Tampere University of Applied Sciences

Fake Reviews in Chinese E-Commerce

The Case of Wine

Geoffrey Brink

MASTER’S THESIS
November 2022

Master’s Degree in International Business Management

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Master’s Degree in Business Administration
International Business Management

GEOFFREY BRINK:
Fake Reviews in Chinese E-Commerce
The Case of Wine
Master's thesis 46 pages
November 2022

The growth of the Chinese wine market has attracted the interest of wine exporters around the world. However, the Chinese market is very large and business conditions vary widely between regions. E-commerce appears to be a solution that is cheaper and simpler than alternatives and to be well established in the country as well. Yet many reports suggest that a large number of online reviews might be fake and are simply commissioned by merchants. This is worrying, as online reviews are crucial to sales performance, influencing purchasing decisions and search result rankings. To examine this problem, over 10,000 wine reviews were scraped from Jingdong, the second largest Chinese online retailer, and analysed using text classifiers created with autoML. The results lend credence to the suspicion that most wine reviews on Jingdong are fake.

Key words: fake review detection, jingdong, autoML

CONTENTS

1 INTRODUCTION

In the last few decades, China has become a major wine market, drawing interest from wine exporters abroad. Its expanse and heterogeneous business conditions pose a challenge to potential entrants, however, which online retail seems the best way to overcome. In principle, selling online would enable merchants to serve the largest market while keeping operations simple and overheads low.

A problem arises in online reviews. They are of great value to users, as they inform them about product quality, helping them to decide which goods to purchase. They are important to merchants as well, especially as a large volume of positive ones results in their products appearing prominently in search rankings and customers being more likely to choose their products over those of competitors. Vendors thus have a strong incentive to do whatever it takes to keep good reviews coming. This leads some of them to commission deceptive reviews.

Academic literature, journalism and widespread public opinion hold that a large number, possibly the majority, of reviews on Chinese e-commerce websites are fake. To explore this possibility, ten thousand wine reviews were scraped from Jingdong, a major Chinese online retailer, and were run through three different text classification models created with no-code automatic machine learning (autoML) frameworks. The vast majority of wine reviews analysed were found to be suspicious. This lends support to the prevailing view that Chinese e-commerce is riddled with fake and deceptive reviews.

2 WINE IN CHINA

2.1 The market

From humble beginnings, China has emerged as one of the most important wine markets in the world:

- The Chinese drink more red wine than any other nation (Sun 2015) and China is tied at sixth place as the largest consumer of all wines (OIV 2022).
- In 2015, the country drank seven times as much wine as Japan, itself the second largest wine consuming nation in Asia, and a hundred times as much as India (Qing et al. 2015).
- China has the third largest amount of vineyard in the world (OIV 2022).
- Between 2000 and 2011, China's import of wine increased 26,000% (Muhammad et al. 2014).
- Between 1970 and 2009, annual growth in Chinese wine consumption increased 145,541% (A. Morrison 2014).

None of this would have seemed plausible a few decades ago. How did it come to pass?

2.2 Background

Alcohol has been consumed in China since time immemorial. But the country's present consumption habits can be traced to the time of Mao and his desire to have his compatriots replace their consumption of hard grain spirits with wine. Wine production was established in the 1960s to meet this objective, and French and Australian firms began to set up vineyards in the 1980s as the country began to open up to the world (Li 2018).

In the same decade, the National Winemaking Conference proposed a Four Changes strategy to cut down national alcohol consumption and to allow more grain to be used for food than liquor production. An important part of this plan was the encouragement of wine consumption (Zeng and Szolnoki, 2017).

2.3 Production

FIGURE 1: Map of Chinese Vineyard (Li and Bardají 2017, p. 2)

Today, China boasts vineyard that stretches from one end of the country to the other, as illustrated in Figure 1. But it was not always so. Li and Bardaji (2017) write that China's production of wine remained meagre throughout the eighties, and its primary wine product was, in fact, a mixture of wine, juice, sugar and water. Progress was made in the next decade, as adulterations containing less than 50% wine were banned in 1994[1]. Moreover, total vineyard increased from 31 600 ha to 830 000 ha between 1980 and 2015. Though roughly 90% of this grape cultivation went into preparing table and dried grapes, wine production increased nonetheless from 0.78 million hectolitres[2] to 11.50 million hectolitres between 1980 to 2015 (ibid).

---

[1] Adulterated wine of any sort was prohibited in 2004 (Li and Bardají 2017.
[2] One hectolitre equals one hundred litres.

2.4 Wine consumption with Chinese characteristics

Research on the behaviour and preferences of Chinese wine drinkers began late, with a seminal paper by Liu and Murphy appearing only in 2007. Several themes emerged in that paper, as well as in other early research, that present a consistent picture of Chinese wine consumption.

Subjects of these studies appeared to know little about wine (Li et al. 2011; Sun 2015). For example, Liu and Murphy’s sample believed all wine to be red. Overall, red wine was preferred to white. Researchers attribute this preference to Chinese culture, in which the colour red is perceived as auspicious, while white is associated with funerals and mortality (Liu and Murphy 2007; Li et al. 2011; Sun 2015).

Chinese wine drinkers liked sweet wines and were prone to mixing wine with soft drinks to improve its flavour and reduce its alcohol content (Somogyi et al. 2011).

Wine was perceived as a status good, the consumption of which was good for one’s image, and the giving and receiving of which as giving face to the parties involved. Consequently, corked bottles were preferred to screwtops (these being perceived as cheap), and one would only buy cheap wines for oneself – expensive wines being reserved as gifts for others (Ibid; (Yu et al. 2009).

After domestically produced wine, French and other Old World wines were (and are) most popular with consumers in China (Qing et al. 2015; Agnoli et al. 2014). Indeed, one in every five bottles of Bordeaux is exported to China today (Wang 2022a).

Wine was also consumed for its putative health benefits (Somogyi et al. 2011; Liu and Murphy 2007; Masson et al. 2017), being seen as good for you in traditional Chinese medicine (TCM) (Somogyi et al. 2011).

The past tense is employed in the above paragraphs because there are signs that the behaviour of Chinese wine drinkers has been changing rapidly and that the picture we have of Chinese wine consumption might no longer be accurate.

Though red wine remains most popular with consumers, tastes have become more sophisticated over the last decade. Certainly everybody knows of the existence of wines other than red wine now (Chu et al. 2020). A study of young consumers found that many of them prefer white to red wine, though they express interest in becoming more discerning drinkers of the latter (Fountain and Zhu 2016). It’s likely that fine wines are bought increasingly for personal consumption rather than only as gifts.

Chinese wine drinkers are better informed about wine, with enormously popular influencers having emerged to educate and advertise to consumers. Perhaps most prominent among them is Lady Penguin, a graduate of an Ivy League university who has millions of followers on social media and whose company’s annual revenue comes to RMB 300 million[3] (Wang 2021).

[3] EUR 437,628,00

FIGURE 2: Photograph of Wang Shenghan, Better Known as Lady Penguin, 2022

Researchers observing wine festivals find attendees to have become progressively more expert in their knowledge of wine as well (Shi et al. 2022).

Finally, experts have disagreed with the idea that the palates of Chinese wine drinkers are distinct from those of wine drinkers elsewhere in the world (Wang 2020a) and research has found far less regional variation in wine consumption habits than was previously taken for granted in a country so vast and so populated (Meiselman 2020). For example, Williamson and co-authors found their sample of Chinese and Australian wine drinkers to have broadly similar palates, though Chinese participants were somewhat more amenable to sweetness and averse to bitterness in their wines. Notable as well was that these study samples varied little in their tastes despite being located in different parts of the country (P.O. Williamson et al. 2012).

2.5 Difficulties

There are a few things to bear in mind when considering wine consumption and the growth of the wine industry in China:

- Statistics attesting to the large size of the Chinese market conceal that per capita consumption remains low (Li and Bardají 2017).

- Counterfeiting is a serious problem that discourages potential consumers from purchasing wine, especially wine from abroad (Muhammad and Countryman 2019).

- Chinese wine consumption peaked in 2017 and diminished markedly during the COVID-19 pandemic. The sector is relatively unproductive and suffers from adverse climates and technological constraints (OIV 2022).

- While wine consumption and production have increased tremendously, beer is vastly more popular and spirit sales are about twice as lucrative, as indicated in Table 2 and 3 below.

TABLES 1, 2 and 3 of Alcohol Consumption, Sales and Production in China (Meiselman 2020, p. 578)

Table 1 World Wine Production 1996 and 2016 in millions of hectolitres (Source OIV 2017)

Country	1996	2016
Italy	54.4	50.9
France	56.3	43.5
Spain	34.1	39.3
USA	21.4	23.9
Australia	7.4	13.0
China	9.6	11.4[a]
South Africa	7.8	10.5
Chile	5.1	10.1
Argentina	13.5	9.4
Germany	10.0	9.0

[a]Wine production was about ten million hectolitres in 2018 (IBIS World 2019)

Table 2 Alcohol sales in China in millions of liters (Passport 2018)

Alcohol type	2012	2013	2014	2015	2016	2017
Spirits	4,940	5,284	5,370	5,586	5,696	5,777
Beer	48,993	50,582	50,582	47,727	45,627	45,535
Wine	**4,218**	**4,208**	**4,108**	**4,350**	**4,581**	**4,777**
RTD's/High strength Premix	31.9	49.5	118.7	152.8	102.4	89.6
Cider/Perry	0.51	0.54	0.61	0.69	0.81	0.95
Alcohol drinks	**58,184**	**60,125**	**59,678**	**57,816**	**56,007**	**56,181**

Table 3 Alcohol sales in China by value in CNY millions (Passport 2018)

Alcohol type	2012	2013	2014	2015	2016	2017
Spirits	925,452	849,797	774,756	804,868	888,591	999,675
Beer	402,116	448.397	481.432	511.069	537,193	572,675
Wine	**396,126**	**375,576**	**368,008**	**405,380**	**441,641**	**480,823**
RTD's/High strength Premix	1,724	2,691	6,077	7,073	4,534	3,910
Cider/Perry	1.1	1.2	1.4	1.6	1.9	2.3
Alcohol drinks	**1,725,419**	**1,676,462**	**1,630,274**	**1,728,392**	**1,871,960**	**2,057,085**

Despite all this, the country remains a major wine consumer, a prodigious producer (see Table 1), and an important export market for the beverage.

3 E-COMMERCE

3.1 Selling wine online in China

Online shopping is highly developed in China and is considered an important channel for selling wine. Estimates of how much wine is sold online vary from a high attribution of 51% of total wine sales, to a lower and perhaps more credible estimate of 29%[4] (Lucy Jenkins 2016; Meiselman 2020).

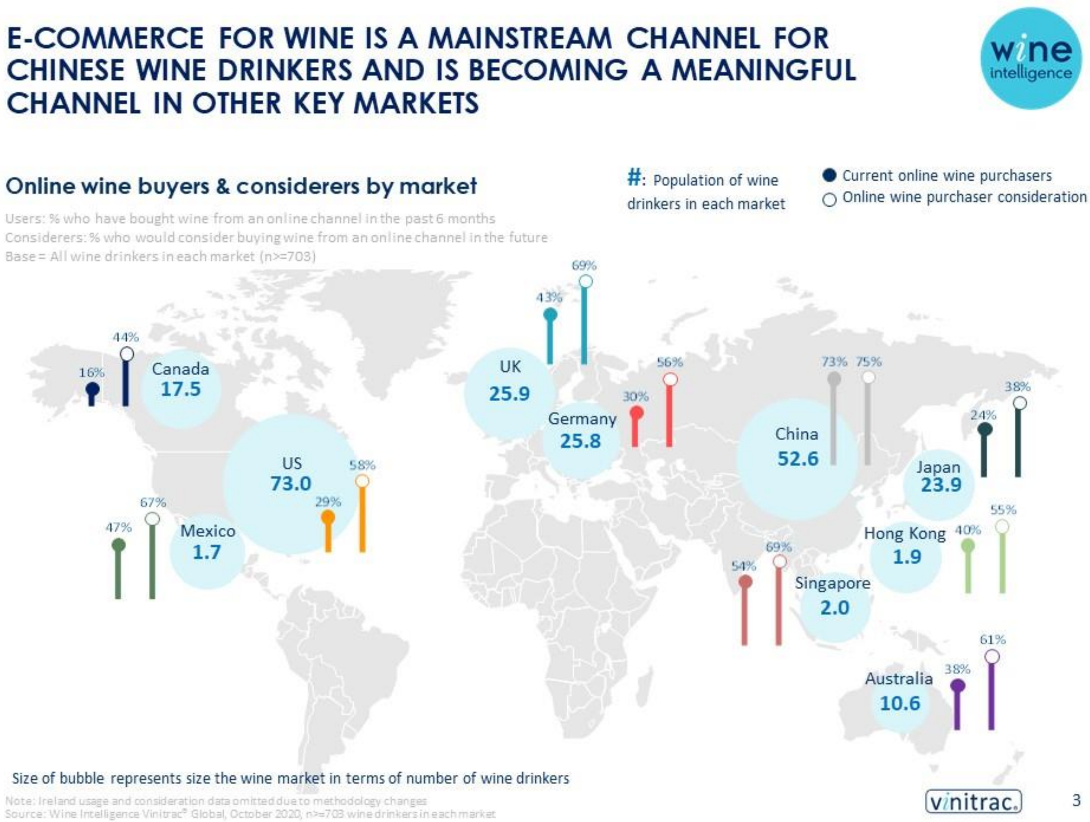

FIGURE 3: Online Wine Sales Around the World (Maxime Lu 2016)

Even this lower estimate is substantial. In countries like Italy and South Africa where viniculture is far more established, the vast majority of wine is still bought at supermarkets and specialist wine shops (Xavier 2016).

---

[4] The lower estimate might be more credible as it was produced in an academic publication. The high estimate is advocated for by researchers in industry who might have an interest in talking up the market.

3.2 Fake reviews

> Amazon has filed a lawsuit against administrators of more than 10,000 Facebook groups it accuses of coordinating fake reviews in exchange for money or free products. In the statement, Amazon said one of the Facebook groups it's targeting, called "Amazon Product Review," had more than 43,000 members. Amazon noted since 2020, it has reported more than 10,000 fake review groups to Meta, the parent company of Facebook. Meta has removed half of these groups and is investigating the others, Amazon said. (Amazon sues admins of 10K Facebook groups over fake reviews 2022)

Reviews are of prime importance in online shopping. It is typical in e-commerce that purchasers can score and post public reviews of the products they buy. Reviews are useful, providing potential buyers with information that enables them to pick out good products from bad (Hajek et al. 2020). Consumers have come to depend on them heavily when deciding what to buy (Plotkina et al. 2020; Eleanor Vaida Gerhards 2015).

Having large numbers of positive reviews is important to vendors as well. Researchers find that consumers are three times as likely to buy a product as not when it has a large number of reviews, and that an increase of a single star in a product's review increases the chance that it will be bought by 26% (Maheshwari 2019). The addition of a single star to a restaurant's Yelp rating can increase its revenue by 5-9%(Luca and Zervas 2016). Merchants have a strong incentive to do whatever they can to increase their good reviews, and some resort to paying companies or freelancers to write fake reviews to do so.

It is unknown how many online reviews are fake, though one expert estimates that half of them could be (Lieber 2021). More authoritative figures are available for specific websites:

- A third of TripAdvisor reviews are estimated to be fake (Ellson 2018)
- Yelp pre-emptively blocks 16% of reviews from being posted and removes a further 30% of reviews successfully posted, suspecting them to be fake (Glazer et al. 2021).

- Liu and co-authors found the content of 46% of Amazon mobile phone reviews to be irrelevant to the product reviewed, suggesting that the reviews were fake (Liu et al. 2017).
- Amazon claims to have prevented 200 million fake reviews from being posted on their websites in 2020 alone (Lieber 2021).

Fake reviews are evidently a serious problem in e-commerce. However, the above statistics referred mostly to Western platforms. What about China?

3.3 Fake reviews in Chinese e-commerce

By all indications, fake reviews are at least as prevalent in Chinese e-commerce as elsewhere. Indeed, fake reviews on Amazon appear to have increased after the platform began inviting Chinese vendors to operate on it, which suggests that the practice of commissioning fake reviews could be even more widespread among them (Thomas Hale 2019).

It is an open secret that faking sales and reviews, or brushing[5] as it is known in China, is the most effective way for online stores, and for new ones in particular, to succeed (Wu Qiuyu 2016; Simin Zhang 2021). Increasing video views, follower counts, posting positive reviews or having negative reviews removed – more or less anything that requires input from users is subject to such manipulation (Han Dandong 2022).

A great deal of money is involved. In 2017 the Ministry of Public Security announced that over RMB 100 million[6] had been spent on brushing in cases it was investigating. In addition, it had shut down over ten thousand websites and identified over ten million brushing posts in the same year (Song Chao 2018). In 2022, the Ministry arrested a further 2000 suspects, shut down more than six million suspicious accounts and dissolved 170 thousand instant messaging groups that had been run to facilitate the practice.

---

[5]刷單, 刷評 (shua dan, shua ping).
[6] EUR 145,876,00.

3.3.1 Modus operandi

The business models underlying brushing and the ways in which fake reviews are commissioned have been reported by journalists, who have in some cases gone undercover to investigate them.

According to these accounts, candidate brushers are invited to join (or must pay to join) instant messaging groups and websites in which admins and merchants post briefs (Yang Jinzhi 2011; Wu Qiuyu 2016).

In some of these groups, all members other than the group administrator are muted, and silently carry out the tasks assigned to them. In others, group members confirm that they will take on the commission advertised and are added to a new group by the admin in which they communicate about the work they are to do. Usually, these groups have around a thousand members. Their names are chosen to sound innocuous and appear unrelated to brushing. After the work has been done, the group is deleted. It is thus difficult for those complicit in this fraud to be tracked down (Han Dandong 2022).

Participants may receive general training before they begin posting deceptive reviews, and they receive instructions specific to their commissions in the project briefs that are posted in the groups. They might, for example, be instructed to wait for five minutes on the page of the product they are to review before posting, and they might be told to browse competitors' products and not to post reviews more than seven times a week. The point is to evade spam filters (Wu Qiuyu 2016).

The most pervasive type of fake review is simple copy-and-pasting, often of generic content posted in groups by their admins. The tone and wording of fake reviews is generic, and their contents often describe the product being evaluated only vaguely (ibid). Brushing is a tiered service, however, and premium fake reviews can be bought too, the texts of which are more elaborate and which have photographs or videos attached. But this is more expensive.

Fake reviewers are paid upon sending screenshots of their having successfully posted reviews (Han Dandong 2022). The average payment is between five and seven Yuan per review[7], although this increases if photographs and links are added to them (Wu Qiuyu 2016; Yang Jinzhi 2011).

Fake reviewers can increase their earnings by recruiting new brushers as well (Yang Jinzhi 2011).

3.3.2 Sanctions

That brushing is so pervasive in China happens in spite of government sanctions and counter-measures taken by e-commerce companies.

Under China's Anti-Unfair Competition Law, merchants found guilty of brushing face fines of up to a million Yuan – up to two million Yuan[8] in serious cases. They also risk losing their business licenses (Liao Jin 2017).

China's main online shopping websites[9] act against brushing in various ways; employing advanced technology to filter it out, collaborating with law enforcement, sanctioning merchants who engage in it, and suing persons and companies they suspect to be involved in it (Simin Zhang 2021; Liao Jin 2017).

Most e-commerce websites also tier their reviewers by purchase and review activity in order to help consumers to identify which reviewers are likely to be more credible. However, this has apparently long been gamed by brushers (Qin Chen 2012).

---

[7] EUR 0.72-1.
[8] EUR 143594, EUR 287188.
[9] Namely Alibaba, Jingdong and Taobao.

A search of patents for technology developed to identify fake reviews suggests that technological solutions for the problem of fake review detection are increasing as well.

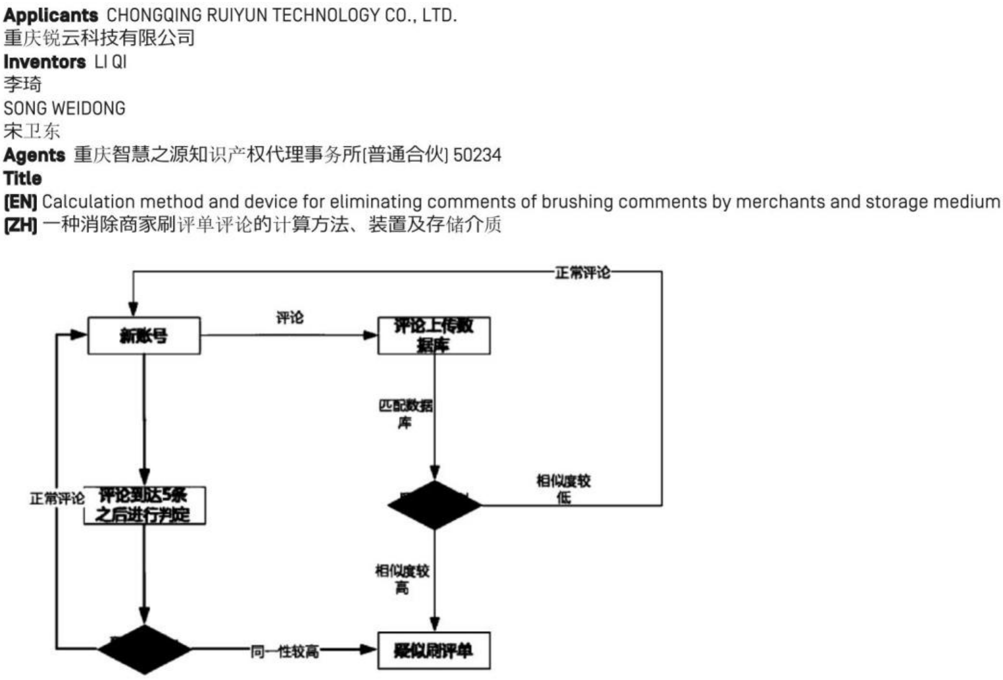
Applicants CHONGQING RUIYUN TECHNOLOGY CO., LTD.
重庆锐云科技有限公司
Inventors LI QI
李琦
SONG WEIDONG
宋卫东
Agents 重庆智慧之源知识产权代理事务所(普通合伙) 50234
Title
[EN] Calculation method and device for eliminating comments of brushing comments by merchants and storage medium
[ZH] 一种消除商家刷评单评论的计算方法、装置及存储介质

FIGURE 4: An Example of a Patent for FRD Technology (CN111310064 Calculation method and device for eliminating comments of brushing comments by merchants and storage medium 2022)

But the problem is deep rooted and all-encompassing. Indeed, on the 12th of July, 2022, a company belonging to Lady Penguin, star of the Chinese wine scene, was found guilty of posting fake negative reviews on the pages of a competitor. Its was fined RMB 120,000[10] and ordered to issue a public apology (Wang 2022b).

[10] EUR 17,505

4 RESEARCH METHODOLOGY

4.1 Objectives

For this thesis, a sample of wine reviews taken from Jingdong was analysed to get a sense of how common deceptive reviews are on the platform, for the category, and in Chinese e-commerce more broadly. Only the content of the reviews was analysed, which means that this analysis might well understate the extent of the problem. The behaviour and metadata of reviewers is out of consideration as it is difficult to observe and collect, and the minority of premium, high quality fake reviews, which are by nature harder to identify by their content alone, are outside the scope of the paper.

4.2 Detecting fake reviews

The first paper to address fake review detection (FRD) is considered to be that of Liu and Jindal, which appeared in 2007 (Heydari et al. 2015). Since then, the FRD literature has flourished. A number of systematic reviews of it exist that describe the field with greater range and in greater detail than can be done in this section, including Paul and Nikolaev 2021, Heydari et al. 2015, Mewada and Dewang 2021 and Vidanagama et al. 2020.

4.2.1 Approaches to fake review detection

The characteristics and limitations of current methods are summarised in the table below:

TABLE 4: Characteristics and Bottlenecks of FRD Approaches (Vidanagama et al. 2020, p. 1348)

Table 9 Characteristics and Bottlenecks of approaches

Approach	Characteristics	Bottlenecks
Content-based approach	1. Depends on the content of the review	1. Consider the linguistic features only
	2. Review based features are used for feature extraction	2. Lack of large amount of labeled data for supervised learning
	3. Machine learning techniques are used for classification	
Behavioral-based approach	1. Depends on the behavior of reviewers	1. Consider the reviewer centric features only
	2. Reviewer profile and behavioral features are used for feature extraction	2. Lack of large amount of labeled data for supervised learning
Graph based approach	1. Analyse the relations among reviews, reviewers, and products through a heterogeneous graph (review graph) to identify review spammers	1. Ignore the textual information of the reviews
	2. Inter-dependent nature of data	2. Inter dependence of reviews, reviewers, and products need to be carefully accounted for spams
	3. Relational nature of problem domain	3. Definitions of spams are much more diverse
		4. Graph-based spam detection algorithms need to be designed not only for effectiveness but also for efficiency and scalability
Neural network approach	1. Better captures the contextual information and is ideal for realizing semantics of long texts	1. Required high computational power
	2. Continuous learning feature produces favorable detection rates	2. Required more data
	3. Require fewer input features	
Temporal and spatial approach	1. Discover unusual temporal patterns, because they allocate a large portion of reviews and can have effect on the product rating trend	1. Ignore the textual features of the reviews

The key points follow:

When attempting to detect fake reviews, one focuses on the content of reviews themselves, or on the reviewers; considering their behaviour; the times when and places where they post; or their connections to merchants and products.

Content analysis appears to work best on low quality reviews: reviews that are short, generic in description, plagiarised, riddled with errors or all of the above. Bad fake reviews are easy for human observers to identify, but automated content analysis proves useful if the reviews are of such a quantity that it is unfeasible for people to look through all of them themselves. The principle limitations of content analysis are that it is likely to miss well-written fake reviews and to ignore behavioural and other useful data that could improve accuracy.

Analysing the reviewers themselves can be very effective in identifying deceptive reviews, as the behaviour of users who post fake reviews is often observably different from that of users who post legitimate reviews. Users who post fake reviews tend to post far more reviews than bona fide users do. They register more accounts than ordinary users (Yu et al. 2019), post a relatively high number of positive reviews (Crawford et al. 2015), post more on weekdays and are typically concentrated in certain locations (Hajek et al. 2020). Many researchers find that methods of FRD that make use of these data are more accurate than those employing only content analysis (Mewada and Dewang 2021).

However, these methods are subject to the so-called cold-start problem: most fake reviews are posted by accounts created only to post a single review (Sihong Xie et al. 2012). Consequently, researchers often lack the types of data mentioned above, impeding the effectiveness of such approaches (Mewada and Dewang 2021).

4.2.2 Databases

A challenge common to all methods of FRD is the lack of verified databases. It is crucial to know which reviews are genuine and which are not, yet this is hard to establish with absolute certainty. Researchers have employed various methods to build their databases, with varying results (Mewada and Dewang 2021; Heydari et al. 2015). These approaches are listed in Table 5 and discussed below.

In their seminal paper, Liu and Jindal classified all duplicate reviews in their database as fakes. Other researchers have relied on experts, volunteers and freelancers to tag reviews they suspect of being fake, or have made use of algorithmic filters to do so.

Reasonable as these approaches might be, they do not establish credibility beyond any shadow of doubt.

Ott and co-authors attempted to solve this problem by commissioning their own fake reviews, employing freelancers from Amazon’s Mechanical Turk service to write deceptive reviews for them (2011). This way, they would be absolutely sure which of the reviews in their database were disingenuous.

The trouble is that, being written under very different circumstances, these fake reviews might be quite unlike real-world examples (Arjun Mukherjee et al. 2013) and models built on such databases are not likely to be more reliable than those built on others. Moreover, the distribution of fake reviews in an artificially produced review dataset might not resemble that of an ordinary review dataset, which could skew the results (Paul and Nikolaev 2021). Databases for which fake reviews were commissioned and paid for by the researchers are also relatively tiny (a few thousand versus a few million reviews: see Table 5), which might limit the validity of any results got from them.

TABLE 5: A List of Datasets Used in the FRD Literature (Mewada and Dewang 2021, p. 6)

Table 2
Different public reviews datasets used in literature.

Annotation method	Dataset name	Dataset description	Dataset fields
Rule-based	Amazon product review	5.8 M reviews and 2.15 M reviewers	Books, DVD, music, products (Jindal and Liu, 2008)
	Amazon book review	6.8 K reviews and 4.8 K reviewers	Books (Fornaciari and Poesio, 2014)
	TripAdvisor booking Review	2.8 K reviews	Hotels (Hammad, 2013)
Human-Based	Epinion	6 K reviews	Products (Li et al., 2011)
	TripAdvisor	3 K Reviews	Hotels (Ren et al., 2014)
Algorithm-Based Filtering	Yelp hotel	67.4 K reviews and 38 K reviewers	Hotels and Restaurants (Mukherjee et al., 2013)
	Yelp hotel	3.59 M reviews and 16 K reviewers	Hotels and Restaurants (Rayana and Akoglu, 2015)
	Yelp hotel	6.086 M reviews and 2.6 M reviewers	Hotels and Restaurants (Rayana and Akoglu, 2015)
	Dianping	10 K reviews and 9 K reviewers	Restaurants (Li and Chen, 2014)
Amazon Turker	TripAdvisor Hotel Review	800 reviews	Hotel (Ott et al., 1107)
	TripAdvisor Hotel Review	1600 reviews	Hotel (Ott et al., 2013)
	TripAdvisor multidomain	3032 reviews	Hotels, Restaurants, Doctors (Li et al., 2014)

In this paper, reviews were classified only as suspicious or credible, rather than fake or real, in order to emphasise that their veracity could not be proven, but rather only suggested plausibly by the models used to analyse them.

4.3 Web scraping and sampling

Web scraping refers to the automated collection of data from webpages using code or software. For this paper, over ten thousand reviews were gathered for analysis using a web scraping tool called Octoparse. It would have been ideal to collect every single wine review hosted on Jingdong, but this would have been unfeasible, as there appear to be nearly sixty nine million of them, and, at the rate Octoparse scrapes webpages (roughly five thousand a day), it would have taken nearly 13,800 days or thirty eight years to collect them.

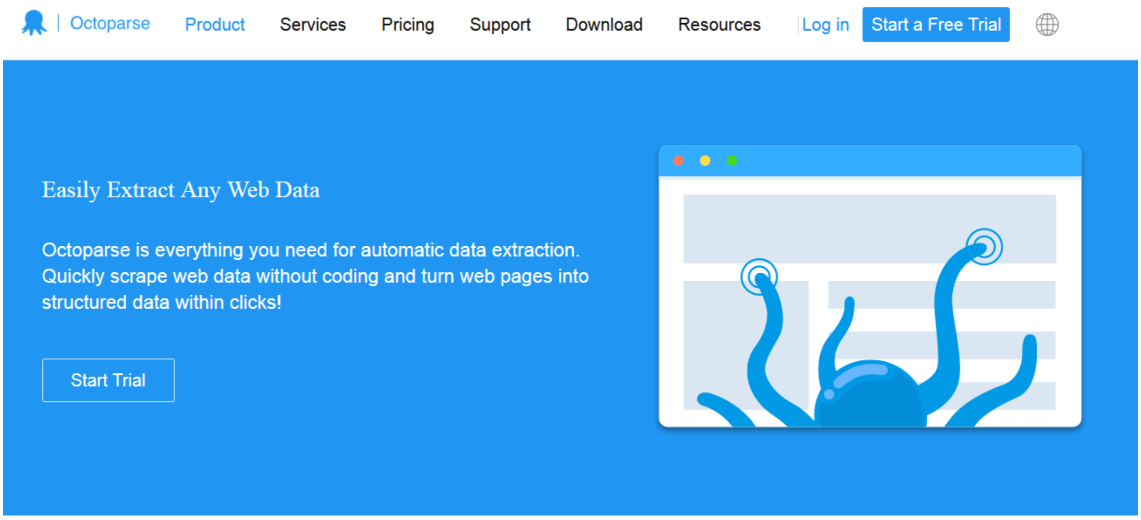

FIGURE 6: Octoparse Homepage (Web Scraping Services & Software | Octoparse 2022)

Ten thousand reviews were collected to strike a balance between practicality and having a large enough dataset to work with. Given that there are individual wines listed on Jingdong that have millions of reviews each, Octoparse was programmed to collect only the first fifty reviews of each wine. As new reviews are listed first, this meant that old reviews were largely neglected. This is likely to bias the sample, as reviews posted years apart might differ from one another, and compromise its statistical reliability. Indeed, many old reviews appeared more obviously fake than new ones. However, reviews that appear immediately during browsing are far more salient and likely to influence browsers' purchasing decisions, thus it seemed a legitimate compromise to make.

4.4 The data source: Jingdong

Jingdong, or JD.com is the second largest online retailer in China after Alibaba (Wang 2020b). It is also a major wine seller, with 35% of Chinese wine drinkers using it to buy imported wine (Lucy Jenkins 2016) and wine sales accounting for a quarter of Jingdong's total alcohol sales (Maxime Lu 2016). In 2015, revenue from Jingdong's wine sales amounted to RMB 500 million.[11]

[11] EUR 71,307,500

Given its importance in China's online wine trade, it makes sense to examine Jingdong's wine reviews when exploring the problem of fake reviews in Chinese e-commerce.

FIGURE 5: Jingdong Homepage (JD.com 2022)

Of crucial importance also is that it was possible to scrape reviews from Jingdong's website. By contrast, Taobao and other online Chinese retailers typically block traffic to their sites upon detecting that visitors are located outside of China. This makes scraping reviews from their websites impractical.

4.5 Machine learning and analysis

Machine learning is a type of programme that can learn and perform tasks autonomously (William L. Hosch 2022). It is efficient at dealing with large quantities of data and is used widely in fake review detection.

As described in section 3.3.1, fake reviews in Chinese E-commerce tend to be repetitive, vague, and generally easy to identify. It seemed reasonable, therefore, to develop a content-analysis model to process the dataset, especially since the sorts of metadata used in other FRD methods were lacking.

A random sample of 500 rows was taken from the main dataset and its contents were manually tagged as suspicious or credible, based on the heuristics and guidelines for identifying fake reviews detailed in section 4.2.1. 62 of the reviews were tagged as credible and 438 as suspicious. Given that tagging was done according to current best practice, it should be fairly valid, statistically speaking. However, it is not possible to confirm or quantify its accuracy as we simply do not know with certainty which reviews are authentic and which are not.

This sub-sample formed the training set that was used to develop text classification models. These models would be capable of classifying the contents of other datasets by identifying their similarities to the reviews tagged as credible or suspicious in the training set.

No-code machine learning platforms were used to develop the text classifiers. This was necessary as the author of this paper lacked the competence to programme such models himself. Models were created using Google Cloud's Vertex AI, AWS Sagemaker Canvas, Microsoft's Azure AutoML, Monkeylearn, Nyckel, PI.EXCHANGE's AI & Analytics, Akkio and Levity.

After trying all of these tools, Akkio and Levity were chosen, as they alone could process Chinese texts, produce functional models, and provide the capacity to process the full dataset without one having to pay for a full subscription. They are described in greater detail below.

Levity employs a particular form of deep neural network algorithm called transfer learning. Whereas traditional machine learning, and neural networks in particular, struggle to become accurate without being given copious training data, transfer learning means that algorithms are derived partly from models that have succeeded at similar tasks (Adrian Goergen 2022). In principle, this should mean that the training set of 500 reviews would be unlikely to hamper the accuracy of the model. However, accuracy can be improved when training data are balanced: a set of 49,000 images of cats and of only 1000 cats could in theory produce a model less accurate than one trained on a dataset of an equal number of the two classes (Levity 2022).

To ensure the best performance, two models were made with Levity. One was given the full training set of 500 classified reviews and another was given a set of 201 reviews, at a 1:2 ratio of credible to suspicious reviews. In theory, this would lead the model to identify credible reviews more accurately while still erring, if at all, on the side of tagging marginal reviews as suspicious. The assumption behind this plan is that fake reviews are preponderant and that it would be better to create a model that tags too many reviews as suspicious rather than too few. This would better protect consumers from deceptive reviewers.

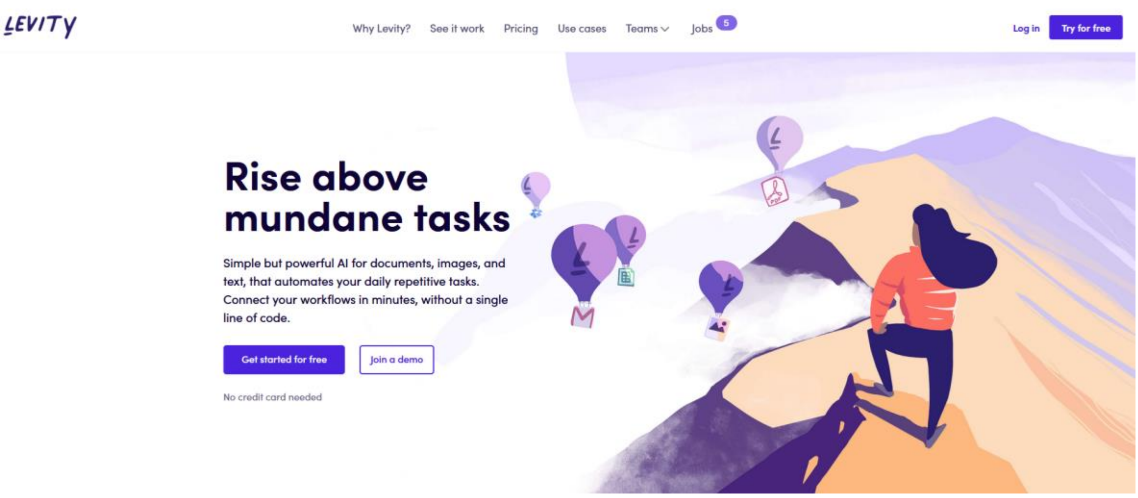

FIGURE 7: Levity Homepage (Levity | No-code AI workflow automation platform 2022)

Akkio functions differently. To create a model, it auditions and selects the top performer from a number of algorithms. It is not wedded to a single algorithm as Levity is to transfer learning. However, it is not clear which algorithm Akkio has ultimately chosen to employ, though it provides a performance report describing the model's effectiveness. This is hardly ideal for research and it hinders replication. This problem is common in machine learning, however, and especially with autoML: the model is often something of a black box and we can understand only partially what it is doing. The statistical validity of the model's results might then be compromised and to an extent that is difficult to quantify.

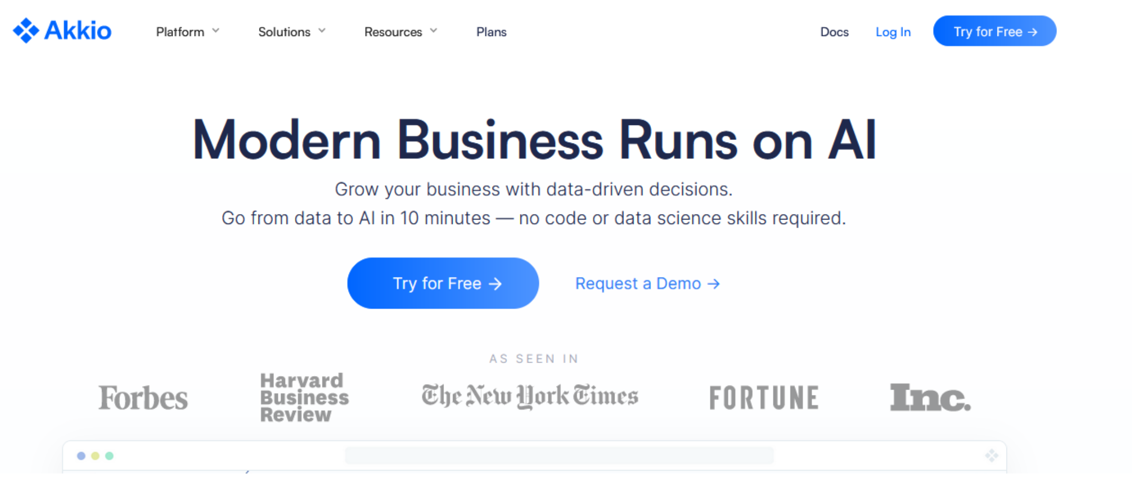

FIGURE 8: Akkio Homepage (Akkio 2022)

5 RESULTS

5.1 Model performance and results

The results of the different models are laid out in the table below.

TABLE 6. Results of Running the Dataset Through Different Classifiers

Model	Credible	Suspicious	Total	Reported Accuracy
Akkio	17	9870	9887[12]	88%
Levity 1[13]	24	10341	10365	73%
Levity 2[14]	1	10364	10365	33%

All of them classified the vast majority of reviews as suspicious. This appears to confirm the conventional wisdom that holds Chinese E-commerce to be flooded with deceptive reviews. However, the models were significantly more pessimistic than even the training set. This suggests that they might be prone to tagging rather too many reviews as suspicious than too few.

TABLE 7. Comparison of Models and Training Set

	Percentage of Reviews Tagged Suspicious
Training Set	86.6%
Akkio	99.8%
Levity 1	99.7%
Levity 2	99.9%

Notably, the Levity model produced with the balanced set of training data (Levity 2) classified almost every review in the dataset as fake. This seems extreme, and given that the model was evaluated by Levity as being only 33% accurate, perhaps this classifier is simply not very good.

---

[12] Note that the total number of reviews evaluated differs between Akkio and Levity, as the former removed a number of them when processing the dataset.
[13] Trained with the full 500 row training set.
[14] Trained with an adjusted training set of 201 items, at a 1:2 ratio of credible to suspicious reviews.

Notwithstanding Levity 2's poor performance, the Akkio model and the first Levity model were rated as being fairly accurate and produced results that were consistent and more realistic. They provide support for the hypothesis that most reviews in Chinese e-commerce are fake.

5.2 Term frequency and analysis

An online text analysis program called Voyant[15] was used to chart the frequency of terms appearing in the reviews. The point of this was to see whether there was a qualitative difference between reviews that were tagged suspicious and those that were tagged credible. The differences are striking.

[15] voyant-tools.org/

TABLE 8.1. Word Clouds Representing Term Frequency in Datasets

Model	Credible
Levity 1	
Akkio	

TABLE 8.2. Word Clouds Representing Term Frequency in Datasets

<table>
<tr><th>Model</th><th>Suspicious</th></tr>
<tr><td>Levity 1</td><td>
</td></tr>
<tr><td>Akkio</td><td>
</td></tr>
</table>

Among the set of reviews Levity 1 identified as credible, some of the most frequent terms are 可以 (fair, passable) 般般 (ordinary) 不 (not) 说实话 (speaking frankly) 有点 (a bit) 但 (but, however). This suggests that the model interprets ambivalent sentiments as reflecting sincerity.

In contrast, the reviews tagged by both Levity 1 and Akkio as suspicious are typically ebullient, with words like 好 (good), 不错 (not bad), 很 (very), 非常

(extremely) appearing frequently. The reviews tend to praise also the 包装 (packaging), 包装 (delivery), 速度 (speed), 服务 (service), and 京东 (Jingdong) itself. While it is possible that some reviewers really were taken with every aspect of their order, the consistency of and the overwhelmingly positive sentiments expressed in these reviews confirms to the stereotype of a fake online review and suggests the models might well be doing a good job of identifying deceptive reviews.

The term frequencies of the set of reviews Akkio tagged as credible are less illuminating than Levity 1’s. Notwithstanding the higher accuracy reported for the model, the reviews that it classified as credible resemble the suspicious reviews quite closely. This makes it difficult to see what differences it found between credible and suspicious reviews and casts doubt on its effectiveness.

To test the consistency of the suspicious reviews and make sure that a minority of very similar reviews were not distorting the results, a random sample of Levity 1’s suspicious reviews, equal in number to its credible reviews, were taken and run through Voyant as well. The results are highly consistent with those of the full set of suspicious reviews, suggesting that fake reviews are very homogeneous indeed.

TABLE 9. Term Frequency in Reviews Identified as Suspicious by Levity

Sample of Levity 1's Suspicious Reviews	
Total Corpus of Levity 1's Suspicious Reviews	

Tables for Levity's results, listing the ten most frequent terms per corpus, quantify these observations below.

TABLE 10.1. Term Frequency of Credible Reviews

Credible Reviews			
Term	Translation	Count	Frequency per Million
可以	All right	8	47058.824
一	One	6	35294.12
般般	Ordinary	6	35294.12
不	No, not	5	29411.766
口感	Mouthfeel	5	29411.766
酒	Alcohol	5	29411.766
味	Flavour	4	23529.412
有点	A bit	3	17647.06
果	Fruit	3	17647.06
一般	The same, in general	2	11764.706

TABLE 10.2. Term Frequency of the Suspicious Review Sample

Suspicious Reviews Sample			
Term	Translation	Count	Frequency per Million
很	Very	35	37353.254
好	Good	22	23479.19
酒	Alcohol	21	22411.953
喝	Drink	18	19210.246
口感	Mouthfeel	11	11739.595
不错	Not bad	10	10672.358
买	Buy	10	10672.358
一	One	9	9605.123
棒	Great	9	9605.123
速度	Speed	9	9605.123
不	No, not	8	8537.887

TABLE 10.3. Term Frequency of the Entire Suspicious Review Corpus

Suspicious Reviews Total			
Term	Translation	Count	Frequency per Million
很	Very	10579	29115.818
好	Good	7889	21712.326
喝	Drink	6316	17383.072
酒	Alcohol	6202	17069.318
口感	Mouthfeel	4856	13364.819
包装	Packaging	4431	12195.121
不错	Not bad	4410	12137.325
也	Also	3381	9305.282
非常	Extremely	3340	9192.441
买	Buy	3339	9189.688

6 DISCUSSION

While the results of this research provide support for the hypothesis that Chinese E-commerce is afflicted with large-scale review falsification, there are limitations to this research that one should be aware of.

6.1 Limitations and suggestions for future research

Despite the apparent suitability of using content analysis to identify fake reviews in this instance, the models would likely be improved were reviewers' behavioural, temporal and spatial data incorporated as well. However, this seems impossible to collect without being provided access to the backend of Jingdong's review platform, and it might be difficult to convince the company to grant outsiders access to this data for two reasons. Firstly, Jingdong has its own measures in place to counter fake reviews (Liao Jin 2017) and might see little reason to hand over operational data to outsiders who might in turn pass them on to competitors. Secondly, Jingdong might not be entirely willing to bring brushing to an end. Site representatives are reported to urge vendors to do whatever it takes to break sales records during shopping festivals, which the merchants interpret as a signal that the company is willing to tolerate fake reviews (Simin Zhang 2021). Jingdong and other e-commerce platforms profit from fake reviews and brushing generally in so far as they increase sales.

Relying on autoML platforms leaves one less room to optimise models than if one were programming them oneself and insight into how they work is typically limited. With greater technical capacity at researchers' disposal, these problems would be eliminated.

Sampling leads to sample bias. With better computational resources, it would be possible to extract the entire statistical population of reviews of interest, obviating the need to sample and the risk of sample bias.

Another benefit of working with complete or larger sets of reviews, and especially with larger training sets, would be the improvement of model accuracy. The models produced in the course of this research seemed to struggle to identify credible reviews. A training set that included a larger number of them would be likely to help.

Ideally, technically proficient researchers would apply state-of-the-art FRD to evaluate complete datasets of verified reviews. Future research would improve to the extent that it approaches this standard.

6.2 Managerial implications

What does this mean for companies interested in selling wine in China? What does it mean that e-commerce reviews in that country are widely regarded as fake and seem in fact to be so?

If consumers believe product reviews are written only to mislead them, then they will rely on them less when making purchases than they do on word of mouth, country of origin, branding, price and other more credible signs of quality.

Given the importance of reviews to succeeding in e-commerce, companies might be tempted to go with the flow and simply hire a brushing agency to generate a steady flow of fake positive reviews to their product pages. But apart from this being plainly unethical and illegal, it is unlikely to help a company and its products stand out from the crowd. Cultivating an air of authenticity could be far better for business.

Granting consumers first-hand experience of the company’s wares at tasting and educational events; providing wines to select on and off-trade partners, such as wine merchants, luxury resorts, wine bars, and upscale restaurants; association with credible influencers – these are forms of marketing that could prove far more effective in attracting customers and increasing sales. It could even motivate satisfied customers to post positive reviews online that others find believable.

That said, it might be wise to place less emphasis on e-commerce altogether. Meiselman (2020) reports that many companies that have been caught in the hype around Chinese e-commerce have entirely neglected building conventional sales channels. A company that focuses instead on developing relationships with ordinary retailers and distributors might gain an advantage over competitors who are concentrate their efforts only in e-commerce.

Online retail should not be entirely neglected, however. While many companies commission fake reviews for their own products, others employ them to tarnish the reputations of rivals, as in the case of Lady Penguin vs. Luoyin. This is a considerable business risk, and plans should be made in advance for how to deal with such an event before it happens.

7 CONCLUSION

The prevalence of fake reviews of wine sold on Jingdong was examined in this paper. The literature review testified to the growth and attractiveness of the Chinese wine market and to the appeal of selling wine online in China, but it also suggested that fake reviews were rife there, and that this poses a problem to companies that rely or intend to rely on on e-commerce to distribute their products in the country.

Ten thousand wine reviews were collected from JD.com, a major Chinese online retailer, and were analysed using models created with autoML to evaluate how many reviews in the sample, and e-commerce by implication, were fake. Though there appears to be room for improving these models, their results support the contention that the majority of Jingdong's wine reviews are fake. This has important implications for companies that operate in the Chinese wine industry or that wish to enter it.

Future research could expand on these implications, and investigate the extent to which deceptive reviews are common in other product categories and on other platforms in Chinese e-commerce.

8 Publication bibliography

A. Morrison (2014): Gradual Catch Up and Enduring Leadership in the Global Wine Industry. In *AAWE Working Paper*, pp. 1–34.

Adrian Goergen (2022): What is Transfer Learning and Why Does it Matter? Available online at https://levity.ai/blog/what-is-transfer-learning, updated on 9/16/2022, checked on 9/16/2022.

Agnoli, Lara; Capitello, Roberta; Begalli, Diego (2014): Geographical brand and country-of-origin effects in the Chinese wine import market. In *J Brand Manag* 21 (7-8), pp. 541–558. DOI: 10.1057/bm.2014.27.

Akkio (2022): Modern Business Runs on AI | No Code AI - Akkio. Available online at https://www.akkio.com/, updated on 9/17/2022, checked on 9/17/2022.

Amazon sues admins of 10K Facebook groups over fake reviews (2022). In *Associated Press*, 7/19/2022. Available online at https://apnews.com/article/technology-lawsuits-amazoncom-inc-03329beafb3b1eba170a13c99ee0d318, checked on 7/22/2022.

Arjun Mukherjee; Abhinav Kumar; Bing Liu; Junhui Wang; Meichun Hsu; Malu Castellanos; Riddhiman Ghosh (2013): Spotting opinion spammers using behavioral footprints.

Chu, Xiaoquan; Li, Yue; Xie, Yimeng; Tian, Dong; Mu, Weisong (2020): Regional difference analyzing and prediction model building for Chinese wine consumers' sensory preference. In *BFJ* 122 (8), pp. 2587–2602. DOI: 10.1108/BFJ-06-2019-0465.

CN111310064 Calculation method and device for eliminating comments of brushing comments by merchants and storage medium (2022). Available online at https://patentscope.wipo.int/search/en/detail.jsf?docId=CN297702382&_cid=P12-L5UVNH-37450-1, updated on 7/21/2022, checked on 7/21/2022.

Crawford, Michael; Khoshgoftaar, Taghi M.; Prusa, Joseph D.; Richter, Aaron N.; Al Najada, Hamzah (2015): Survey of review spam detection using machine learning techniques. In *Journal of Big Data* 2 (1). DOI: 10.1186/s40537-015-0029-9.

Eleanor Vaida Gerhards (2015): Your store is gross! How recent cases, the FTC, and state consumer protection laws can impact a franchise system's response to negative, defamatory, or fake online reviews. In *Franchise Law Journal* 34 (4).

Ellson, Andrew (2018): ‘A third of TripAdvisor reviews are fake’ as cheats buy five stars. In *The Times*, 9/22/2018. Available online at https://www.thetimes.co.uk/article/hotel-and-caf-cheats-are-caught-trying-to-buy-tripadvisor-stars-027fbcwc8, checked on 7/28/2022.

Fountain, J.; Zhu, M. (2016): Young Chinese Consumers’ Wine Socialization, Current Wine Behavior and Perceptions of Wine. In Roberta Capitello, Stephen Charters, David Menival (Eds.): The wine value chain in China. Global dynamics, marketing and communication in the contemporary Chinese wine market / Roberta Capitello ; edited by Steve Charters, David Menival. Oxford: Chandos Publishing, pp. 63–79.

Glazer, Jacob; Herrera, Helios; Perry, Motty (2021): Fake Reviews. In *The Economic Journal* 131 (636), pp. 1772–1787. DOI: 10.1093/ej/ueaa124.

Hajek, Petr; Barushka, Aliaksandr; Munk, Michal (2020): Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. In *Neural Comput & Applic* 32 (23), pp. 17259–17274. DOI: 10.1007/s00521-020-04757-2.

Han Dandong (2022): 记者卧底 10 多个网络水军群发现 网购一致好评可能是刷出来的_任务_评论_文案. With assistance of Wang Yitian (王意天). 法治日报. Available online at http://news.sohu.com/a/521530841_100001593, updated on 7/21/2022, checked on 7/21/2022.

Heydari, Atefeh; Tavakoli, Mohammad ali; Salim, Naomie; Heydari, Zahra (2015): Detection of review spam: A survey. In *Expert Systems with Applications* 42 (7), pp. 3634–3642. DOI: 10.1016/j.eswa.2014.12.029.

JD.com (2022). Available online at https://www.jd.com/?cu=true, updated on 9/20/2022, checked on 9/20/2022.

Levity (2022): How can I improve the performance of my machine learning model? Available online at https://help.levity.ai/how-can-i-improve-the-performance-of-machine-learning-models, updated on 9/11/2022, checked on 9/15/2022.

Levity | No-code AI workflow automation platform (2022). Available online at https://levity.ai/, updated on 9/16/2022, checked on 9/16/2022.

Li, H. (2018): Worlds of wine: Old, new and ancient. In *Wine Economics and Policy* 7 (1), 178–182.

Li, Jia-Gui; Jia, Jin-Rong; Taylor, Dennis; Bruwer, Johan; Li, Elton (2011): The wine drinking behaviour of young adults: an exploratory study in China. In *British Food Journal* 113 (10), pp. 1305–1317. DOI: 10.1108/00070701111177700.

Li, Yuanbo; Bardají, Isabel (2017): A new wine superpower? An analysis of the Chinese wine industry. In *Cah. Agric.* 26 (6), p. 65002. DOI: 10.1051/cagri/2017051.

Liao Jin (2017): 《反不正当竞争法》出台严惩刷单刷评价 京东对"刷单门"再回应 [Anti-Unfair Competition Law to Harshly Sanction Brushed Orders and Brushed Reviews: Jingdong Responds to Brushers Again. 中央新闻网站 [Central News Website]. Available online at https://news.ifeng.com/c/7faDhuYgrjd, updated on 7/20/2022, checked on 7/20/2022.

Lieber, Dave (2021): With so many fake reviews on Amazon, how do you know whom to trust before you buy? In *The Dallas Morning News*, 6/25/2021. Available online at https://www.dallasnews.com/news/watchdog/2021/06/25/with-so-many-fake-reviews-on-amazon-how-do-you-know-who-to-trust-before-you-buy/, checked on 7/22/2022.

Liu, Fang; Murphy, Jamie (2007): A qualitative study of Chinese wine consumption and purchasing. In *International Journal of Wine Business Research* 19 (2), pp. 98–113. DOI: 10.1108/17511060710758669.

Liu, Pan; Xu, Zhenning; Ai, Jun; Wang, Fei (2017): Identifying Indicators of Fake Reviews Based on Spammer's Behavior Features. In : 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C). 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C). Prague, Czech Republic, 2017-07-25 - 2017-07-29: IEEE, pp. 396–403.

Lucy Jenkins (2016): E-commerce driving China's wine sales - The Drinks Business. Available online at https://www.thedrinksbusiness.com/2016/08/e-commerce-driving-chinas-wine-sales/, updated on 7/23/2022, checked on 7/23/2022.

Maheshwari, Sapna (2019): When Is a Star Not Always a Star? When It’s an Online Review. In *The New York Times*, 11/28/2019. Available online at https://www.nytimes.com/2019/11/28/business/online-reviews-fake.html, checked on 7/22/2022.

Masson, Josselin; Sánchez, Carlos Raúl Sánchez; Celhay, Franck (2017): Is Mianzi the only Face of Chinese Consumers of Wine?: A Typology of Chinese Consumers of Imported Wine. In *International Journal of Market Research* 59 (5), pp. 625–654. DOI: 10.2501/IJMR-2017-044.

Maxime Lu (2016): JD.com: Mapping the landscape of online wine sales in China (Part I). Decanter China. Available online at https://www.decanterchina.com/en/knowledge/people/importers/jd-com-mapping-the-landscape-of-online-wine-sales-in-china-part-i, updated on 7/23/2022, checked on 7/23/2022.

Meiselman, Herbert L. (2020): Handbook of Eating and Drinking. Cham: Springer International Publishing.

Mewada, Arvind; Dewang, Rupesh Kumar (2021): Research on false review detection Methods: A state-of-the-art review. In *Journal of King Saud University - Computer and Information Sciences*. DOI: 10.1016/j.jksuci.2021.07.021.

Muhammad, Andrew; Countryman, Amanda M. (2019): In Vino ‘No’ Veritas: impacts of fraud on wine imports in China. In *Aust J Agric Resour Econ* 63 (4), pp. 742–758. DOI: 10.1111/1467-8489.12333.

Muhammad, Andrew; Leister, Amanda M.; McPhail, Lihong; Chen, Wei (2014): The evolution of foreign wine demand in China. In *Aust J Agric Resour Econ* 58 (3), pp. 392–408. DOI: 10.1111/1467-8489.12029.

OIV (2022): State of the World Vine and Wine Sector 2021.

P.O. Williamson; J. Robichaud; I.L. Francis (2012): Comparison of Chinese and Australian consumers' liking responses for red wines. Australian Journal of Grape and Wine Research (18).

Paul, Himangshu; Nikolaev, Alexander (2021): Fake review detection on online E-commerce platforms: a systematic literature review. In *Data Min Knowl Disc* 35 (5), pp. 1830–1881. DOI: 10.1007/s10618-021-00772-6.

Photograph of Wang Shenghan, Better Known as Lady Penguin (2022), updated on 7/27/2022, checked on 7/27/2022.

Plotkina, Daria; Munzel, Andreas; Pallud, Jessie (2020): Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews. In *Journal of Business Research* 109, pp. 511–523. DOI: 10.1016/j.jbusres.2018.12.009.

Qin Chen (2012): 京东现评价“专业户”:1 分钟评 5 款商品被疑刷评[Jingdong's Current 'Professional Account' Reviews: Five Product Reviews in One Minutes Leads to Brushing Suspicions]. 中国新闻网 [China News Network]. Available online at https://it.sohu.com/20120214/n334711131.shtml, updated on 7/11/2022, checked on 7/21/2022.

Qing, Ping; Xi, Aiqin; Hu, Wuyang (2015): Self-Consumption, Gifting, and Chinese Wine Consumers. In *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 63 (4), pp. 601–620. DOI: 10.1111/cjag.12086.

Shi, Fangfang; Ji, Shaojun; Weaver, David; Huang, Ming-Feng (2022): From curious to connoisseur: a longitudinal segmentation of attendees at a Chinese wine festival. In *IJCHM* 34 (3), pp. 885–907. DOI: 10.1108/IJCHM-03-2021-0331.

Sihong Xie; Guan Wang; Shuyang Lin; Philip S. Yu (2012): Proceedings of the 21st international conference companion on World Wide Web. New York, NY: ACM. Available online at http://dl.acm.org/citation.cfm?id=2187980.

Simin Zhang (2021): How to stay honest (or not) as an e-commerce merchant in China | KrASIA. Available online at https://kr-asia.com/how-to-stay-honest-or-not-as-an-e-commerce-merchant-in-china, updated on 7/2/2021, checked on 7/22/2022.

Somogyi, Simon; Li, Elton; Johnson, Trent; Bruwer, Johan; Bastian, Susan (2011): The underlying motivations of Chinese wine consumer behaviour. In *Asia Pacific Journal of Marketing and Logistics* 23 (4), pp. 473–485. DOI: 10.1108/13555851111165039.

Song Chao (2018): 公安部部署打击首起网络水军平台案：三打哈网被查封 (First Case of Ministry of Public Security Striking at a Water Army Platform: Sandaha Shut Down). 新京报 (The Beijing News). Available online at http://finance.sina.com.cn/chanjing/cyxw/2018-02-05/doc-ifyremfz4969720.shtml, updated on 7/22/2022, checked on 7/22/2022.

Sun, Benjamin (2015): China: Navigating the lucrative Chinese wine market. In *Wine & viticulture journal* 30 (2), pp. 68–70.

Thomas Hale (2019): Amazon and the problem of fake reviews. Financial Times. Youtube. Available online at https://www.youtube.com/watch?v=eDrK7qKFU1Y, updated on 8/2/2022, checked on 8/2/2022.

Vidanagama, Dushyanthi U.; Silva, Thushari P.; Karunananda, Asoka S. (2020): Deceptive consumer review detection: a survey. In *Artif Intell Rev* 53 (2), pp. 1323–1352. DOI: 10.1007/s10462-019-09697-5.

Wang, Natalie (2020a): China's own wine rating system adopted by JD.com. In *Vino Joy Limited*, 6/8/2020. Available online at https://vino-joy.com/2020/06/08/chinas-own-wine-rating-system-adopted-by-jd-com/, checked on 3/29/2022.

Wang, Natalie (2021): Lady Penguin appointed as 'Chinese Wine Ambassador'. In *Vino Joy Limited*, 4/15/2021. Available online at https://vino-joy.com/2021/04/15/lady-penguin-appointed-as-chinese-wine-ambassador/, checked on 7/21/2022.

Wang, Natalie (2022a): One in five Bordeaux wines destined to China in 2021. In *Vino Joy Limited*, 3/10/2022. Available online at https://vino-joy.com/2022/03/10/one-in-five-bordeaux-wines-destined-to-china-in-2021/, checked on 7/25/2022.

Wang, Natalie (2022b): China's biggest wine influencer Lady Penguin caught in smear campaign scandal. In *Vino Joy Limited*, 7/17/2022. Available online at https://vino-joy.com/2022/07/17/chinas-biggest-wine-influencer-lady-penguin-caught-in-smear-campaign-scandal/, checked on 7/21/2022.

Wang, Xiaodong (2020b): The Impact of E-Commerce Development on an Example of Jingdong Company. In *Russian and Chinese Studies* 4 (3), pp. 205–213. DOI: 10.17150/2587-7445.2020.4(3).205-213.

Web Scraping Services & Software | Octoparse (2022). Available online at https://www.octoparse.com/product, updated on 9/17/2022, checked on 9/17/2022.

William L. Hosch (2022): machine learning | artificial intelligence. Available online at https://www.britannica.com/technology/machine-learning, updated on 8/25/2022, checked on 9/23/2022.

Wu Qiuyu (2016): 淘宝销量评价被指随意买卖 守信卖家反被挤出市场 [Taobao Sales and Reviews Accused of Being Bought and Sold at One's Discretion: Honest Vendors Being Squeezed Out of the Market]. People's Daily. Available online at http://news.e23.cn/shehui/2016-07-29/2016072900107.html, updated on 7/20/2022, checked on 7/20/2022.

Xavier (2016): Chinese ECommmerce Leads a New Way for Wine Consumption. 中国电商引领葡萄酒消费新方式-红酒世界网. Available online at https://www.wine-world.com/culture/sc/20160608085626267, updated on 7/23/2022, checked on 7/23/2022.

Yang Jinzhi (2011): 网络水军牟利之道：分 25 级 接单每帖最低 7 毛 [Online Water Army's Way to Profit: Divided into 25 Levels and Each Commision Worth at Least 7 Bucks. Available online at https://news.qq.com/a/20110109/000825.htm, updated on 7/22/2022, checked on 7/22/2022.

Yu, Chuanming; Zuo, Yuheng; Feng, Bolin; An, Lu; Chen, Baiyun (2019): An individual-group-merchant relation model for identifying fake online reviews: an empirical study on a Chinese e-commerce platform. In *Inf Technol Manag* 20 (3), pp. 123–138. DOI: 10.1007/s10799-018-0288-1.

Yu, Ying; Sun, Huihui; Goodman, Steve; Chen, Shangwu; Ma, Huiqin (2009): Chinese choices: a survey of wine consumers in Beijing. In *International Journal of Wine Business Research* 21 (2), pp. 155–168. DOI: 10.1108/17511060910967999.

Zeng, L., & Szolnoki, G. (2017). Some fundamental facts about the wine market in China. In R. Capitello, S. Charters, D. Menival, & J. Yuan (Eds.), The wine value chain in China. Kidlington: Elsevier.