

Article

# Uncovering Key Factors That Drive the Impressions of Online Emerging Technology Narratives

Lowri Williams \* , Eirini Anthi  and Pete Burnap 

School of Computer Science & Informatics, Cardiff University, Cardiff CF24 4AG, UK; anthies@cardiff.ac.uk (E.A.); burnapp@cardiff.ac.uk (P.B.)

\* Correspondence: williams110@cardiff.ac.uk

**Abstract:** Social media platforms play a significant role in facilitating business decision making, especially in the context of emerging technologies. Such platforms offer a rich source of data from a global audience, which can provide organisations with insights into market trends, consumer behaviour, and attitudes towards specific technologies, as well as monitoring competitor activity. In the context of social media, such insights are conceptualised as immediate and real-time behavioural responses measured by likes, comments, and shares. To monitor such metrics, social media platforms have introduced tools that allow users to analyse and track the performance of their posts and understand their audience. However, the existing tools often overlook the impact of contextual features such as sentiment, URL inclusion, and specific word use. This paper presents a data-driven framework to identify and quantify the influence of such features on the visibility and impact of technology-related tweets. The quantitative analysis from statistical modelling reveals that certain content-based features, like the number of words and pronouns used, positively correlate with the impressions of tweets, with increases of up to 2.8%. Conversely, features such as the excessive use of hashtags, verbs, and complex sentences were found to decrease impressions significantly, with a notable reduction of 8.6% associated with tweets containing numerous trailing characters. Moreover, the study shows that tweets expressing negative sentiments tend to be more impressionable, likely due to a negativity bias that elicits stronger emotional responses and drives higher engagement and virality. Additionally, the sentiment associated with specific technologies also played a crucial role; positive sentiments linked to beneficial technologies like data science or machine learning significantly boosted impressions, while similar sentiments towards negatively viewed technologies like cyber threats reduced them. The inclusion of URLs in tweets also had a mixed impact on impressions—enhancing engagement for general technology topics, but reducing it for sensitive subjects due to potential concerns over link safety. These findings underscore the importance of a strategic approach to social media content creation, emphasising the need for businesses to align their communication strategies, such as responding to shifts in user behaviours, new demands, and emerging uncertainties, with dynamic user engagement patterns.

**Keywords:** emerging technologies; natural language processing; statistical modelling; social media data analysis; technology adoption



**Citation:** Williams, L.; Anthi, E.; Burnap, P. Uncovering Key Factors That Drive the Impressions of Online Emerging Technology Narratives. *Information* **2024**, *15*, 706. <https://doi.org/10.3390/info15110706>

Academic Editor: Paulo Quaresma

Received: 26 September 2024

Revised: 30 October 2024

Accepted: 4 November 2024

Published: 5 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the rapidly evolving domain of emerging technologies, businesses face the dual challenge of harnessing opportunities for growth and innovation while managing associated risks like compliance and cybersecurity. Effective decision-making in this context therefore requires a nuanced understanding of technological potentials and risks [1,2], often necessitating advanced analytical tools and expert insights.

Social media platforms, such as Facebook, Twitter (now re-branded as 'X'), Instagram, and LinkedIn, are increasingly seen not just as communications tools, but as barometers for public sentiment and opinion, influencing factors from brand perception to strategic

business decisions [3,4]. Such platforms offer a rich source of data from a global audience on their behaviours, preferences, and opinions, generating demand for advanced analytical strategies such as social media data analytics [5,6]. Such analytics can provide insights into market trends, consumer behaviour, and attitudes towards specific technologies, as well as monitoring competitor activity, including their adoption and application of technologies. As a result, the insights gathered can help businesses understand and make informed decisions about which technologies to invest in and how to market them, provide valuable insights into competitive strategies and potential market opportunities, as well as identify potential risks and respond swiftly to mitigate them [4,7–10].

The concept of online engagement encompasses a variety of dimensions and interpretations [11,12]. Johnston [13] defines online engagement as a dynamic process which captures the psychological and behavioural connections and interactions between individuals and organisations. In the context of social media, such engagements are conceptualised as immediate and real-time behavioural responses measured by likes, comments, and shares [14,15]. To monitor such engagements, social media platforms have now introduced new tools that allow users to analyse and track the performance of their posts and understand their audience. These analytics tools allow users to monitor key performance indicators such as the engagement, reach, and conversions of their narratives, which allow valuable insights into the success of their social media campaigns and strategies.

Recent studies have identified Twitter as an effective platform to support organisations in their business decision making processes. While these studies offer valuable insights, they often focus primarily on sentiment analysis to help assess consumer reactions and feedback (e.g., [16]) and interactive metrics, like retweets, which help capture information propagation, overlooking the broader impact of content exposure. Another metric, impressions, may be defined as the potential number of times a post is displayed on to a user, regardless of whether they interact with it. This exposure can significantly influence perceptions around topics of interest, even if users do not actively engage through retweets or likes.

Current analytics predominantly focus on measuring direct user interactions, such as likes and shares, which do not fully capture the breadth of influence that contextual content elements exert on public perception. This oversight can lead organisations to miss subtle yet powerful opportunities to shape market trends. This paper addresses this limitation by developing a comprehensive framework that not only analyses the features of high-impression tweets related to emerging technologies, but also how varying such features can attract or repel user attention, and how these dynamics shift over time. Powered by the automatic collection and analysis of social media discourse containing references to such technologies, this information may not only support business decision making, but also the strategic crafting of effective online narratives, which has the potential to play an important role in driving consumers towards engaging with such narratives and their referenced technologies beyond the social media space.

The main contributions of the work presented herein are as follows:

- A data-driven and scalable methodology for analysing the impact of nuanced features on the impressions of tweets concerning emerging technologies, and how the contribution of such factors change over time, providing a more detailed understanding of how specific content elements influence public perception.
- While existing analytics tools focus on quantitative metrics, such as likes, shares, and views, the framework presented herein incorporates the analysis of contextual features in posts, such as the number of words used and the sentiment expressed, as well as account-based attributes, such as the number of followers the publisher had at the time of posting the tweet. This allows for a more comprehensive understanding of the factors influencing the impressions of technology-related tweets.
- The insights derived from this study may not only support traditional business decision-making, but also offer strategies for effectively shaping online narratives. By identifying which aspects of content resonate most with audiences, organisations

can enhance their social media strategies to better align with user behaviours and emerging market trends. This capability is critical for fostering favourable perceptions of new technologies and facilitating their adoption.

Williams et al. [17] present a scalable and automated framework for tracking the likely adoption and/or rejection of new technologies from a large landscape of adopters. To support such experiments, textual data referencing emerging technology terms were collected from Twitter. Using this dataset to support the experiments in this paper, the study was designed as shown in Figure 1: (1) divide the texts based on their publication date; (2) for each dataset in (1), automatically extract the technology aspects from the text segments, as well as additional content-based features, such as the number of sentences and hashtags they include; (3) calculate the impressions of each text segments in the social media space; (4) apply statistical data modelling to identify which features positively and negatively contribute to the impressions of text segments; and (5) visualise and analyse the results.

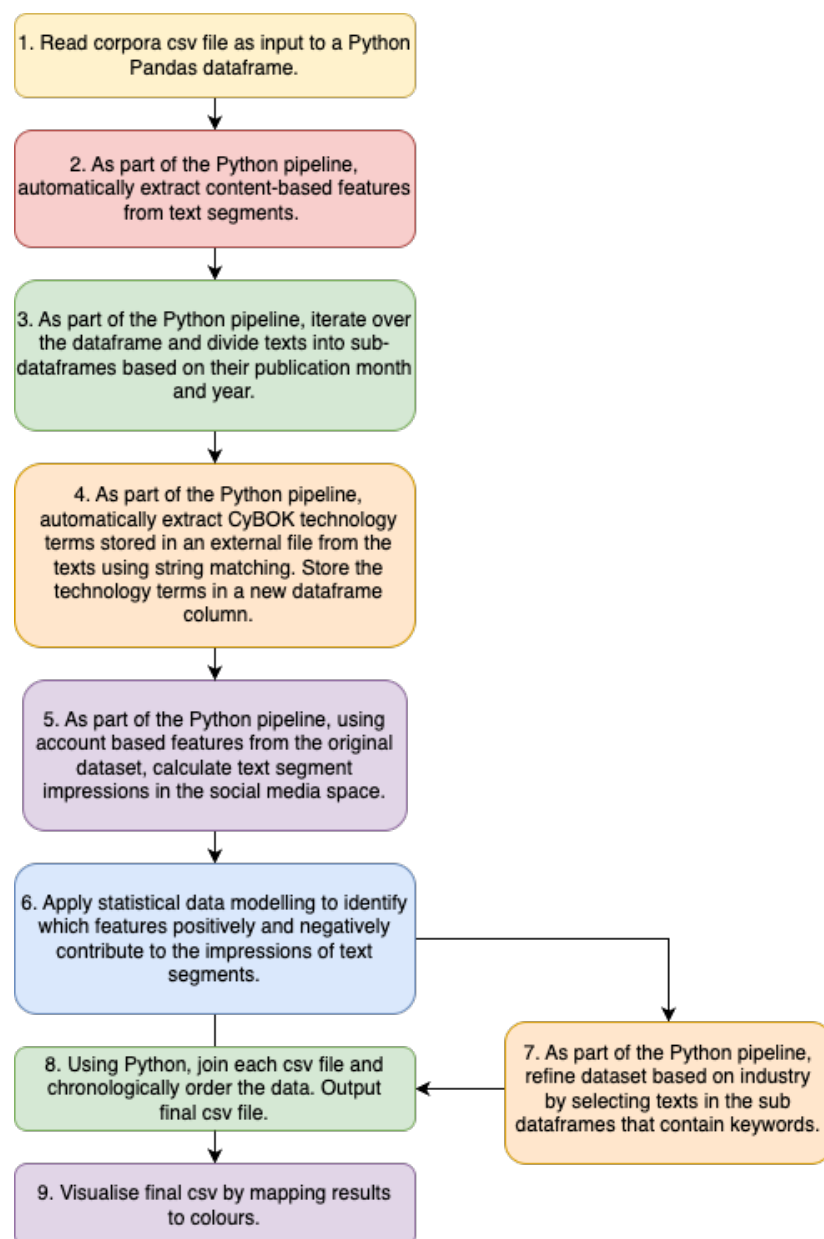


Figure 1. An overview of the study design.

The remainder of this paper is structured as follows: Section 2 presents the related work, Section 3 discusses the corpus of tweets used to support the experiments herein and the extraction of the independent textual features associated with tweets, as well as the features associated with the user account in which shared the text online (steps 1–3 in Figure 1), Section 4 discusses the calculation of a tweet’s impact in the social media space (point 5 in Figure 1), Section 5 discusses the statistical data model used to measure the contribution of the independent features towards the impact of tweets (point 6 in Figure 1), Section 6 presents and discusses the results following the analyses (generated from point 9 in Figure 1), Section 7 concludes the paper and, finally, Section 8 discusses future work.

## 2. Related Work

In recent years, research in the domain of social media analytics has gathered substantial interest. Numerous studies have explored the use of social media platforms, particularly Twitter, as tools for information dissemination, discerning consumer behaviour, market dynamics, and intelligence gathering [4,18–20] that can significantly aid in making a well-informed business decision. Studies range from focusing on identifying correlations between tweet frequency and stock market performance [21], predicting product sales [22], predicting Bitcoin prices [23], and analysing ways in which CEOs communicate via Twitter to help develop guidelines for effective tweeting strategies that can leverage the platform in leadership communication [24]. Such studies demonstrated how social media data can be utilised to assess public attitudes and forecast economic indicators, thereby providing businesses with valuable decision making information.

Previous studies also report that emotional messages are more effective than non-emotional messages, as emotions and sentiment influence the visibility and shareability of messages on Twitter, contributing to increased public attention and feedback [25]. With the rise of advanced analytics, there has been an increased emphasis on understanding the sentiment expressed in technology-related tweets. Many studies have demonstrated the value of sentiment analysis for businesses, such as predicting future stock market movements [26], understanding what investors think about a certain firm and, as a consequence, about the relative stock [27], assessing consumer reactions to product launches and/or adoption, such as ChatGPT [16], autonomous vehicles [28], open-source software [29], and Bitcoin [30,31], assessing consumer reactions to product features [32], such as smartphones and their applications, screens, cameras, etc., and assessing the impact and advancements in technology on employment [33].

Despite the acknowledged significance of social media analytics in the business decision making process, these works typically focus on attributes, such as the sentiment expressed, at a high level. As a response to such limitations, a selection of studies have investigated the influence of textual features on user engagement in social media posts. For example, Zhang et al. [15] understand how nonprofit organisations effectively engage with the public on social media by examining the effects of features such as the inclusion of URLs, hashtags, and mentions on the retweetability of tweets. Ji et al. [34] investigate how corporate Facebook posts’ functional traits, such as the number of likes and shares it has received, and its emotional traits (e.g., emotion presence, valence, and strength) impact public engagement online.

As a result, there is an opportunity to expand on existing research by creating a framework that allows for the examination of a greater range of features, their contribution to the impressions of emerging technology-related narratives in the social space, and how such contributions change over time. This study will not only add to the existing body of knowledge by broadening the scope of analysis and allowing for greater customisation, but it also provides valuable insights that allow for the strategic crafting of effective online narratives at specific times, subsequently optimising social media strategies for better communication and engagement with audiences. This approach fills a critical gap in the literature by integrating the analysis of content-level and account-level features, thereby providing more granular insights that can further support essential business decisions.

### 3. Text Corpus and Independent Features

To support the experiments presented in this paper, the dataset collected in [17] was used. The dataset consists of English tweets published between 1 January 2016 and 31 December 2021, containing the hashtags “IoT” or “Internet of Things” resulting in a dataset of 4,520,934 tweets. While hashtags can be a valuable tool for identifying relevant tweets, they are not without their limitations. By focusing solely on specific hashtags, such datasets may unintentionally exclude some of the broader conversations. This can occur because not all users consistently use hashtags, especially in informal or spontaneous conversations, users may employ different variations or misspellings of a hashtag, making it difficult to capture all relevant content, and popular hashtags can attract a large volume of irrelevant or spam content, making it challenging to filter out noise [35]. Despite these limitations, using hashtags remains a widely used method for collecting Twitter data (e.g., [15,17,28]), as it provides a structured way to identify and analyse specific topics of interest within the platform’s vast corpus of information, making them a valuable tool when exploring targeted conversations.

In [17], to analyse narratives surrounding specific emerging technology aspects, technology terms were automatically extracted from tweets using a direct string matching approach, where tweets were mapped against the Cybersecurity Body of Knowledge (CyBOK) [36], a resource which provides an index of cybersecurity reference terms. The final dataset consists of 514,459 tweets.

As described in Section 2, some works have conducted experiments that have looked at the contribution of different tweet features towards their impact in the social space. These studies have found that factors such as the content of the tweet, including the use of keywords and hashtags, the sentiment expressed, the timing of the tweet, the account posting the tweet, and the sentiment expressed in the tweet, all play a role in determining the virality of a tweet. In this case, Sections 3.1 and 3.2 discuss the extraction of content-based features and account-based features, respectively.

#### 3.1. Content Based Features

There is some evidence to suggest that expressing sentiment in a tweet can affect its retweetability (i.e., the physical act of sharing other users’ posts on Twitter). For example, Tsugawa and Ohsaki [37] found that tweets expressing negative sentiment are likely to be retweeted more rapidly and more frequently than positive and neutral ones. Additionally, Mahdikhani [38] show that tweets with higher emotional intensity are more popular than tweets containing information on the COVID-19 pandemic. Similarly, Javed et al. [39] report a relationship between negative emotions, such as fear, and malware propagation in the social space.

In [17], it is hypothesised that the expression of positive sentiment in tweets that referenced emerging technology infers an increase in the likelihood of impacting a technology user’s acceptance to adopt, integrate, and/or use the technology, and negative sentiment infers an increase in the likelihood of impacting the rejection of emerging technologies by adopters. In this case, the dataset in [17] includes the sentiment class (positive, negative, and neutral) extracted from such tweets, and will be used in this study to measure the contribution of the sentiment expressed towards a tweet’s impression.

Online user writing styles can vary greatly depending on a variety of factors, such as the platform they are using (e.g., social media, forums, blogs), the purpose of their writing (e.g., personal communication, marketing, journalism), the audience they are writing for (e.g., friends, strangers, customers), their education, gender, and vocabulary [40]. Some users may adopt a more casual and informal writing style, using abbreviations, slang, and emoticons, while others may use more formal language and grammar. In addition, users may employ different writing styles depending on the topic or tone of their writing, such as using a more serious or humorous tone.

To capture different writing styles, the contextual features, i.e., the context in which words and phrases are used, otherwise known as Part of Speech (POS), such as the sur-



rounding words, sentence structure, and semantic relationships used in tweets were extracted using Python’s natural language package, Natural Language Toolkit (NLTK) [41] (version 3.4.1).

In addition to POS features, the number of words (tokens), sentences, exclamation marks, alphanumeric data, capitalisation, and word extensions were extracted using Python’s RegEx (version 2020.9.27). The remaining features, such as the number of URLs, hashtags, and user mentions in a tweet, as well as the number of times a tweet is retweeted, were collected as part of the data collection described in Section 3. Table 1 describes the features extracted from tweets.

**Table 1.** Tweet content-based features.

Content-Based Feature	Description
Adjectives (ADJ)	Number of adjectives in a tweet (e.g., new, good, high, special, big, local)
Adposition (ADP)	Number of adpositions in a tweet (e.g., on, of, at, with, by, into, under)
Adverbs (ADV)	Number of adverbs in a tweet (e.g., really, already, still, early, now)
Conjunctions (CONJ)	Number of conjunctions in a tweet (e.g., and, or, but, if, while, although)
Determiner (DET)	Number of determiners in a tweet (e.g., the, a, some, most, every, no, which)
Nouns (NOUN)	Number of nouns in a tweet (e.g., year, home, costs, time, Africa)
Particle (PRT)	Number of particles in a tweet (e.g., at, on, out, over per, that, up, with)
Pronouns (PRON)	Number of pronouns in a tweet (e.g., he, their, her, its, my, I, us)
Verbs (VERB)	Number of verbs in a tweet (e.g., is, say, told, given, playing, would)
Numeral (NUM)	Number of numbers in a tweet (e.g., twenty-four, fourth, 1991, 14:24)
URL Count	Number of URLs in a tweet
Token Count	Number of words in a tweet
Hashtag Count	Number of hash-tagged words in a tweet (e.g., #iot, #hardware, #5g)
Retweet Count	Number of times a tweet has been retweeted
User Mentions	Number of users mentioned in a tweet
Sentences Count	Number of sentences in a tweet
Exclamation Count	Number of exclamation marks (!) used in a tweet
Alphanumeric Count	Number of numerical digits (0–9) in a tweet
Capitalisation Count	Number of capitalised characters in a tweet
Word Extensions Count	Number of trailing characters in a word (e.g., “lolllll”)

### 3.2. Account Based Features

Studies also suggest that features associated with the account from which a tweet is posted, such as the number of followers the account has or the number of times a tweet has been favoured, affect a tweet’s virality. For example, studies have shown that tweets from

users with more followers tend to be more likely to be retweeted than tweets from users with fewer followers [42,43]. This is intuitive, as tweets from users with more followers are likely to be seen by more people, and thus have a greater potential to be shared.

In this case, in the analysis herein, account-based features are also considered. As part of the data collection process in [17], features associated with the account which shared the tweet containing references to emerging technologies were also extracted. Table 2 describes the six account-based features used in the study herein.

**Table 2.** Twitter account-based features.

Account-Based Feature	Description
Listed Count	The number of public lists that the user is a member of
Friends Count	The number of users the account is following
Favourite Count	Indicates approximately how many times a tweet has been liked by Twitter users
Statuses Count	The number of tweets (including retweets) issued by the user
Followers Count	The number of followers the user account currently has
Favourites Count	The number of tweets a user has liked in the account’s lifetime

To summarise, Table 3 describes the mean, standard deviation, minimum, and maximum values for each of the aforementioned features extracted from the dataset.

**Table 3.** Description of the dataset.

Feature	Mean	Std. Dev.	Min	Max
Content-Based Features				
ADJ	2.39	1.83	0	32
ADP	1.76	1.51	0	12
ADV	0.59	0.91	0	10
CONJ	0.61	1.02	0	28
DET	1.03	1.25	0	13
NOUN	13.56	5.86	1	125
PRT	0.54	0.79	0	8
PRON	0.52	0.96	0	15
VERB	2.42	2.18	0	21
NUM	0.67	1.34	0	22
URL Count	1.05	0.46	0	4
Token Count	21.83	9.13	1	89
Hashtag Count	4.71	4.26	1	43
Retweet Count	2.23	15.94	0	6777
User Mentions	0.31	0.69	0	12
Sentences Count	2.92	1.08	1	21
Exclamation Count	0.10	0.37	0	16
Alphanumeric Count	1.89	1.49	0	25
Capitalisation Count	18.67	10.03	0	227
Word Extensions Count	0.01	0.13	0	6

Table 3. Cont.

Feature	Mean	Std. Dev.	Min	Max
Neutral	0.37	0.48	0	1
Negative	0.04	0.13	0	0.98
Positive	0.25	0.29	0	0.99
Account-Based Features				
Listed Count	591.52	1498.79	0	26,683
Friends Count	5720.71	24,126.94	0	795,309
Favourite Count	2.61	57.71	0	31,181
Statuses Count	58,613.40	149,062.75	0	2,504,864
Followers Count	17,166.26	89,689.98	0	11,957,332
Favourites Count	13,421.51	50,594.47	0	1,406,517

#### 4. Dependent Variable

As noted by Saxton and Waters [14], in the context of social media, online engagements are conceptualised as immediate and real-time behavioural responses measured by likes, comments, shares, etc. The impact of such narratives, however, refers to the influence, reach, and effect such posts have on its audience [44]. In the context of Twitter, the impact of a tweet can be measured in various ways and is not limited to one metric. The following provides descriptions of the metrics used to measure tweet impact:

- **Reach:** The potential number of unique users who saw the tweet. It can be estimated using tools like Twitter Analytics, third-party analytics platforms, or social media management tools.
- **Engagement:** The number of likes, retweets, comments, and shares a tweet receives. This can be a good indicator of the level of interest and involvement of the audience with the tweet's content.
- **Impressions:** The potential number of times a tweet is displayed on a user's timeline, regardless of whether they engage with it. This can provide a good estimate of the total visibility of a tweet.
- **Click-Through Rate (CTR):** Measures the number of clicks a tweet receives as a proportion of the number of impressions it receives. This can give an idea of the effectiveness of the call to action in the tweet.
- **Audience demographics:** Demographic data such as gender, age, location, and interests can provide insights into the audience that is most engaged with a tweet. This information can help tailor future tweets to better target the desired audience.
- **Hashtag performance:** The use of hashtags in tweets can increase their visibility and reach. Measuring the performance of specific hashtags can give insights into the topics and conversations that are resonating with the audience.

While retweetability, or the likelihood of a tweet being shared, is often used as a measure of a tweet's impact, it does not necessarily provide a comprehensive view of a tweet's overall reach or influence. The act of retweeting signifies a certain level of engagement, indicating that the content resonated with a user to the point where they felt compelled to share it with their followers. However, not every user who sees or is influenced by a tweet will necessarily retweet it. Some users may not be active sharers, or they might consume the content without feeling the need to broadcast it further. Moreover, a tweet can have a significant impact by appearing in users' feeds, being included in search results, or being read directly on the tweet author's profile—these impressions all contribute to the tweet's overall impact, but they would not be reflected in the retweet count. For example, a tweet about a new technology may not receive many likes or retweets, but could still contribute to users' knowledge of the topic by being displayed on their timelines. Thus,



impressions help capture both active engagement and passive reach, providing a more comprehensive picture of a tweet’s impact. Furthermore, impressions can highlight the true potential reach of content, especially when users with a large following tweet about the same topic multiple times. This allows the measure of not just the direct effects of user interaction, but also the potential scope of influence within technology-driven discussions.

In this case, in the experiments herein, the contribution of features towards the impact of a tweet is measured using their impressions. Twitter’s developer portal (<https://developer.twitter.com>, accessed on 5 July 2024) allows users to register for an account to retrieve a set of keys to use alongside the API to access Twitter data. Given that the account used to collect data in [17] was under the academic category, and not an enterprise account, the impression metric was not retrievable. Therefore, in this paper, the impression score of each tweet was calculated by multiplying the number of followers of the user who had posted the tweet by the number of times the user had tweeted about each technology. For example, a user with 2811 followers tweeted twice about ‘drones’; therefore, the impression score was calculated as 5622. When more than one technology term was referenced, the impression score of the tweet was calculated as the sum of the impression scores for each technology referenced. For example, the aforementioned user tweeted about the ‘cyber attack’ launched on ‘drones’. Given that the impression score for ‘drones’ is 5622, and as they had only ever tweeted about the ‘cyber attack’ once, the overall impression score for the given tweet was calculated as 8433. Table 4 describes the statistics surrounding the dependent variable.

**Table 4.** Description of the dependent variable, where N defines number of instances.

Mean	Std. Dev.	Min Score	Max Score	N = 0	N > 0
102,281.68	733,036.30	1	87,496,227	6386	500,364

## 5. Model Selection

In the dynamic field of social media analytics, the transformation of raw count data into actionable insights is pivotal for guiding strategic decision-making in technology, marketing, and consumer engagement. However, there are inherent complexities in social media data, which makes choosing a statistical model for this task challenging. Such complexities include rapid fluctuations in how much and how quickly content spreads across social media [45]. That is, some posts may go viral and receive a high level of engagement (likes, shares, comments, etc.), while others may see very little activity. Therefore, the effective selection of a statistical model that can handle the unique properties of social media data and the frequent absence of zero counts is important.

In this paper, and as shown in Table 4, the dependent variable, which represents the number of impressions of a tweet, is a positive non-zero number; that is, 500,364 tweets have an impression score greater than 0. Count data, by definition, represents the number of occurrences of an event within a fixed period or space. Therefore, standard linear regression models are not appropriate for such data, due to the violation of the assumption of normality. Instead, models specifically designed for count data, such as the Poisson and Negative Binomial models, are more suitable.

The Poisson model is often the starting point for count data analysis due to its simplicity and underlying assumption that the mean and variance of the count data are equal. However, real-world data often violates this assumption, exhibiting over-dispersion, where the variance exceeds the mean. Over-dispersion can lead to underestimating the standard errors in the Poisson model, resulting in overly optimistic *p*-values and confidence intervals [46]. On inspecting Table 4, it can be observed that the mean and variance of the impression score are 102,281.68 and 733,036.30, respectively, indicating over-dispersion and rendering the Poisson model inappropriate for this analysis.

Given the presence of over-dispersion, the Negative Binomial model, which is a generalisation of the Poisson model, becomes a more appropriate choice. The Negative Binomial

model introduces an additional parameter to account for the over-dispersion by allowing the variance to exceed the mean, making it more flexible and better suited for data where the variance is not constant [47]. The dataset in this study consists only of tweets that have received at least one impression, meaning there are no zero counts in the dependent variable. This zero-truncation characteristic further complicates the use of standard count models, as they typically assume the possibility of zero occurrences. The presence of zero-truncation violates this assumption, potentially biasing the results if not properly addressed.

To handle this, the Zero-Truncated Negative Binomial (ZTNB) model is an off-the-shelf model specifically designed to account for the absence of zeros in the data. This model adjusts the likelihood function to consider only positive counts, providing a more accurate estimation of the relationship between the independent variables and the count of impressions. This model has also been notably applied in similar studies, such as by Javed et al. [39], whose work focuses on investigating the features that influence the retweetability of tweets containing malware.

$$Pr(y_i|y_i > 0) = \frac{(\Gamma(y_i + \alpha^{-1})/y_i!\Gamma(\alpha^{-1}))(\alpha^{-1}/(\alpha^{-1} + \mu_i))^{\alpha^{-1}}(\mu_i/(\alpha^{-1} + \mu_i))^{y_i}}{1 - (1 + \alpha\mu_i)^{\alpha^{-1}}} \tag{1}$$

$$E(y_i|y_i > 0) = \frac{\mu_i}{Pr(y_i > 0)} = \frac{\mu_i}{1 - (1 + \alpha\mu_i)^{\alpha^{-1}}} \tag{2}$$

$$Var(y_i|y_i > 0) = \frac{E(y_i|y_i > 0)}{Pr(y_i > 0)^\alpha} [1 - Pr(y_i = 0)^{\alpha+1} E(y_i|y_i > 0)] \tag{3}$$

$$L = \prod_{i=1}^N Pr(y_i|y_i > 0) = \prod_{i=1}^N \frac{(\Gamma(y_i + \alpha^{-1})/y_i!\Gamma(\alpha^{-1}))(\alpha^{-1}/(\alpha^{-1} + \mu_i))^{\alpha^{-1}}(\mu_i/(\alpha^{-1} + \mu_i))^{y_i}}{1 - (1 + \alpha\mu_i)^{\alpha^{-1}}} \tag{4}$$

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \tag{5}$$

In a ZTNB model, the probability of the dependent variable size is based on positive count data of the independent variables using Bayes' Theorem [47–49]. Equations (1)–(5) report how the probability mass function, mean, variance, likelihood function, and response surface of a ZTNB model are calculated, respectively, where  $Pr(y_i|y_i > 0)$  is the probability mass function of the ZTNB distribution,  $E(y_i|y_i > 0)$  is the expectation of ZTNB distribution,  $Var(y_i|y_i > 0)$  is the variance of ZTNB distribution,  $\alpha$  is the over-dispersion parameter,  $L$  is the likelihood function,  $\mu_i$  is the estimated impression count for the  $i$ th observation,  $y_i$  is the observed impression count for the  $i$ th observation,  $k$  is the parameter coefficient of the  $k$ th predictor variable ( $k = 0$  for intercept), and  $X_{ki}$  is the value of the  $k$ th predictor variable (independent variables) for the  $i$ th observation [39].

## 6. Results and Discussion

The Incident Rate Ratio (IRR) was used to evaluate the results of each analysis, rather than the coefficients of predictor variables. IRR values are calculated by exponentiating the ZTNB regression coefficients, allowing the strength of the relationship between the independent variables discussed in Section 3 and the dependent variable to be determined.

The values reported in the IRR column are expressed as a percentage change in the impression of a tweet and consider that all other factors are held constant. An  $IRR > 1$  indicates an increase in the likelihood that a tweet will be impressionable. The percentage change in the IRR is calculated as  $(IRR - 1) \times 100$ . Therefore, when  $IRR = 1.096948$ , and when all other factors are constant, tweets containing a higher use of that particular independent variable are 9.7% more likely to have greater impressions.

Likewise,  $IRR < 1$  indicates a decrease in the likelihood that a tweet will be impressionable. The percentage change in the IRR is calculated as  $(1 - IRR) \times 100$ . Therefore, when  $IRR = 0.9686518$ , and when all other factors are constant, tweets containing a greater use of that particular independent variable are 3.1% less likely to have impressions.

The IRR percentage increase and decrease is capped at 100% when it is greater than 100%, as it ensures that the value remains within a realistic range and provides a meaningful measure of the relationship between the independent and dependent variable.

Table 5 reports ZTNB results across the dataset. For impressions, 14 features derived from the tweets were reported as being statistically significantly associated with the dependent variable size. Results demonstrate that increasing the number of content-based features used in a tweet, particularly the number of words (IRR = 1.019128,  $z = 8.71$ ,  $p < 0.05$ ), the number of pronouns (IRR = 1.013692,  $z = 2.91$ ,  $p < 0.05$ ), and the number of URLs (IRR = 1.10283,  $z = 3.4$ ,  $p < 0.05$ ) positively contribute to a tweet’s impression by 1.9%, 1.4%, and 2.8%, respectively. However, the most notable feature to positively affect a tweet’s impression is when a negative sentiment is expressed. With IRR = 1.096948,  $z = 2.83$ ,  $p < 0.05$ , a tweet expressing negative sentiment is more likely to be more impressionable by 9.7%. This may be explained, due to negativity bias, by the fact that negative news is considered as being more likely to elicit an emotional response [50], which may increase the likelihood to spark a debate or conversation, which can also increase user engagement and tweet virality.

**Table 5.** ZTNB results across the whole dataset for the dependent variable.

Feature	IRR	Std. Err.	z	p z	Percent
Content Based Features					
ADJ	0.9959427	0.0029206	−1.39	0.166	0.40573
ADP	0.9686518	0.0035066	−8.8	0	3.13482
ADV	1.008247	0.0046779	1.77	0.077	0.8247
CONJ	0.9877936	0.0035892	−3.38	0.001	1.22064
DET	0.9731641	0.0039463	−6.71	0	2.68359
NOUN	1.000149	0.0020763	0.07	0.943	0.0149
PRT	0.9937308	0.0052185	−1.2	0.231	0.62692
PRON	1.013692	0.0047367	2.91	0.004	1.3692
VERB	0.9929429	0.003038	−2.31	0.021	0.70571
NUM	0.9902476	0.0053631	−1.81	0.07	0.97524
URL Count	1.0283	0.0084348	3.4	0.001	2.83
Token Count	1.019128	0.0022178	8.71	0	1.9128
Hashtag Count	0.9968387	0.0011928	−2.65	0.008	0.31613
Retweet Count	0.9993111	0.000252	−2.73	0.006	0.06889
User Mentions	0.9869311	0.0052843	−2.46	0.014	1.30689
Sentences Count	0.9888629	0.0037508	−2.95	0.003	1.11371
Exclamation Count	0.997554	0.0092245	−0.26	0.791	0.2446
Alphanumeric Count	1.008149	0.004226	1.94	0.053	0.8149
Capitalisation Count	0.9996738	0.0005604	−0.58	0.561	0.03262
Word Extensions Count	0.9144226	0.0231311	−3.54	0	8.55774
Neutral	0.9963644	0.0112812	−0.32	0.748	0.36356
Negative	1.096948	0.0358416	2.83	0.005	9.6948
Positive	0.9769754	0.0194333	−1.17	0.242	2.30246

Table 5. Cont.

Feature	IRR	Std. Err.	z	p z	Percent
Account Based Features					
Listed Count	0.9999951	0.00000453	−1.09	0.274	0.00049
Friends Count	1	0.00000201	0.66	0.512	0
Favourite Count	0.9999929	0.0001268	−0.06	0.955	0.00071
Statuses Count	1	0.000000284	−1.69	0.092	0
Followers Count	1	0.000000082	1.91	0.056	0
Favourites Count	1	0.0000000993	4.39	0	0

Conversely, 10 of the 12 features demonstrated to negatively affect the impression of a tweet. When a tweet was retweeted and included more hashtags, verbs, sentences, conjunctions, user mentions, determiners, and adpositions, the less likely the tweet was to be impressionable. The most notable feature to negatively contributes to a tweet’s impression is when it contains a higher number of words with trailing characters (word extensions). It was reported, by increasing the use of this feature in the text, tweets are less likely to be impressionable by 8.6% (IRR = 0.9144226,  $z = -3.54$ ,  $p < 0.05$ ). It was also observed that the account-based feature, favourites count, was statistically significant; however, it did not affect the impression of a tweet (IRR = 1,  $z = 4.39$ ,  $p < 0.05$ ).

Balanced datasets are useful for statistical analysis because they ensure that each class or group is equally represented. In the case of social media data analysis, however, balancing the dataset may not be essential or even desirable because such interactions are inherently imbalanced, with some topics or events generating far more discussion than others. Attempting to balance the dataset by including an equal amount of tweets across all themes or events would not adequately reflect the natural distribution of social media conversation [51]. When assessing the data, it is critical to account for the inherent imbalance in the dataset by employing proper statistical approaches that can handle unbalanced data, such as ZTNB [52,53]. Figure 2 reports the number of tweets across an excerpt of emerging technologies used for further analysis herein. With 38,455 and 1122 tweets, ‘software’ and ‘5G network’ have the most and least tweets in the dataset, respectively.

Figures 3–7 present IRR percentage results following analyses using the ZTNB method. In these figures, the statistically significant IRR percentage decrease in a tweet’s impression is indicated in green, with darker shades of green representing a higher contribution and lighter shades indicating a lower contribution. Conversely, to easily compare the magnitude between the contribution of features towards a tweet’s impression, the IRR percentage decrease that is statistically significant is shown as negative values and is shown in red, with darker shades of red representing a greater negative contribution and lighter shades indicating a less negative contribution. The legend indicates the percentage increase, which ranges between 0 and 100, and percentage decrease, which ranges between 0 and −100.

As discussed in Section 3, in [17], technology terms were automatically extracted from tweets using a direct string matching approach, where tweets were mapped against CyBOK [36] (e.g., ‘Machine Learning’, ‘5G Network’, and ‘WiFi’). In this case, in this paper, analysing such datasets allows a further understanding on the contribution of features towards the impression of tweets referencing specific technologies. Figure 3 reports ZTNB results across technology specific datasets, where a model was built to analyse the contribution of features towards the impression of tweets referencing an excerpt of 40 terms relating to emerging technologies. Similar observations to the aforementioned results in Table 5 can be found, where for a selection of emerging technologies, such as ‘Android’ and ‘autonomous vehicles’, the high usage of content based features, such as the number of word extensions and URLs, positively contribute to a tweet’s impression by 100%. Likewise, for other technologies, such as ‘cryptocurrency’ and ‘Linux’, the use of such features may negatively or have no contribution towards a tweet’s impression.

A further interesting observation from Figure 3 is the contributions of expressing positive and negative sentiment towards a tweet’s impression. Expressing a positive sentiment alongside technologies (e.g., ‘data science’, ‘deep learning’, and ‘machine learning’) which often can positively impact, improve, and streamline various industries and daily tasks demonstrates to increase a tweet’s impression in the social space by 100%. Conversely, when negative sentiment is expressed towards technology-related terms, such as ‘hacker’, the tweet’s impression is significantly increased by 100%, as they may be more in line with the perceived severity of the issue.

However, positively charged tweets referencing technology-related terms which are often viewed negatively (e.g., ‘attack, botnet, malware’ and ‘threat’) were reported to significantly decrease the tweet’s impression. This is intuitive, as such terms are related to cybersecurity attacks and threats that are often correlated with serious consequences for individuals and organisations. Therefore, expressing positive sentiment alongside them may be perceived as inappropriate or not in line with the gravity of the situation. Expressing negative sentiment alongside consumable technologies such as ‘hardware’, ‘smartphones’, and ‘iOS’ also positively contribute to the tweet’s impression as they may be more in line with common complaints or issues that users may have during events such as new updates or product releases, and may be seen as more relatable or relevant to users.

To gain a deeper understanding of what features have contributed to the impressions of tweets referencing specific technology terms and changes in such metrics over time, the original dataset can be refined to datasets of tweets which focus on individual technologies and the specific month and year across the five year period in which they were posted online. A ZTNB model can then be applied to each refined dataset allowing a chronological analysis of the contribution of features to the impressions of tweets in the social space.

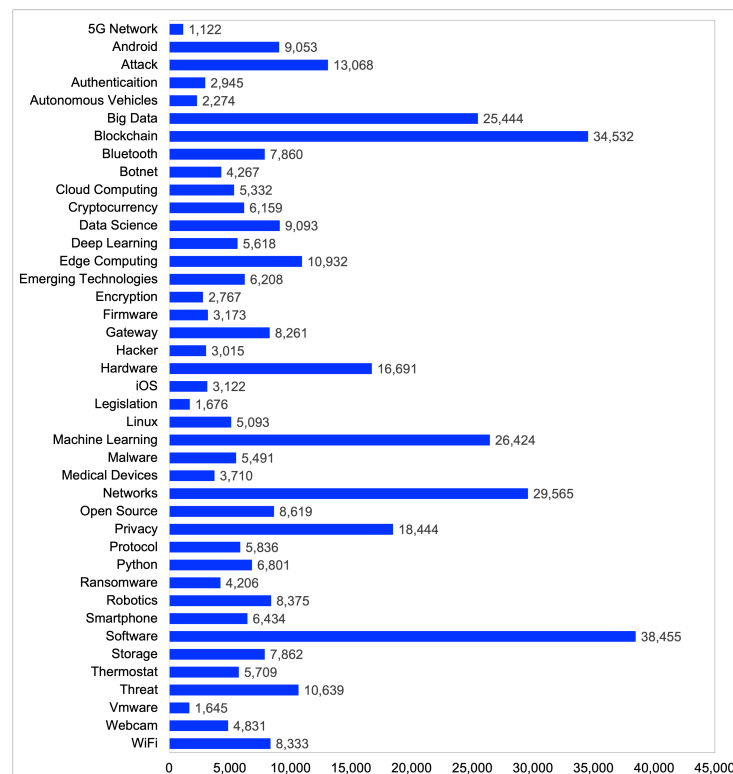


Figure 2. Number of tweets across an excerpt of emerging technologies.



Figure 3. ZTNB results across an excerpt of emerging technologies.



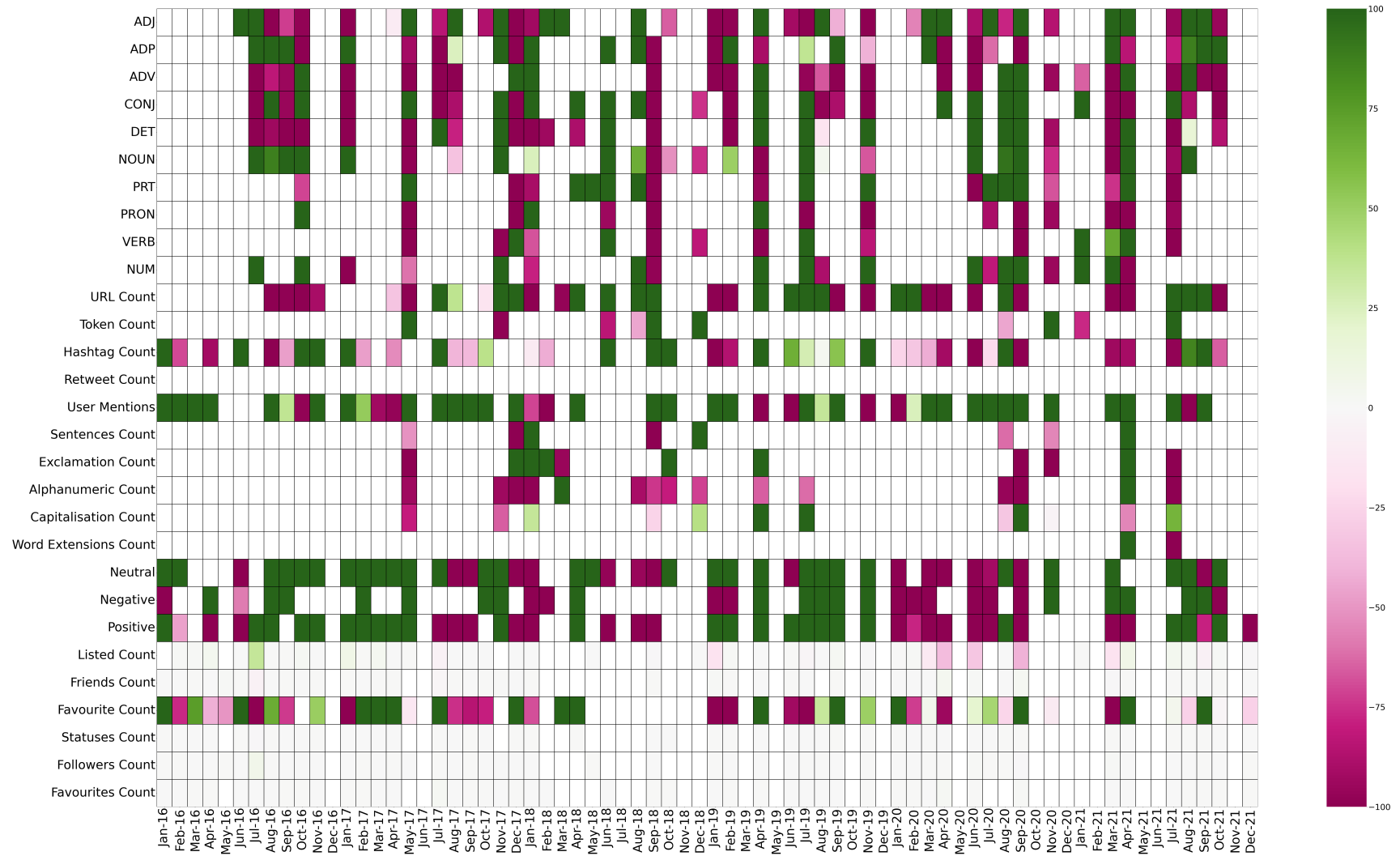


Figure 4. Chronological ZTNB results across tweets referencing the technology term ‘data science’.

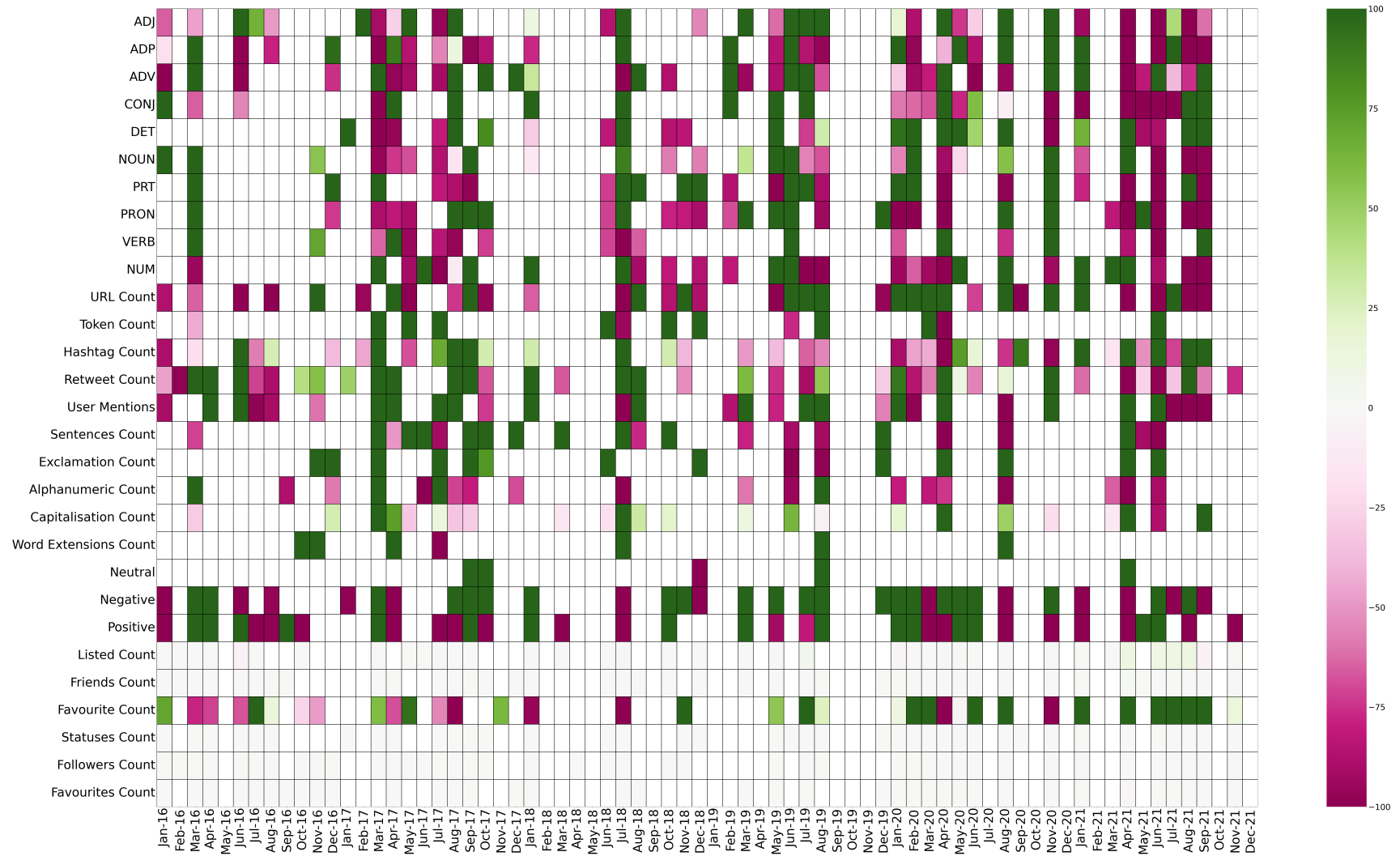


Figure 5. Chronological ZTNB results across tweets referencing the technology term 'attack'.

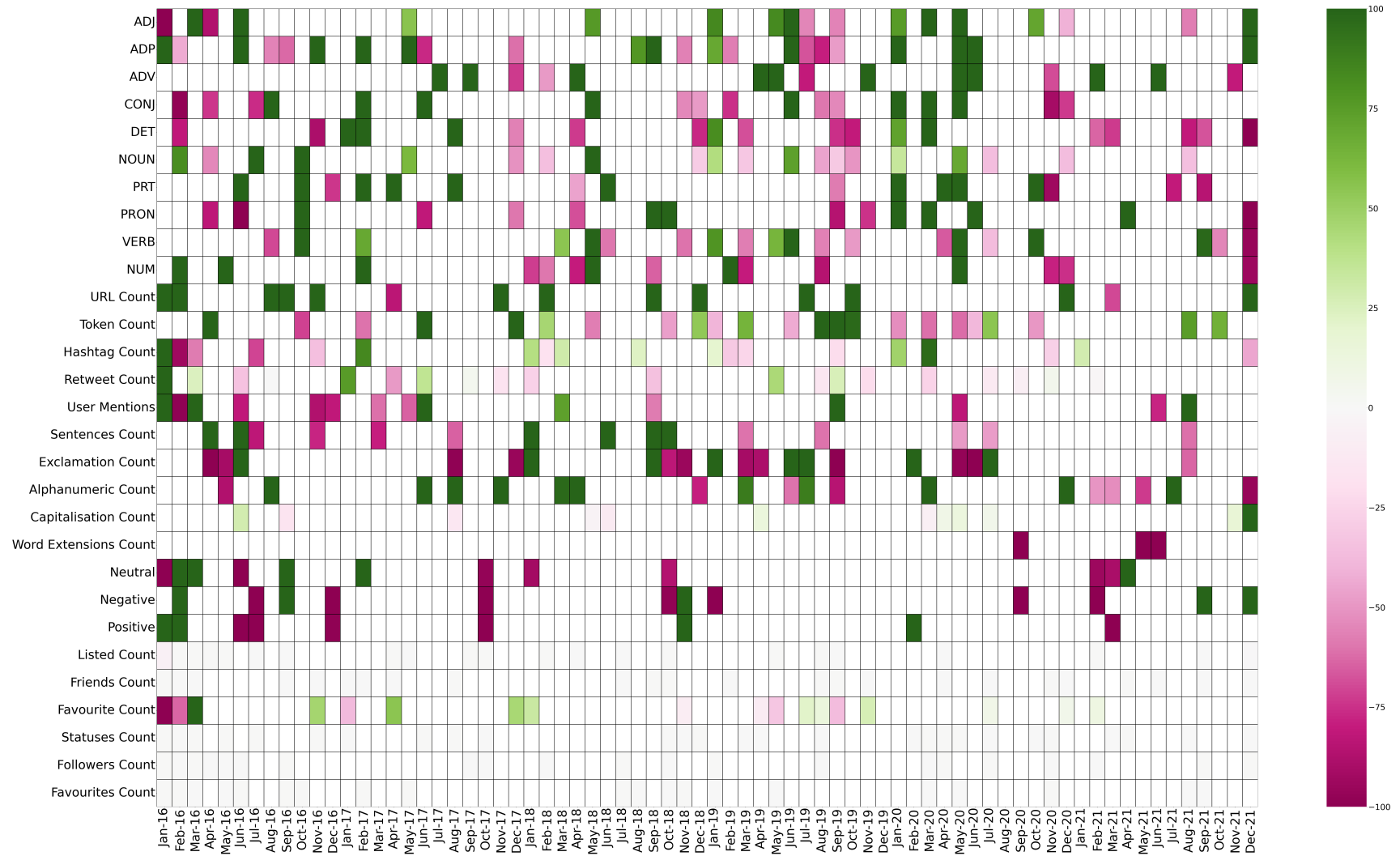


Figure 6. Chronological ZTNB results across tweets referencing the technology term ‘machine learning’.



Figure 7. Chronological ZTNB results across tweets referencing the technology term ‘software’.

Figures 4–7 illustrate chronological ZTNB results across the impressions of tweets referencing the technology terms ‘data science’, ‘attack’, ‘machine learning’, and ‘software’ respectively. For each technology, it is reported that different features have positively and negatively contributed towards the impressions of tweets across the timeline. A potential reason behind this is that social media conversations are dynamic and constantly evolving; thus, the sentiment and tone of conversations can shift rapidly in response to changing events, opinions, and circumstances [54]. The variation in the contribution of features to tweet impression may also be attributed to the diverse interests and concerns of the communities discussing these technologies, reflecting the unique preferences and priorities of each group in the context of the respective technology.

An analysis of all four datasets reveals different impression patterns based on the features used. For example, when analysing the relationship between the sentiment expressed in such tweets, neutral tweets referencing ‘attacks’ have rarely contributed to tweet impressions, whereas neutral ‘data science’-referencing tweets have, overall, positively contributed to the impression of tweets over time. The difference in such impressions may vary depending on context and the audience. One may assume that, generally, neutral tweets about ‘data science’ have more impressions than those about ‘cyber attacks’, as it is a relatively broad and popular topic that is relevant to many industries and fields of study, while ‘cyber attacks’ are a more specific and potentially niche topic that is primarily of interest to a smaller audience of security professionals and experts. In addition, such tweets, generally, may be more focused on reporting negative news and highlighting potential risks and vulnerabilities. This is shown in Figure 5, where negative sentiment expressed in ‘cyber attack’-related tweets across the timeline has, overall, positively contributed to the impressions of tweets, and those that express positive sentiment have produced opposite results.

In the context of business decision making and emerging technologies, it is also important to consider the stages of adoption, and how such stages can considerably impact communication and marketing strategies. During the initial phases of diffusion, when the primary audience consists of innovators and early adopters [55], organisations should focus on leveraging positive sentiment surrounding terms as such individuals are typically risk-takers and are more likely to respond favourably to optimistic narratives regarding emerging technologies. This may be illustrated in Figure 4, where a year’s worth of positive sentiment expressed in ‘data science’-related tweets (from July 2016 to July 2017), a field and term which is growing in popularity, have positively contributed to the impressions of tweets. Organisations may modify their communication strategies as the technology advances to the early and late majority stages. During this phase, the audience is frequently more concerned with practical applications and problem-solving [55]. Thus, narratives that convey neutral or negative sentiments regarding issues and challenges may resonate more strongly with these users, thereby enhancing their impressions. This reporting emerges in Figure 4 between January 2019 and November 2019. Laggards are typically the last to adopt new technology, and require substantial evidence of its utility and dependability [55]. In this situation, organisations should strive to provide a balanced narrative, which includes both positive discussions of the benefits of ‘data science’ and honest discussions of the challenges it poses. Such a balanced narrative can effectively enhance this group’s perceptions of technology-related narratives. This behaviour is reported between July 2021 and December 2021.

Other patterns emerge where, for example, the inclusion of user mentions in ‘data science’-related tweets have shown to frequently positively contribute to the impression of tweets over the timeline when compared with those referencing ‘software’, which have predominantly not contributed or negatively affected the overall impression of the tweet. This result indicates the difference in the behaviour of users when sharing tweets referencing different technology-related terms. This insight can guide businesses in engaging with their audience. For instance, in collaborative fields such as ‘data science’, mentioning other users, especially experts or influencers, could enhance the credibility and usefulness of the tweet

for the audience, making it more informative and engaging. However, due to the nature of the subject, organisations or individuals may choose to not publicise information about 'cyber attacks' to prevent damage to their reputation or to protect sensitive information.

The inclusion of URLs in tweets across the timeline has both positively and negatively contributed to the impressions of tweets. For tweets referencing 'data science', 'machine learning', and 'software', URLs can provide additional information and context, linking to articles, reports, or other sources of information that provide additional context and detail. However, users may be more hesitant to engage with tweets containing URLs about 'attacks', as they might be concerned about the credibility and safety of the linked content. The fear of clicking on malicious links or being exposed to scams could lead to lower engagement, and in the case of Figure 5, consistently decreases tweet impressions.

There are no obvious patterns where other content-based features have consistently contributed to the impression of tweets from the datasets. However, such results suggest organisations have some flexibility to experiment with different content styles, formats, and themes without being confined to a specific structure or format, which subsequently allows unique and customised content that resonates with their specific audience. This finding also emphasises the importance of continuously monitoring and analysing social media data to understand what content is driving the impressions of posts, and that organisations should not rely on preconceived notions of what kind of content will drive impressions but should instead use data to guide their content strategies accordingly.

For all datasets, account-based features, such as the number of followers or the favourites count, have had very little or no contribution to the impression of tweets. This may be an indication that, although having a large social media following helps increase the visibility and reach of tweets [56], other factors, such as sharing relevant, interesting, and quality content at the right time, may contribute more towards their impressions. That is, a large social media following may not be needed to construct impressionable emerging technology-related narratives which have the intention to impact its readers.

## 7. Conclusions

Social media platforms play a significant role in facilitating business decision making, especially in the context of emerging technologies. These platforms offer a rich source of data from a global audience, allowing organisations to gather insights into the landscape of their market, as well as enabling direct communication with consumers, which opens up opportunities to foster relationships and gather information about their attitudes and experiences. As a result, several studies have investigated social media data towards supporting business decision making processes. However, such studies and tools do not present a comprehensive picture of which features contribute to the impressions of emerging technology-related narratives. Therefore, organisations may not have a complete understanding of what specific elements of their content are resonating with their audience and when such elements are most impressionable.

This paper presents a data-driven and scalable framework for automatically identifying key factors which maximise the impact of emerging technology-related tweets in the social media space. Such a framework is powered by the automatic collection and analysis of social media discourse containing references to emerging technologies. In addition to the collection of the features associated with the user account which shared the tweets online, content-based features used in tweets, such as the number of sentences, URLs, and hashtags, as well as the sentiment expressed in the text, were also extracted. Then, for each tweet, the dependent variable, i.e., the number of impressions a tweet received, was determined as the metric to measure tweet impact. The statistical data model, Zero-Truncated Negative Binomial, was applied to measure the contribution of the independent features towards the dependent variable.

The findings highlight the dynamic nature of social media where the contribution of features to emerging technology referenced tweet impressions can vary significantly



over time, reflecting shifts in user behaviour and interests. The key findings summarised herein are:

- Certain content-based features, such as the number of words and pronouns, showed a positive correlation with tweet impressions, increasing impressions by up to 2.8%.
- Tweets expressing negative sentiments were more likely to be impressionable, potentially due to the negativity bias where such content elicits stronger emotional responses, thus driving higher engagement and virality.
- Several features were found to negatively impact tweet impressions, including the use of more hashtags, verbs, sentences, conjunctions, user mentions, determiners, and adpositions. Notably, a higher number of words with trailing characters significantly reduced tweet impressions by 8.6%.
- Features based on the user's account, like the number of followers or favourites count, showed little to no contribution to the impressionability of tweets, suggesting that content quality and relevance might outweigh the popularity of the account.
- Positive sentiments associated with technologies perceived as beneficial (e.g., data science, machine learning) significantly contributed to tweet impressions, whereas positive sentiments towards negatively viewed technologies (e.g., cyber threats) decreased impressions.
- The inclusion of URLs in tweets had a mixed impact on impressions, enhancing engagement for tweets about general technologies, but potentially reducing engagement for tweets about sensitive topics (e.g., cyber attacks), due to possible concerns over link safety.

The success of these efforts is influenced by various factors, including the size and engagement of the audience, the timing and frequency of the posts, the content of the tweets, and the existing level of interest and awareness of the technology. The findings of this study can be situated within the broader theoretical frameworks of social media effects, diffusion of innovations, and social media marketing. From a media effects perspective, the research supports the initiation of agenda-setting and framing theories, which suggest that the way information is presented—such as the choice of pronouns, negative sentiments, or use of URLs—can shape public perceptions and engagement with technology-related content. In particular, the role of negative sentiments highlights the negativity bias, where emotionally charged content tends to garner more attention and impressions, aligning with established psychological theories [57]. Additionally, the study's exploration of emerging technologies connects to the diffusion of innovations framework [58], which assumes that social media serves as a channel for disseminating new technologies and fostering early adoption, with certain content features (like the perceived benefits of data science) enhancing this process. By identifying the specific factors that influence the visibility and impact of social media posts, this research advances knowledge in the field by offering a scalable, data-driven framework that organisations can apply to optimise their social media strategies. It also bridges a gap in existing literature by demonstrating how content features vary over time in driving impressions, thus offering a dynamic perspective on how organisations can stay competitive in the fast-evolving social media landscape, particularly during the critical early stages of technology adoption.

## 8. Future Work

Given the positive findings of this initial study, there are several pathways for future research that would not only broaden the scope of our current understanding, but also refine the application of data analytics and machine learning in optimising social media strategies. A crucial development would involve creating and testing algorithms capable of predicting and optimising the timing and format of social media posts for maximum user engagement. Selecting appropriate machine learning models is important, with options ranging from regression models for continuous outcomes, classification models for engagement levels, to more complex neural networks for handling sequential data. Additionally, an important aspect of future work will be to compare different metrics of measuring engagement to

identify which provide the most holistic and actionable insights. This comparative analysis could lead to a deeper understanding of which metrics best correlate with successful outcomes, thus refining the tools available for digital marketers to tailor content effectively. The integration of these models into existing digital marketing tools can significantly augment the decision-making process, allowing for real-time adjustments and strategic planning to maximize social media engagement.

Further comparative analysis across different social media platforms would help in identifying the most effective strategies tailored to each platform's unique audience and capabilities. This could include testing the same content on platforms like Twitter, LinkedIn, and Facebook to gauge differential impacts and understand the nuances of platform-specific content efficacy, particularly for emerging technologies. For organisations, such insights could pinpoint which content types perform best on each platform and adjust their strategies accordingly. For example, video content might resonate more on Facebook due to its robust multimedia support, while concise, timely updates might perform better on Twitter due to its fast-paced nature. LinkedIn could be more effective for in-depth articles or professional insights due to its professional network context. Moreover, an in-depth exploration of paid versus organic reach is essential. Quantitative measures of the effectiveness of paid promotions compared to organic strategies could provide actionable insights that help organisations optimise their social media budgets more effectively. Additionally, longitudinal studies on the long-term effects of social media strategies on brand recognition and engagement could yield significant insights into the temporal dynamics of digital marketing success. Analysing the life cycle of posts to comprehend how content maintains relevance or contributes to a sustained digital presence over time would be particularly valuable. Network analysis can also provide insights into how posts are shared and how it influences user engagement. By examining the connections between tweets through retweets, mentions, and replies, and mapping out the relationships between users who engage with these tweets, we can identify influential nodes (users or tweets that have disproportionate influence), the reach of specific tweets, and how information flows through complex social networks. This approach allows us to understand the structural properties of the network—such as density, centrality, and clustering. These metrics can tell us not just which tweets are popular, but why they are popular, revealing patterns that might not be apparent from statistical analysis. For example, a tweet that acts as a bridge between two otherwise distinct user communities might have greater strategic value, influencing diverse groups and spreading information across different segments of the network.

Lastly, predictive analytics could also play a pivotal role in forecasting trends and preparing strategies that align with future engagement patterns, ultimately enhancing the strategic use of social media for maximum impact. There is also room to delve deeper into the subjectivity of social media impressions by performing detailed sentiment analysis of the responses associated with highly engaged tweets. This investigation would involve expanding the dataset to include user interactions and comments directly linked to significant posts. This analysis would help understand the nuances of user engagement and the impact of different sentiments on the perceived value of the content. To provide a much deeper insight into audience preferences and content effectiveness for more targeted and effective social media strategies, other available data, such as demographics and content themes, could also be explored.

**Author Contributions:** Conceptualization, L.W.; Methodology, L.W. and P.B.; Validation, L.W.; Investigation, L.W.; Data curation, L.W.; Writing—original draft, L.W. and E.A.; Writing—review and editing, L.W. and E.A.; Funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** For the purpose of open access, the author has applied a CC BY public copyright licence (where permitted by UKRI, 'Open Government Licence' or 'CC BY-ND public copyright licence' may be stated instead) to any Author Accepted Manuscript version arising. This work was funded by the Economic and Social Research Council (ESRC), grant 'Discribe—Digital Security by Design (DSbD) Programme'. REF ES/V003666/1.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original data used in this study is presented in [17] and is available to access here: [https://github.com/LowriWilliams/IoT\\_referenced\\_tweets](https://github.com/LowriWilliams/IoT_referenced_tweets) (accessed on 24 March 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Chesbrough, H.W. *Open Innovation: The New Imperative for Creating and Profiting from Technology*; Harvard Business Press: Brighton, MA, USA, 2003.
- Breedveld, L. Combining LCA and RA for the integrated risk management of emerging technologies. *J. Risk Res.* **2013**, *16*, 459–468. [[CrossRef](#)]
- He, W.; Wang, F.K.; Akula, V. Managing extracted knowledge from big social media data for business decision making. *J. Knowl. Manag.* **2017**, *21*, 275–294. [[CrossRef](#)]
- Zhang, H.; Zang, Z.; Zhu, H.; Uddin, M.I.; Amin, M.A. Big data-assisted social media analytics for business model for business decision making system competitive analysis. *Inf. Process. Manag.* **2022**, *59*, 102762. [[CrossRef](#)]
- Yang, J.; Xiu, P.; Sun, L.; Ying, L.; Muthu, B. Social media data analytics for business decision making system to competitive analysis. *Inf. Process. Manag.* **2022**, *59*, 102751. [[CrossRef](#)]
- Ghani, N.A.; Hamid, S.; Hashem, I.A.T.; Ahmed, E. Social media big data analytics: A survey. *Comput. Hum. Behav.* **2019**, *101*, 417–428. [[CrossRef](#)]
- Dwivedi, Y.K.; Ismagilova, E.; Hughes, D.L.; Carlson, J.; Filieri, R.; Jacobson, J.; Jain, V.; Karjaluoto, H.; Kefi, H.; Krishen, A.S.; et al. Setting the future of digital and social media marketing research: Perspectives and research propositions. *Int. J. Inf. Manag.* **2021**, *59*, 102168. [[CrossRef](#)]
- Algharabat, R.; Rana, N.P.; Dwivedi, Y.K.; Alalwan, A.A.; Qasem, Z. The effect of telepresence, social presence and involvement on consumer brand engagement: An empirical study of non-profit organizations. *J. Retail. Consum. Serv.* **2018**, *40*, 139–149. [[CrossRef](#)]
- Kaur, P.; Dhir, A.; Rajala, R.; Dwivedi, Y. Why people use online social media brand communities: A consumption value theory perspective. *Online Inf. Rev.* **2018**, *42*, 205–221. [[CrossRef](#)]
- Lal, B.; Ismagilova, E.; Dwivedi, Y.K.; Kwayu, S. Return on investment in social media marketing: Literature review and suggestions for future research. In *Digital and Social Media Marketing: Emerging Applications and Theoretical Development*; Springer: Cham, Switzerland, 2020; pp. 3–17.
- Dhanesh, G.S. Putting engagement in its PRoper place: State of the field, definition and model of engagement in public relations. *Public Relations Rev.* **2017**, *43*, 925–933. [[CrossRef](#)]
- Johnston, K.A.; Taylor, M. Engagement as communication: Pathways, possibilities, and future directions. In *The Handbook of Communication Engagement*; Wiley-Blackwell: Hoboken, NJ, USA, 2018; pp. 1–15.
- Johnston, K.A. Toward a theory of social engagement. In *The Handbook of Communication Engagement*; Wiley-Blackwell: Hoboken, NJ, USA, 2018; pp. 17–32.
- Saxton, G.D.; Waters, R.D. What do stakeholders like on Facebook? Examining public reactions to nonprofit organizations' informational, promotional, and community-building messages. *J. Public Relations Res.* **2014**, *26*, 280–299. [[CrossRef](#)]
- Zhang, Y.; Dong, C.; Cheng, Y. How do nonprofit organizations (NPOs) effectively engage with the public on social media? Examining the effects of interactivity and emotion on Twitter. *Internet Res.* **2022**, *33*, 550–577. [[CrossRef](#)]
- Taecharungroj, V. "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data Cogn. Comput.* **2023**, *7*, 35. [[CrossRef](#)]
- Williams, L.; Anthi, E.; Burnap, P. A scalable and automated framework for tracking the likely adoption of emerging technologies. *Information* **2024**, *15*, 237. [[CrossRef](#)]
- Kaplan, A.M.; Haenlein, M. Users of the world, unite! The challenges and opportunities of Social Media. *Bus. Horizons* **2010**, *53*, 59–68. [[CrossRef](#)]
- Kietzmann, J.H.; Hermkens, K.; McCarthy, I.P.; Silvestre, B.S. Social media? Get serious! Understanding the functional building blocks of social media. *Bus. Horizons* **2011**, *54*, 241–251. [[CrossRef](#)]
- Felix, R.; Rauschnabel, P.A.; Hinsch, C. Elements of strategic social media marketing: A holistic framework. *J. Bus. Res.* **2017**, *70*, 118–126. [[CrossRef](#)]
- Bollen, J.; Mao, H.; Zeng, X. Twitter mood predicts the stock market. *J. Comput. Sci.* **2011**, *2*, 1–8. [[CrossRef](#)]
- Asur, S.; Huberman, B.A. Predicting the future with social media. In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, ON, Canada, 31 August–3 September 2010; Volume 1, pp. 492–499.
- Zaman, S.; Yaqub, U.; Saleem, T. Analysis of bitcoin's price spike in context of Elon Musk's twitter activity. *Glob. Knowl. Mem. Commun.* **2022**, *72*, 341–355. [[CrossRef](#)]

24. Wu, T.; Reynolds, J.; Wu, J.; Schlegelmilch, B.B. CEOs as corporate ambassadors: Deciphering leadership communication via Twitter. *Online Inf. Rev.* **2022**, *46*, 787–806. [[CrossRef](#)]
25. Stieglitz, S.; Dang-Xuan, L. Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *J. Manag. Inf. Syst.* **2013**, *29*, 217–248. [[CrossRef](#)]
26. Nguyen, T.H.; Shirai, K.; Velcin, J. Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.* **2015**, *42*, 9603–9611. [[CrossRef](#)]
27. Corea, F. Can twitter proxy the investors' sentiment? The case for the technology sector. *Big Data Res.* **2016**, *4*, 70–74. [[CrossRef](#)]
28. Kwarteng, M.A.; Ntsiful, A.; Botchway, R.K.; Pilik, M.; Oplatková, Z.K. Consumer Insight on Driverless Automobile Technology Adoption via Twitter Data: A Sentiment Analytic Approach. In Proceedings of the Re-Imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation: IFIP WG 8.6 International Conference on Transfer and Diffusion of IT, TDIT 2020, Tiruchirappalli, India, 18–19 December 2020; Proceedings, Part I; Springer: Berlin/Heidelberg, Germany, 2020; pp. 463–473.
29. Ikram, M.T.; Butt, N.A.; Afzal, M.T. Open source software adoption evaluation through feature level sentiment analysis using Twitter data. *Turk. J. Electr. Eng. Comput. Sci.* **2016**, *24*, 4481–4496. [[CrossRef](#)]
30. Mardjo, A.; Choksuchat, C. HyVADRf: Hybrid VADER–Random Forest and GWO for Bitcoin Tweet Sentiment Analysis. *IEEE Access* **2022**, *10*, 101889–101897. [[CrossRef](#)]
31. Caviggioli, F.; Lamberti, L.; Landoni, P.; Meola, P. Technology adoption news and corporate reputation: Sentiment analysis about the introduction of Bitcoin. *J. Prod. Brand Manag.* **2020**, *29*, 877–897. [[CrossRef](#)]
32. Chamlerwat, W.; Bhattarakosol, P.; Rungkasiri, T.; Haruechaiyasak, C. Discovering Consumer Insight from Twitter via Sentiment Analysis. *J. Univers. Comput. Sci.* **2012**, *18*, 973–992.
33. Qaiser, S.; Yusoff, N.; Ahmad, F.K.; Ali, R. Sentiment analysis of impact of technology on employment from text on twitter. *Int. J. Interact. Mob. Technol.* **2020**, *14*, 88–103. [[CrossRef](#)]
34. Ji, Y.G.; Chen, Z.F.; Tao, W.; Li, Z.C. Functional and emotional traits of corporate social media message strategies: Behavioral insights from S&P 500 Facebook data. *Public Relations Rev.* **2019**, *45*, 88–103.
35. Kim, A.E.; Hansen, H.M.; Murphy, J.; Richards, A.K.; Duke, J.; Allen, J.A. Methodological considerations in analyzing Twitter data. *J. Natl. Cancer Inst. Monogr.* **2013**, *2013*, 140–146. [[CrossRef](#)]
36. Hallett, J.; Lata Nautiyal, B.S.A.R. *The Cyber Security Body of Knowledge (CyBOK)—CyBOK Mapping Reference Version 1.1*; University of Bristol: Bristol, UK, 2022.
37. Tsugawa, S.; Ohsaki, H. On the relation between message sentiment and its virality on social media. *Soc. Netw. Anal. Min.* **2017**, *7*, 1–14. [[CrossRef](#)]
38. Mahdikhani, M. Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of COVID-19 pandemic. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100053. [[CrossRef](#)]
39. Javed, A.; Burnap, P.; Williams, M.L.; Rana, O.F. Emotions behind drive-by download propagation on Twitter. *ACM Trans. Web (TWEB)* **2020**, *14*, 1–26. [[CrossRef](#)]
40. Suh, J.H. Comparing writing style feature-based classification methods for estimating user reputations in social media. *Springer-Plus* **2016**, *5*, 261. [[CrossRef](#)] [[PubMed](#)]
41. Bird, S. NLTK: The natural language toolkit. In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, Sydney, Australia, 17–18 July 2006; pp. 69–72.
42. Cha, M.; Benevenuto, F.; Haddadi, H.; Gummadi, K. The world of connections and information flow in twitter. *IEEE Trans. Syst. Man Cybern. Part A Syst. Humans* **2012**, *42*, 991–998.
43. Suh, B.; Hong, L.; Pirolli, P.; Chi, E.H. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In Proceedings of the 2010 IEEE Second International Conference on Social Computing, Minneapolis, MN, USA, 20–22 August 2010; pp. 177–184.
44. Rivadeneira, L.; Yang, J.B.; López-Ibáñez, M. Predicting tweet impact using a novel evidential reasoning prediction method. *Expert Syst. Appl.* **2021**, *169*, 114400. [[CrossRef](#)]
45. Iacopini, M.; Santagiustina, C.R. Filtering the intensity of public concern from social media count data with jumps. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2021**, *184*, 1283–1302. [[CrossRef](#)]
46. Hilbe, J. *Modeling Count Data*; Cambridge University Press: Cambridge, UK, 2014.
47. Hilbe, J. *Negative Binomial Regression*; Cambridge University Press: Cambridge, UK, 2011.
48. Gurmu, S. Tests for detecting overdispersion in the positive Poisson regression model. *J. Bus. Econ. Stat.* **1991**, *9*, 215–222. [[CrossRef](#)]
49. Long, J.S. *Regression Models for Categorical and Limited Dependent Variables (Advanced Quantitative Techniques in the Social Sciences)*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 1997; Volume 7.
50. de Hoog, N.; Verboon, P. Is the news making us unhappy? The influence of daily news exposure on emotional states. *Br. J. Psychol.* **2020**, *111*, 157–173. [[CrossRef](#)]
51. Jenders, M.; Kasneci, G.; Naumann, F. Analyzing and predicting viral tweets. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 657–664.
52. Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*; Cambridge University Press: Cambridge, UK, 2013; Volume 53.
53. Winkelmann, R. *Econometric Analysis of Count Data*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.

54. Kouloumpis, E.; Wilson, T.; Moore, J. Twitter sentiment analysis: The good the bad and the omg! In Proceedings of the Fifth International AAAI Conference on Web and Social Media, Barcelona, Spain, 17–21 July 2011; Volume 5, pp. 538–541.
55. Rogers, E.M. *Diffusion of Innovations*; Free Press: New York, NY, USA, 2003; Volume 576.
56. Cha, M.; Haddadi, H.; Benevenuto, F.; Gummadi, K. Measuring User Influence in Twitter: The Million Follower Fallacy. In Proceedings of the Fourth International AAAI Conference on Web and Social Media, Washington, DC, USA, 23–26 May 2010; Volume 4, pp. 10–17.
57. Baumeister, R.F.; Bratslavsky, E.; Finkenauer, C.; Vohs, K.D. Bad is stronger than good. *Rev. Gen. Psychol.* **2001**, *5*, 323–370. [[CrossRef](#)]
58. Rogers, E.M.; Singhal, A.; Quinlan, M.M. Diffusion of innovations. In *An Integrated Approach to Communication Theory and Research*; Routledge: Abingdon-on-Thames, UK, 2014; pp. 432–448.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.