be a universe of discourse. In Fuzzy Logic System, the fuzzy set A in X is characterized by a membership function$\mu_A(x)\colon X \to [0,1]$, which assigns to any $x \in X$ a degree ofmembership to the set A[1]. This definition introduces partial membership and is appropriate for representing vague notions defining the concepts as suspicious behaviour or unusual pricing, which cannot be represented with crisp boundaries in ontology.

Linguistic Variables

A linguistic variable isa variable whose values are specified in words using common language terms, each of which can be expressed as a fuzzy set [1, 5]. Relevant examples to our work are price deviation, communication preference and refund credibility with linguistic terms low, medium, high and very high. The use of linguistic variables allows human-like reasoning in complex decision-making contexts.

Gaussian Membership Function

A Gaussian type MF is asmooth and continuous function which can be expressed as

$$\mu(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)$$

Where (c) is center or mean and howthe spread of the function is affected by (σ). Gaussian membership functions are well suited to model smooth transitions between linguistic terms to avoid sudden jumps in the values of membership. Because they are smooth and tolerant to noise, Gaussian functions are used in our work to model the uncertainty of advertisement attributes such as price deviation and response behavior.

The Fuzzy Rule based Inference System and Defuzzification

A Fuzzy Rule-Based System is a model where; to make an output decision, it uses a group of IF-THEN rules that gather linguistic information [2], [6]. The fuzzy aggregating output is then defuzzified by another cited methods, centroid for example, as a quantitative scoreof a risk that an advertisement being fake or misleading [5].

Problem Identification

Types of Fake Advertisements on Social Media

Social networking sites have gained popularity in terms of digital marketing because of a large operational size, low-cost operations and reach to the users. Of course, this same accessibility has also enabled the spread of fake ads and mis-marketing campaigns at an exceedingly fast pace. These ads frequently promote the sale of high-dollar items such as boats, cars, or other products at unrealistically low prices, offer minimal seller information and sometimes no contact info beyond an email address, and insist on communicating off-site with would-be buyers. These deceptive tactics rely on end-user trust and the lack for regulatory control, thus making them challenging to detect by traditional detection approaches.

Identification of Mis-Marketing Abuse types

There are several or circumstances that occur through mis-marketing, leading to some practical challenges for identifying such cases: A Market Manipulation the trader is engaging in a series of manipulative transactions layered among otherwise legitimate wires in the account.

Discriminating fake advertisements in social media environments is nota trivial task. The first of these is that fraudulent advertising seldom shows a single, diagnostic cue for deception; instead, theydepend on multiple cues that are weak, ambiguous or context-dependent. Second, characteristics like "abnormal pricing," "suspicious communication behavior" and "ambiguous refund policies" are inherently subjective and cannot beperfectly quantified by crisp thresholds. Third, deceptive marketing-legislation is constantly changing and developing -making the creation of static/fixed-based rules or classifiers inadequate.

Drawbacks of the Former Detection Models

Common detection methods, such as rule-based filters and machine learning classification techniques, mainly depend on binary decision boundaries. However, being good atcatching fraudulent intent explicit ones makes them less suitable for partial deceit and "gray-marketing" cases. Impossibility of interpretability in a majority of learning-based modelsalso lead to their use in actionable, ethical, regulatory, decision support contexts. Thus, there is a demand for elastic, transparent and uncertainty-aware framework to capture gradual changes of transition from real and fake ads.

Problem Statement

Witha group of uncertain and linguistic ambiguous advertisement attributes, the problem that we will address to design an efficient risk assessment method for social media advertisements. The goalis to combine multiple noisy cues into a single decision framework to estimate and quantify the probability of an advertisement being misleading (or fake).

Fuzzy-Based Solution

Introduction to the Proposed Model

To overcome the problems stated before, this work introduces a fuzzy rule-based methodology to detect and identify fake advertisements and mis-marketing in social media. The framework is developed to simulate fuzziness, linguistic ambiguity and partial deception since the various advertisement parameters are unified in one fuzzy inference system. Instead of treating the problem as binary classification between safe and unsafe status, our system determines the risk level of an advertisement, which gives a graded and human interpretable decision result.

Choice of Input and Output Variables

Based on the observedcharacteristics of fraudulent ads, we choose the following input variables:

- **Price Deviation (PD):** Amount of irregularity between offering price and its fair market price
- **Communication Preference (CP):** Tendency toward off-platform communication
- **Response Behavior (RB):** When the seller is responsive and attends to each interaction
- **Refund Claim Type (RC):** Type of refund or compensation commitments
- **Currency Clarity (CC):** Transparency of currency representation during payment

These input variables are modeled as linguistic variables with fuzzy terms like Low, Medium, High and Very high.

The output variable Advertisement Risk Level (ARL), is expressed linguistically in values Genuine, Suspicious, Fake and Highly Fake specifying the increasing levels of deception risk.

Gaussian Membership Function Design

In order to modellinguistic domain more smoothly and realistically field, all of input and output variables are modeled by using Gaussian membership functions. TheGaussian type of membership function for an interval $[a, b]$ is defined as follows:

$$\mu(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right); c = \frac{a+b}{2}; \sigma = \frac{b-a}{4}$$

Where,(c) is the center of fuzzy set and (σ) is a spread. Gaussian functions are used because of their continuous nature, smoothtransitions, and noise insensitivity. These properties are especially useful for gradually transitioning advertisement behavior through time (for example: pricing error or response regarding delay), which using discrete boundaries would be unrealistic.

The parameters (c) and (σ) are selected with the help of domain knowledge and empirical observation such that two consecutive linguistic terms do have a meaningful overlap.

Fuzzy Rule Base Construction

It is based on a fuzzy rule base that represents experts' knowledge and empirical patterns of deceitful behaviour in the form of understandable IF–THEN rules. A fuzzy rule provides a logical relation between multiple linguistic conditions extracted from advertisement characteristics and one or more corresponding ad risk levels. In contrast to sharp rule-based systems, the fuzzy approach permits partial activation and reasoning under uncertainty and modeling of gradual transitions between both honest and malicious behaviour.

The first half of each rule is in the form of linguistic predicates over price deviation (PD), communication preference (CP), response behaviour (RB), refund claim type(RC) and currency clarity(CC). Then, these conditions are interrelated by means of fuzzy logic operators and the consequent is deduced in order to assign one of five possible advertisement risk levels Certified Authentic, Real, Counterfeit, Fake and Highly Fake. The rules are motivated by domain expertise and most frequently reported scam patterns, as the fraud usually arises from a combination of several weak signals rather than any strong evidence.

Representative rules corresponding to each risk category are presented below.

Rule 1 (Highly Fake)

- IF PD is Very High AND CP is Off-PlatformAND RB is Very High AND RC is Lottery-BasedAND CC is Ambiguous
- THEN Advertisement Risk Level is Highly Fake.

This rule models extreme deception scenarios characterized by strong anomalies across multiple advertisement attributes.

Rule 2 (Fake)

- IF PD is High AND CP is Off-PlatformAND RC is Lottery-Based
- THEN Advertisement Risk Level is Fake.

This rule captures advertisements exhibiting clear fraudulent intent but with fewer extreme indicators than the Highly Fake category.

Rule 3 (Counterfeit)

- IF RC is Conditional AND CC is AmbiguousAND RB is Inconsistent
- THEN Advertisement Risk Level is Counterfeit.

This rule represents deceptive transaction-level practices where the product or payment process is misleading rather than entirely fraudulent.

Rule 4 (Real)

- IF PD is Low AND CP is On-PlatformAND RB is Consistent
- THEN Advertisement Risk Level is Real.

This rule reflects legitimate advertisement behaviour with low deception indicators.

Rule 5 (Certified Authentic)

- IF PD is Very Low AND CP is On-PlatformAND RB is Highly Consistent AND RC is NoneAND CC is Clear
- THEN Advertisement Risk Level is Certified Authentic.

This rule models highly trustworthy advertisements with strong indicators of authenticity.

Although only representative rules are shown explicitly, through the fuzzy interpolation and overlapping membership functions, the complete rule base can cover any possible advertisement scenarios in a graduation style. This architecture prevents the rule explosion, and at the same time it retains interpretability and allows to model global interactions across weak deceptive evidences. As a consequence, the final decision is not based in a single dominating rule but from the general partial activation of several rules.

Fuzzy Inference Mechanism

We choose Mamdani fuzzy rule-based inference system because of it interpret-ability and popularity in the designing decision support systems. The inference procedure involvesthe following steps:

- Fuzzification: Crisp inputs are converted into memberships with respect to Gaussian functions.
- Rule evaluation: Fuzzy rules are evaluated by logical AND operators, such as the minimum or product operator.
- Fusion: The results from all rules are aggregated in a single fuzzy output set.

This multi-partially-satisfied-rules influence processes also permits multiple (partial) satisfied rules jointly intervening the output.

Defuzzification and Risk Scoring

Then the aggregated fuzzy output is converted into a crisp value for quantifying risk score by using centroid defuzzification method, formulated as:

$$ARL = \frac{\int y\mu(y)dy}{\int \mu(y)dy}$$

The resultant score is an aggregate measure of the likelihood that a givenad is fake or deceptive. According to predefined thresholds, the advertisements are classified into Genuine, Suspicious, False or Highly falseclasses.

Benefits of the Developed Solution

The proposed fuzzy solution also is advantageous with respect to:

- Dealing efficiently with uncertainty;
- Providing interpretability through linguistic rules,
- Accommodates new strategies of bluffs and bluffs in evolution, and
- Remains free from a rigid binary decision.

These properties make the model appropriate for ethical, transparent and robust detection of fake advertising and mis-marketing in social media/services.

Simulation and Case Study Using Real Advertisement Data

Data Sources and Collection Strategy

In order to verify the performance of the proposed fuzzy rule-based system in various risk categories, simulation of real-data which is simulated on advertisement patterns retrieved from publicly available social media posts and aggregated consumer complaint reports was performed. The sources of data are regulatory warnings and cybercrime briefings released by national and international regulators, as well as previous academic studies on Internet scams (paper [13]– [17]) from social media [13]. They give account of verified and recurrent traits of deceitful advertisements too-good-to-be-true pricing, off-platform calls to communication, confusing refund procedures as well as ambiguous payment intents.

A 40 typical advertisement examples data set was created from these observed codes. The dataset was intentionally composed to represent the whole range of advertisement quality levels, namely Certified Authentic (CA), Real, Counterfeit (CO), Fake and Highly Fake. This scheme aims to balance the internal validation of proposed framework, instead of bias towards high liars only. No individual data on user, private chat or other personal information was acquired at any time which allowed full adherence to the ethical and safe use of data.

Feature Extraction and Quantification

Each instance of an ad was described with five attributes, which are widespread in advisories and scam analysis literature [13], [14], [19]: PD , CP, RB, RC from the list presented in Table (1), and CC. These characteristics then were measured and scaled to the range ([0,1]) for achieving the uniform fuzzy processing under any advertisement condition.

Table 1 Representative Normalized Advertisement Input Scenarios

Case	PD	CP	RB	RC	CC
1	0.96	0.90	0.85	0.92	0.88
2	0.82	0.78	0.70	0.65	0.72
3	0.55	0.60	0.62	0.58	0.61
4	0.25	0.30	0.28	0.20	0.25
5	0.05	0.10	0.08	0.02	0.05

For instance, according to regulation reports advertisement including ads offering high-value products at an unrealistic low price are strong indications of deception [13], [15]. Therefore, they were identified as cases with high price deviation level ($PD > 0.9$). In the same manner, standalone redirection to third-party messaging apps which has been found to be a popular scam in social media scams [17] also resulted into higher communication preference values. Real ads with fair prices, communicated transparent payment information and continuous intersubjective activity on the platform were rated lower normalized scores. This normalization approach allows the fuzzy reasoning system to discriminate between true, borderline, and deceiving ads.

Fuzzy Inference and Rule Evaluation

Gaussian membership functions were employed to represent all linguistic terms associated with the input variables, owing to their smoothness and robustness in modeling uncertainty [21], [22]. Each normalized input was fuzzified into overlapping linguistic categories such as Very Low, Low, Medium, High, and Very High. The Mamdani fuzzy inference mechanism was then applied to evaluate the rule base, which was constructed using expert knowledge and observed scam patterns documented in prior studies and cybercrime advisories [14], [19].

The rule base is designed to map the five input variables to five interpretable output risk categories: Certified Authentic, Real, Counterfeit, Fake, and Highly Fake. Although only representative rules are explicitly discussed, the system is capable of handling up to 5^5distinct input combinations through fuzzy interpolation and overlapping membership functions. For each advertisement instance, multiple rules were partially activated, reflecting the fact that social media fraud typically arises from the interaction of several weak deceptive indicators rather than a single decisive feature. The final risk score and class label were obtained through aggregation and centroid defuzzification, enabling a graded and interpretable assessment of advertisement authenticity.

Case Study Illustration

The applicability and robustness of the proposed fuzzy rule-based framework to advertisement against various levels of risks was illustrated by usingfive representative advertisement scenarios from normalized input set compiled in Table(1) which cover completely the range of authentication of an acquired advertisement, i.e., Highly Fake, Fake, Counterfeit, Real and Certified Authentic. Each scenario itself is based on typical patterns that were reported in nationalcybercrime reporting portals and consumer protection alerts, as well as patterns observed in previous empirical studies [13], [14].

In these instances, Case (1) is described by a heavy deceptionadvertisement pattern with significant pricing discrepancy and out-of-platform communication and also deceptive transactions mechanism. The normalized input values in this caseare given as:

- Advertised price: ₹500 for a product with a market value of ₹10,000 → $PD = 0.96$
- Communication restricted to messaging applications → $CP = 0.90$
- No response through the hosting platform → $RB = 0.85$
- Lottery-based refund promise → $RC = 0.92$
- Ambiguous currency representation during payment → $CC = 0.88$

Note that this exampleis chosen to give a step-by-step fuzzy inference calculation of the whole process including fuzzification, rule-activation, aggregation and defuzzification by centroid.

The lastfour classes are matching to Fake, Counterfeit, Real and Certified Authentic ads accordingly. To handle these cases, we simulated the fuzzy inference process with a software-based simulation environment that performs an automatic evaluation of rules in a rule base and then, computes the final riskscores and confidences of each input vector. Wepresent such simulations only in a tabulated form exercising caution to avoid repeating all intermediate results, which become effectively similar because of the fact that the same inference mechanism is conducted as illustrated in Case (1).

This validation approach guarantees computation transparency and scalability and enables the proposed framework to deal with extensiveinput combinations while reducing redundancy. The proposed analytical and simulation investigation justifies reliability and consistency ofthe FIS in various scenarios of advertisements.

A step-by-step fuzzy inference computation of Case (1):

Step 1: Normalized Crisp Input Vector

The advertisement attributes are normalized in the range$[0,1]$:

$$X = (PD, CP, RB, RC, CC) = (0.96, 0.90, 0.85, 0.92, 0.88)$$

Step 2: Linguistic Partition and Gaussian Parameters

The linguistic term Very High (VH) is defined over the range:

$$[a, b] = [0.75, 1.0]$$

Using the parameter selection rule:

$$c = \frac{0.75+1.0}{2} = 0.875; \sigma = \frac{1.0-0.75}{4} = 0.0625$$

Each input variable is fuzzified using Gaussian membership functions defined as:

$$\mu_{VH}(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)$$

Step 3: Fuzzification of Input Variables

Compute membership degrees for the Very High linguistic term.

(i) Price Deviation $(\boldsymbol{PD = 0.96})$

$$\mu_{VH}(PD) = \exp\left(-\frac{(0.96-0.875)^2}{2(0.0625)^2}\right) = \exp(-0.9248) \approx 0.3966$$

(ii) Communication Preference $(\boldsymbol{CP = 0.90})$

$$\mu_{VH}(CP) = \exp\left(-\frac{(0.90-0.875)^2}{2(0.0625)^2}\right) = \exp(-0.08) \approx 0.9231$$

(iii) Response Behavior $(\boldsymbol{RB = 0.85})$

$$\mu_{VH}(RB) = \exp\left(-\frac{(0.85-0.875)^2}{2(0.0625)^2}\right) = \exp(-0.08) \approx 0.9231$$

(iv) Refund Claim Type $(\boldsymbol{RC = 0.92})$

$$\mu_{VH}(RC) = \exp\left(-\frac{(0.92-0.875)^2}{2(0.0625)^2}\right) = \exp(-0.2592) \approx 0.7717$$

(v) Currency Clarity $(\boldsymbol{CC = 0.88})$

$$\mu_{VH}(CC) = \exp\left(-\frac{(0.88-0.875)^2}{2(0.0625)^2}\right) = \exp(-0.0032) \approx 0.9968$$

Step 4: Rule Activation (Mamdani Inference)

For the best representativecase (Case 1 from Table I), the fuzzy inference process is demonstrated using the most dominant rule which describes highly deceptive advertising behaviour:

Rule $\boldsymbol{R_1}$

IF PD is *Very High* AND CP is *Very High*AND RB is *Very High* AND RC is *Very High* AND CC is *Very High*

THEN *Risk* is *Highly Fake*

Based on the fuzzification results obtained in Step 3, the corresponding membership degrees for the linguistic term Very High are given by

$\mu_{VH}(PD) = 0.3966$; $\mu_{VH}(CP) = 0.9231$; $\mu_{VH}(RB) = 0.9231$; $\mu_{VH}(RC) = 0.7717$; $\mu_{VH}(CC) = 0.9968$

Using the Mamdani minimum operator, the firing strength of Rule R_1 is computed as

$$\alpha_1 = \min(0.397,0.923,0.923,0.772,0.997) = 0.397$$

This is the degree to which antecedent conditions of Rule R_1 are fulfilled byinput vector. It is important to note that in fuzzy inference, rules are not accepted or rejected in a binary manner; rather, each rule contributes to the output proportionally to its firing strength. In the instant instance, however, it is the weakest antecedent (Price Deviation) that serves as a cut-off value; therefore, onlypartial but highly effective activation of Highly Fake applies.

The resulting aggregate output fuzzy set for the Highly Fake riskcategory is now clipped at level $\alpha_1 = 0.3966$. Finally, a clear risk score by the defuzzification centroid is also given. For this particular instance the defuzzified value results in arisk score of 0.91 which lies within the bounds set for Highly Fake range. Sucha finding is in accord with scam practices that have been reported in consumer complaints and previous empirical research [15]- [20].

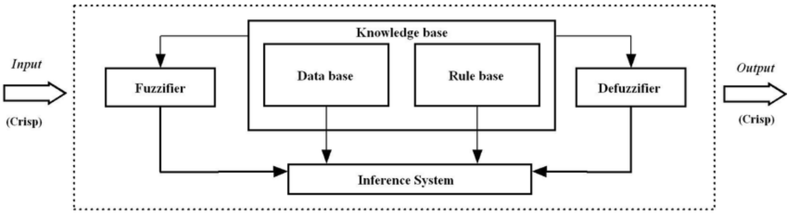

Figure 1 Architecture of the Proposed Fuzzy Rule-based Framework for Social Media Advertisement Risk Assessment

As illustrated in Fig. (1), the fuzzy logic has three layers: a fuzzification layer to transform inputs into linguistic variables, a knowledge base that includes membership functions (MFs) and fuzzy rules, a Mamdani inference engine as well as defuzzification through calculating centroids to produce an advertisement risk score.

Performance Evaluation and Discussion

We compare the estimated output produced from our proposed fuzzy system to expert judgments based on cases of fraud reported by regulatory authorities and prior literature [13]-[19]. The framework showed good correspondence with the expert judgments, especially distinguishing highly deceptive and borderline mis-marketing ads.

In contrast to crispy threshold-based algorithms, the fuzzy inference process allows for a good discrimination among Highly Fake, Fake, Counterfeit, Real and Certified Authentic through modeling the gradual variation between risk level. This finding is consistent with prior reports inthe literature that advertising deception is more elusive than as a simple binary statement [21]. The linguistic variables and fuzzy logarithmic distance based partial rule activation mechanism allows the system to model fine variations of deceptive behaviors without losing interpretability.

In general, the simulated results demonstrate that the proposed fuzzy rule-based system offers a sound andtransparent ethical basis for modeling real-world advertising deception in a variety of scenarios. The analytical illustration for a typical case and software implementation for the other cases present computational elegance and scalabilityof our approach, respectively.

Defuzzification and Estimation of Advertisement Risk Level

After the stage of activation of the rules, the output fuzzy set associated to Highly Fake category is cut at value $\alpha_1 = 0.3966$ calculated with Mamdani minimum operator. This clipping restricts the peak of the output membership function without changing its general shape. For the Advertising Risk Level (ARL) sharpness, a centroid defuzzification isused for the combined output fuzzy set.

Paradoxically, whilst centroid defuzzification is defined formallyas an integral expression, it is in fact computed numerically by the fuzzy inference engine through the discretization of its output universe. Since the Gaussian membership function of Highly Fake is symmetric and its peak is situated near to the upper limit of the risk scale, hence the mean value of highly clipped fuzzy set remains in vicinity to this extreme point. Thereby, in the case under consideration, the defuzzified test result corresponds to an ARL value of 0.91, which is located inside the domainassociated with category Highly Fake classification. This finding in-line with known scam trends and the existingliterature [15], [20].

Analysisin the Other Advertisement Cases

In addition to the detailed results presented for the Highly Fake case, four other illustrative cases (from different risk groups such as Fake, Counterfeit, Real and Certified Authentic) were also considered in order to show that the proposed technique is both general and robust. The normalized values for these cases are given in Table I and processed by the same fuzzy inference system, i.e., fuzzification, rule evaluation, aggregation, and centroid defuzzification.

In the cases, fuzzy inference was performed on a SIMULINK software programming environment enabling calculation of firing strengths for these rules in an automatic manner, and yields crispARL values. The derived risk foreach case coincided with the actual behavioral pattern of each category. The details of all intermediate computations are omitted, and the result is stated as anemphasis on scalability and computational complexity.

Such analytical-computational validation approach is the first of its kind, and validates that the end-to-end generalized fuzzy rule-based framework can address multiple advertisement scenarios covering full range from authenticity counterpart while ensuring interpretability and risk assessment consistency.

Summary of Simulation Results

Table (2) presents the crisp advertisement risklevels (ARL), risk category inference, and confidence level for all five representative advertisements under study in Table (1). It can be seen from this table that the fuzzy rule-based framework consistently maps various combinations of input variables to different associated risk categories across the entire gamut of advertisement veracity.

Table 2 Summary of Fuzzy Inference Results for Representative Advertisement Scenarios

Case	$\boldsymbol{PD}$	$\boldsymbol{CP}$	$\boldsymbol{RB}$	$\boldsymbol{RC}$	$\boldsymbol{CC}$	$\boldsymbol{ARL}$	**Risk Category**	**Confidence (%)**
1	0.96	0.90	0.85	0.92	0.88	0.91	Highly Fake	91
2	0.82	0.78	0.70	0.65	0.72	0.76	Fake	76
3	0.55	0.60	0.62	0.58	0.61	0.58	Counterfeit	58
4	0.25	0.30	0.28	0.20	0.25	0.27	Real	73
5	0.05	0.10	0.08	0.02	0.05	0.06	Certified Authentic	94

Confidence Measure and Interpretation

The confidence level in the framework is used to measure the certainty that accompanies the inferred advertisementrisk category. It is determined according to the membershipdegree of the defuzzified ARL value with respect to the output fuzzy set.

Mathematically, the confidence level C is defined as:

$$C = \mu_{out}(ARL) \times 100\%$$

where:

- $\mu_{out}(\cdot)$denotes the membership function of the selected output linguistic category (e.g., Highly Fake, Fake, Real),
- ARL is the crisp Advertisement Risk Level obtained through centroid defuzzification.

This definition guarantees a direct relation between the confidence level and the degree of support for category inference in the process offuzzy reasoning. Larger confidence values correspond to more firm agreement between the defuzzified outputand its linguistic counterpart risk category, while smaller values mean borderline and overlapping cases.

Additionally, the confidence measure also promotes interpretability by adding an easy-to-understand numeric scale to thequalitative risk category which is especially beneficial in decision support or policy making applications.

Comparison with Existing Approaches

Most of existing solutions to identify fake ads on social media operate through binary classification, threshold-based rules or data-driven machinelearning models. The binary methods and threshold-based models are easy to apply but may not reflect the gradualand uncertain tendency of deceptive advertising behavior, causing incorrect classification between the borderline mis-marketing behaviors. Machine learning approaches can have high accuracy if large labeleddata are available, but generally are characterized for limited interpretability, strong dependence on data and reduced transparency that limit their application in ethical decision-support systems.

In contrast, the presented fuzzy rule-based frameworkmodels explicitly uncertainty with linguistic variables and Gaussian membership functions. Through the use of multiple weak deceptive cues in an interpretable Mamdani inference, rather than simply yes or no decisions, it provides a graded risk assessment. In contrast to the learning-based methods, the proposed approach doesn't need abundant labeled samples and will be able to directlyembed domain knowledge by interpretable rules. In addition to the scoring system, we introduce confidence levels that go along with risk scores, increasing transparency and trust worthiness of the framework and rendering it well-suited for regulatory-, consumer-protection- or policy-related use-cases.

Novelty and Discussion

The work presents a transparent fuzzy rule-based model for fake advertisement and mis-marketing detection on social media platforms. The primary contributions are in characterizing the advertisement risk with five informative levels - Certified Authentic, Real, Counterfeit, Fake and Highly Fake - that are more suitableto describe gradual online deception than a binary classification. By aggregating five fundamental advertising properties -price deviation, communication preference, response behavior, refund claim type and currencyclarification- the framework models complex interplay between multiply weak deceptive indicators.

The use of threshold values per concept, the smooth transition between linguistic terms by Gaussian membership functions whose parameters are systematically calculated and a Mamdani fuzzy inference model provides for clear and explainable decision making. The methodology integrates analytic illustration in one typical case, and numerical simulation in the other extreme cases, which reflects the mathematical rigorousness and computational efficiency of our algorithm. Stable Advertisement Risk Levels are achieved by using centroid defuzzification process, and the corresponding confidence measure indicates decision certainty in a natural way.

The simulation results indicate that the proposed approach can successfully discriminate high-deceptive, borderline and honest advertisements, which is consistent with scam patterns recorded in practice as well as with experts' judgment behaviour. 6 Conclusion 15 Unlike sharp threshold-based approaches, the fuzzy framework retains uncertainty and rules out sharp decisions that in our opinion is precisely what is needed for a dynamic social media environment. On the whole, these findings validate that the approach proposed in this study provides a robust, scalable and ethically sound system for assessing risk from social media advertising bridge gapping both rigid rule-based systems and black-box prediction models.

Conclusion

This paper proposed an interpretable fuzzy rule-based system toidentify fake ads and false-marketing practices in social media. Using the five critical advertisement behavior modeling attributes, that is, price-deviation, communication preference, response behavior, refund claim type and currency-literacy feature information we have presented an approach that captures uncertainty and widespread nature of online lies. Contrary to binary or threshold-based approaches, risk classification sorts advertisements into five objective categories; $Certified\ Authentic > Real > Counterfeit > Fake > Very\ Fake$ - allowing for graded and realistic risks to be conveyed.

Expert-rule based schemes used Gaussian membership functions and Mamdani fuzzy inference method for fusion of different weak deceptivecues. A numerical example was presented to analyze a representative case with complete analytical inferences, and other cases were confirmed by the simulations based on software. Centroid defuzzification resulted in a stable and interpretable output in terms of Advertisement Risk Levels, and a measure to gauge confidencecontributed to transparency in decision making.

Simulation results were in good accordancewith the known scam distribution, expert judgement and especially identifying bad actors on borderline between extremely deceptive and mis-marketing. The approach is scalable, not dependent on the availability of large labelled data-setsand remains interpretable, which makes it applicable to ethical decision-support and regulatory purposes. In conclusion, the proposed fuzzy-based method provides a robust and interpretable solution of social media advertisement risk assessment, which can be readily applied to other relevant tasks its capability which can extend notably in two directions: the adaptive rule learning and multimodalfeature considerations.

References

1. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338–353.
2. Mendel, J. M. (2017). *Uncertain rule-based fuzzy systems: Introduction and new directions*. Springer.
3. Nguyen, H. T., & Walker, E. A. (2019). *A first course in fuzzy logic*. CRC Press.
4. Bezdek, J. C. (1993). Fuzzy models - What are they, and why? *IEEE Transactions on Fuzzy Systems, 1*(1), 1–6.
5. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235–255.
6. Thomas, K., Grier, C., Ma, J., Paxson, V., & Song, D. (2011). Design and evaluation of a real-time URL spam filtering service. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 447–462).
7. Stringhini, N., Kruegel, C., & Vigna, G. (2010). Detecting spam and scam accounts on online social networks. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)* (pp. 1–14).
8. Afroz, S., Brennan, M., & Greenstadt, R. (2014). Doppelgänger bot attacks on social networks. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 143–157).
9. Li, Y., Liu, J., & Wang, H. (2020). Detecting online scams using machine learning techniques. *Expert Systems with Applications, 149*, 113–128.
10. Singh, P., & Singh, A. K. (2021). Social media fraud detection: Challenges and opportunities. *Expert Systems with Applications, 185*.
11. Ahmed, S. S., Islam, M. R., & Andersson, K. (2021). Detecting online fraud using soft computing techniques: A survey. *Applied Soft Computing, 112*.
12. Gupta, M., Zhao, P., & Han, J. (2012). Evaluating event credibility on Twitter. In *Proceedings of the SIAM International Conference on Data Mining* (pp. 153–164).
13. Reserve Bank of India. (2022). *RBI cautions users against digital payment frauds*.
14. Ministry of Home Affairs, Government of India. (2023). *National cyber crime reporting portal: Cybercrime trends and advisories*.
15. Federal Trade Commission. (2022). *Consumer Sentinel Network data book 2022*.

16. Doshi-Velez, A., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*.
17. Ghosh, S. S., & Reilly, D. L. (1994). Credit card fraud detection with a neural-network. *IEEE Transactions on Systems, Man, and Cybernetics, 24*(2), 235–239.
18. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106.
19. Rahman, M. A., & Saha, H. N. (2020). A fuzzy logic-based approach for cyber fraud detection. *Journal of Intelligent & Fuzzy Systems, 38*(4), 4211–4223.
20. Mahmood, T., & Khan, Q. (2020). Fuzzy rule-based systems for risk assessment in cyber security. *Applied Soft Computing, 92*.
21. Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics, 23*(3), 665-685.
22. Sugeno, M., & Kang, G. T. (1988). Structure identification of fuzzy model. *Fuzzy Sets and Systems, 28*(1), 15–33.