Article

Applying Machine Learning in Marketing: An Analysis Using the NMF and k-Means Algorithms

Victor Gallego [1,2], Jessica Lingan [1], Alfons Freixes [1], Angel A. Juan [2,*] and Celia Osorio [2]

[1] Business Analytics Research Group, Euncet Business School, 08225 Terrassa, Spain; afreixes@euncet.com (A.F.)
[2] Research Center on Production Management and Engineering, Universitat Politècnica de València, 03801 Alcoy, Spain
[*] Correspondence: ajuanp@upv.es

Abstract: The integration of machine learning (ML) techniques into marketing strategies has become increasingly relevant in modern business. Utilizing scientific manuscripts indexed in the Scopus database, this article explores how this integration is being carried out. Initially, a focused search is undertaken for academic articles containing both the terms "machine learning" and "marketing" in their titles, which yields a pool of papers. These papers have been processed using the Supabase platform. The process has included steps like text refinement and feature extraction. In addition, our study uses two key ML methodologies: topic modeling through NMF and a comparative analysis utilizing the k-means clustering algorithm. Through this analysis, three distinct clusters emerged, thus clarifying how ML techniques are influencing marketing strategies, from enhancing customer segmentation practices to optimizing the effectiveness of advertising campaigns.

Keywords: machine learning; digital marketing; algorithms; artificial intelligence

1. Introduction

The impact of marketing on businesses is crucial to ensure their growth and development [1]. Marketing enables brand awareness and recognition, customer acquisition and loyalty, differentiation from competitors, and increased sales, thus contributing to the sustainability of business profitability [2]. Systematic data management and data processing is essential to obtain timely and accurate information about the market, products and customers, as well as the business environment, as it facilitates decision making in the different marketing processes. With the advancement of information technologies, companies need to apply new methods for data processing and analysis [3]. Machine learning (ML) can be used to gain new insights into consumer behavior, improve the performance of marketing actions [4], formulate strategic marketing decisions [5], identify competitors, effectively manage products and prices, etc. [6]. Customer reviews are considered essential for both strategy development and consumer attraction. Likewise, topic modeling algorithms have emerged as a popular analytical tool, enabling companies to make decisions on (i) product enhancement and development; (ii) service recovery and improvements; (iii) pricing; and (iv) customer recruitment and retention [7]. The use of ML offers great possibilities for companies, marketing professionals, and academics [8]. However, it has become a challenge to identify the methods, techniques, and tools being applied to different marketing processes. Additionally, determining the level of application and technological adaptation by companies is complex [9]. This includes marketing activities, commercial management, improving customer experience, optimizing operations, and creating new business models.

The increasing application of ML in marketing strategies not only shows technological advancements in this field, but also brings to light the diverse academic perspectives that investigate both the impacts and the potential of ML in marketing. Studies such as the one by Herhausen et al. [4] highlight the integration of ML in targeting consumers and customizing offers, while also addressing the challenges of data privacy concerns and the

opacity of algorithms. Likewise, Arasu et al. [10] explore how data analytics can predict customer purchase preferences by proposing an integrated ML approach to social media marketing. Their approach employs text mining and analysis techniques to process and understand large volumes of social media data, thereby enhancing marketing effectiveness. Similarly, Bonab et al. [11] focus on the challenge of adapting product recommendations across diverse markets by developing multiple models. They introduce FOREC, a novel approach designed to optimize market adaptation through structured steps of pre-training, forking, and fine-tuning of market data. The study also considers models such as the Generalized Matrix Factorization, Multi-Layer Perceptron, and Neural Matrix Factorization. Tirunillai and Tellis [12] introduce Latent Dirichlet Allocation (LDA) to marketing research, utilizing it to analyze online product reviews. Their application of LDA revealed significant dimensions of product quality, referred to as topics, which captured diverse aspects of consumer satisfaction. Another application of ML in marketing is studied by Kalinić et al. [13]. Specifically, they use Artificial Neural Networks and Structural Equation Modeling in order to model consumer satisfaction, thus identifying factors influencing consumer satisfaction.

Building on the analysis of existing research, our study will contribute to the literature by examining both academic publications and case studies to further explore the impact and potential of ML techniques on marketing practices [14]. To this end, the following objectives have been established: (i) to identify the ML methods, techniques, and tools that are more frequently employed in marketing practices; (ii) to find out marketing processes in which ML techniques are being applied in an intensive manner; and (iii) to establish successful cases of the use of ML techniques in marketing. To achieve the aforementioned research objectives, the study makes use of the Scopus database using two precise keywords, "machine learning" and "marketing", in order to obtain relevant data. After identifying some of the most representative scientific manuscripts on the subject, they are systematically processed. Two analysis methodologies are employed: (i) topic modeling using the non-negative matrix factorization (NMF) algorithm [15] and (ii) a comparative analysis using the k-means clustering algorithm. The NMF has allowed us to discover and understand the predominant themes in the dataset of selected articles. Meanwhile, the purpose of using k-means clustering is to analyze and validate the results of the NMF. Hence, one methodological novelty of our work is the combined use of NMF and the k-means clustering algorithm to analyze academic literature at the intersection of ML and marketing. While both algorithms are already well known, their combined application to review and compare existing papers on marketing represents an innovative approach.

The rest of the document is structured as follows. Section 2 identifies the research methodology that explains the process of obtaining and systematizing the information based on the main keywords of the study. Section 3 introduces the first cluster of papers obtained, which describes the applications of ML algorithms in marketing. Section 4 refers to a second cluster, which refers to the application of big data and ML techniques in marketing processes and their impact on consumer behavior. Section 5 describes the results obtained in a third cluster, which refers to the applications of ML in digital marketing. Section 6 introduces future trends related to the application of ML in marketing practices. Finally, Section 7 summarizes the main findings of the article.

2. Methodology

This section describes the methodology employed in our study, from the data collection process to data management and preprocessing, feature extraction, topic modeling, comparative analysis, and evaluation.

2.1. Data Collection Process: Literature Search Procedure

The first step of this study is the identification of academic articles in which ML has been employed to develop marketing applications. We utilized the Scopus database, known for its wide range of peer-reviewed literature, to conduct the search. In order to ensure that

only the most relevant studies are included, the search criteria were precise. Hence, we searched for articles including both terms, "machine learning" and "marketing", in their title, also refining our search with several filters. (i) Document type was set as 'Article', thus excluding non-peer-reviewed materials; (ii) source type was restricted to 'Journal', thus ensuring sources were reputable academic journals; (iii) language was set to 'English', hence ensuring the findings are widely accessible; and (iv) the period considered is from 2018 to 2023. This approach resulted in a collection of 66 papers, marking a significant corpus of work at the intersection of ML and marketing. Regarding the time period chosen for the analysis of these articles, it was determined based on the availability of relevant articles rather than a preconceived timeframe.

2.2. Data Management and Storage: Database Development and Record Storage

After identifying relevant literature, we organized and stored these data systematically. We chose Supabase (https://supabase.com, accessed on 19 June 2024), an open-source platform for database storage, to manage our collection. We created an SQL database table to catalog each article, including fields for (i) ID (a unique identifier for each paper); (ii) created-at (the date and time of the record's database entry); (iii) title (the article's title as it appears in the source); (iv) abstract (a summary of the research and its outcomes); (v) year (the publication year of the article); (vi) cited (the citation counts of the article, indicating its influence); (vii) title-clean (a formatted version of the title for analysis); and (viii) abstract-cleaned (a standardized version of the abstract for uniform analysis). Notice that 'cleaned title' refers to the process of refining and preparing the titles of documents in the dataset by removing any unnecessary or extraneous information, such as special characters, formatting issues, or irrelevant words. This ensures that the titles are standardized and ready for analysis. Similarly, 'cleaned abstract' refers to the abstracts of documents in the dataset that have undergone a similar refinement process. This involves removing any noise or irrelevant information from the abstracts, such as filler words, repetitive phrases, or formatting inconsistencies. Figure 1 shows a bar chart of journals including at least 2 articles with the terms 'machine learning' and 'marketing' in their title.
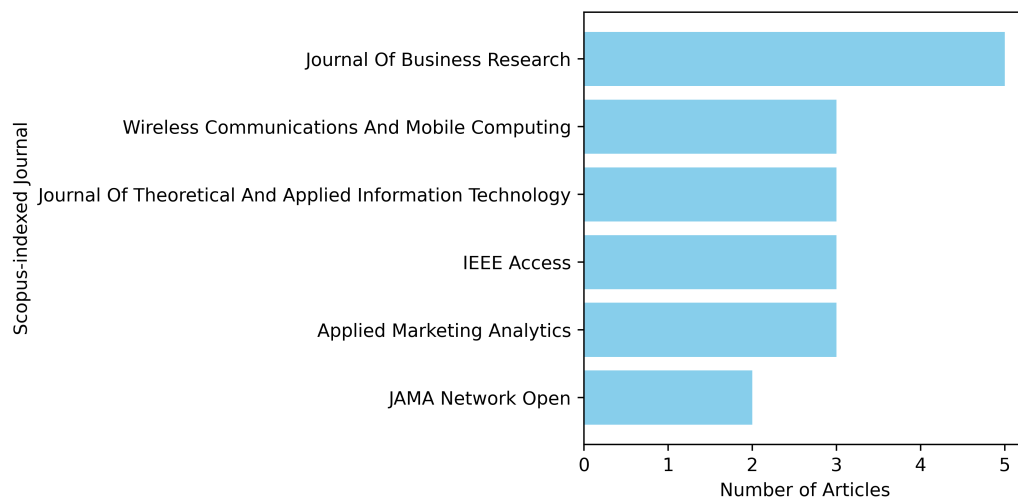

Figure 1. Journals with at least 2 articles including 'machine learning' and 'marketing' in their titles.

Additionally, Figure 2 presents the same analysis but the search criteria are expanded to include articles that contain the terms "machine learning" and "marketing" either in the title or in the abstract.
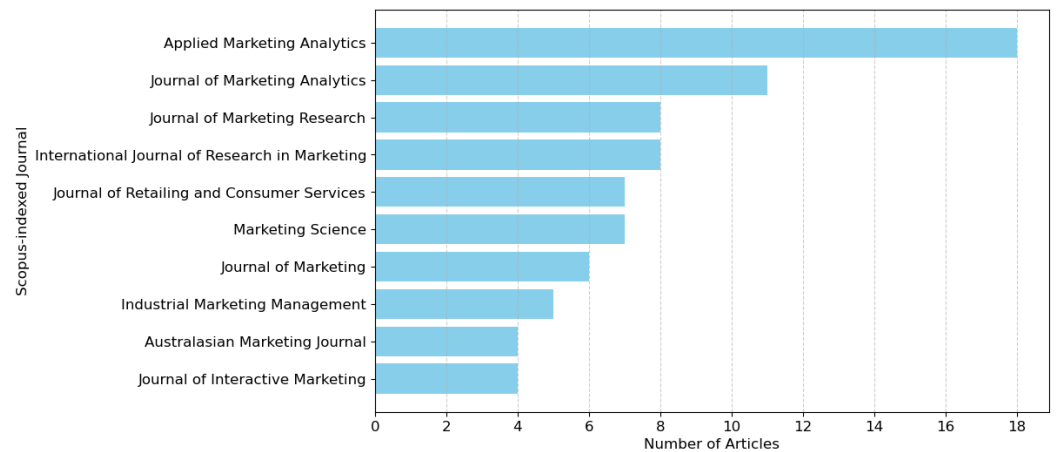
Figure 2. Journals with articles including 'machine learning' and 'marketing' in their titles or abstract.

Figure 3 displays the number of publications per year during the period 2018 to 2023. The blue line refers to articles in which both "machine learning" and "marketing" appear in their title, while the orange line also includes the possibility that these terms appear in the articles' abstracts.
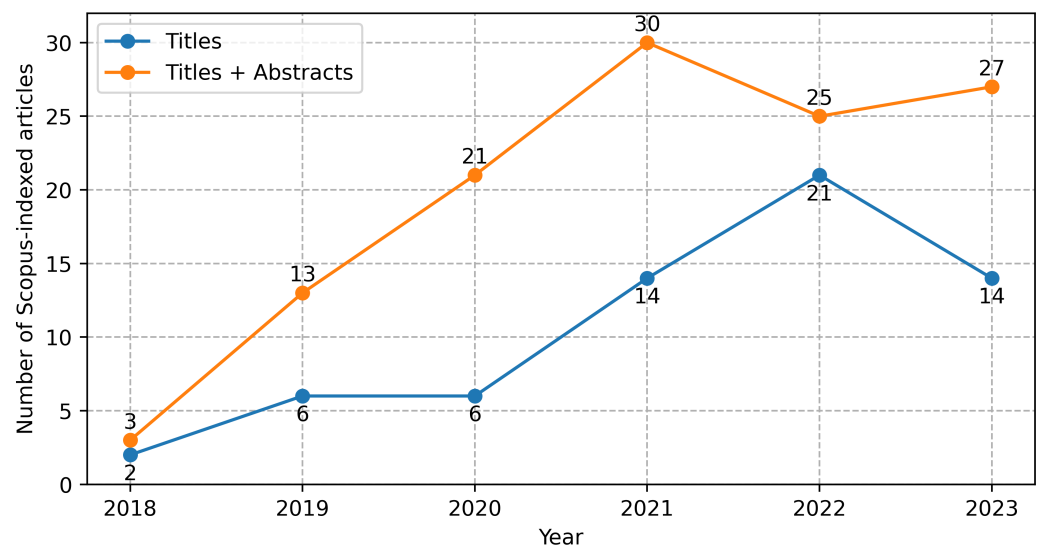

Figure 3. Annual evolution of Scopus-indexed journal articles with ML and marketing in their title (blue line) or in their title or abstract (orange line).

2.3. Data Preprocessing: Text Data Cleaning

The subsequent phase of our study was dedicated to text data cleaning, a key process aimed at enhancing the quality and consistency of the data extracted from the article titles and abstracts. This stage involved multiple processes to refine the text for subsequent analysis: (i) inclusion of custom stop words—the text cleaning procedure was augmented by incorporating a custom list of stop words, which included domain-specific terms such as "ai" and "digit", along with the standard stop words typically used in text preprocessing; (ii) tokenization—we segmented the text into discrete words or tokens to facilitate detailed thematic analysis; (iii) case normalization—all textual data were converted to lowercase to ensure uniformity across the dataset; (iv) numerical character removal—numerical characters were removed as they were considered irrelevant to our topic analysis; (v) contraction expansion—we expanded contractions to their full forms to enhance the clarity and consistency of the text; (vi) stemming—words were reduced to their root form using stem-

ming algorithms, which simplified the analysis by reducing the complexity of the dataset; (vii) punctuation and enhanced stop word removal—punctuation marks were removed, and the text was further refined using our expanded stop word list to eliminate unnecessary tokens; and (ix) exclusion of short words and spaced tokens—we excluded words shorter than two characters and tokens containing spaces, as their contribution to our analysis was small.

These meticulous preprocessing steps transformed the raw text into a structured format, ready for deeper analysis and ensuring the integrity and relevance of data. To illustrate the effectiveness of these procedures, Tables 1 and 2 show examples of cleaned titles and abstracts, respectively, from our dataset.

Table 1. Examples of cleaned titles from our dataset.

Title	Title Cleaned
Machine Learning in Marketing	machin learn market
Algorithmic bias in machine learning-based marketing models	algorithm bias machin learning-bas market model
Marketing analytics stages: Demystifying and deploying machine learning	market analyt stage demystifi deploy machin learn

Table 2. Examples of cleaned abstracts from our dataset.

Abstract	Abstract Cleaned
The widespread impacts of artificial intelligence (AI) and machine learning (ML) in many segments of society have not yet been felt strongly in the marketing field. Despite such shortfall, ML offers a variety of potential benefits, including the opportunity to apply more robust methods for the generalization of scientific discoveries. Trying to reduce this shortfall, this monograph has four goals. First, to provide marketing...	widespread impact artifici intellig machin learn ml mani segment societi felt strong market field despit shortfal ml offer varieti potenti benefit includ opportun appli robust method general scientif discoveri tri reduc shortfal monograph goal provid market...

2.4. Feature Extraction: Generation and Storage of Vector Embeddings

In the next step of our methodology, the cleaned textual data are transformed into high-dimensional vector embeddings. These embeddings will be used by a k-means clustering algorithm, which will allow us to categorize the papers into groups based on their intrinsic textual similarities. We use the 'text-embedding-ada-002' model to convert text into vectors that efficiently encapsulate semantic content. To ensure optimal representation while maintaining semantic integrity across large datasets, we pair this model with the 'cl100k-base' encoding scheme. The process adheres to a maximum token limit of 8000 per text segment, a threshold determined by the model's capacity. The initial steps for generating vector embeddings are outlined in the the Python code snippet shown in Listing 1.

Listing 1. Python code for embedding model and encoding parameters.

```python
# Define embedding model and encoding parameters
embedding_model = "text-embedding-ada-002"
embedding_encoding = "cl100k_base"
max_tokens = 8000

# Load data set and apply encoding
df = get_data_to_supabase(supabase=supabase, name_table='web_results')
encoding = tiktoken.get_encoding(embedding_encoding)
```

Post-generation, the vector embeddings are systematically stored within a Supabase database. This integration supports NLP tasks within our ongoing analytical processes. Listing 2 shows the integration process and data storage:

Listing 2. Python code for calculating tokens, filtering data, and generating embeddings.

```python
# Calculate number of tokens and filter data
df["n_tokens_title"] = df["title_clean"].apply(
    lambda x: len(encoding.encode(x)))
df = df[df.n_tokens_title <= max_tokens]
df["n_tokens_abstract"] = df["abstract_clean"].apply(
    lambda x: len(encoding.encode(x)))
df = df[df.n_tokens_abstract <= max_tokens]

# Generate vector embeddings for titles and abstracts
df["embedding_title"] = df["title_clean"].apply(
    lambda x: get_embedding(x, engine=embedding_model))
df["embedding_abstract"] = df["abstract_clean"].apply(
    lambda x: get_embedding(x, engine=embedding_model))

# Save embeddings to database
new_df = df[['id','embedding_title','embedding_abstract']]
stored = save_embeddings_to_supabase(supabase=supabase, df=new_df)
```

To illustrate the application of these methods, Tables 3 and 4 display examples of the embedding vectors generated from the cleaned title and abstract, respectively.

Table 3. Example of embedding vectors of a cleaned title.

Title Cleaned	Embedding Title
machin learn market	$(-0.04582874, -0.00920587, 0.0051676217, \ldots, -0.019451953)$ The vector dimension is 1536.

Table 4. Example of embedding vectors of a cleaned abstract.

Abstract Cleaned	Embedding Abstract
widespread impact artifici intellig machin learn ml mani segment societi felt strong market field despit shortfal ml offer varieti potenti benefit includ opportun appli robust method general scientif discoveri tri reduc shortfal monograph goal provid market...	$(-0.024209673, -0.011041806, 0.02789942, \ldots, -0.022495108)$ The vector dimension is 1536

2.5. Topic Modeling Using Non-Negative Matrix Factorization

Topic extraction is an effective unsupervised data extraction technique for exploring the relationship between text subtexts. Although there are many different approaches, this paper focuses on NMF. The next stage in our methodological framework involved the application of NMF to discover thematic topics within the clean abstracts. The following steps defined this process:

1. Data preparation: We retrieved the cleaned text data from our Supabase database to ensure a robust analysis foundation. The first row of this database is shown in Table 5.
2. Word count analysis and common word identification: The initial analysis focused on the distribution of word usage in the abstract, identifying the most frequently occurring words in the dataset to guide the topic modeling process. The mean word value for abstract paper is approximately 120.59, and the mean standard deviation is approximately 44.63, indicating a wide range of values. The 25th percentile (Q1) is 94 words, the average (Q2) is 109.5 words, and the 75th percentile (Q3) is 142.75 words. The top words by frequency, among all papers after processing the text, are 'market', 'learn', 'data', 'machine', and 'use' (Figure 4).
3. Vectorization of text: The NMF is an unsupervised technique, so there is no labeling of the topics in which the model is trained. It decomposes (or factorizes) high-dimensional vectors into low-dimensional representations. These lower-dimensional vectors are not negative, which means that their coefficients are also not negative.

In our example, high-dimensional vectors are tf-idf weights [16], but they can be anything, including word vectors or a simple number of raw words. By converting the abstract text of each document into numerical form, we can use it to create functions (Listing 3). After processing the documents, we have a little more than 1553 unique words, so we set max-features to include only the top 750 terms per term frequency in all documents to further reduce the features.

4. Model fitting and topic extraction: The only parameter required is the number of components, i.e., the number of topics that we want (Listing 4). From this point on, the model is run to obtain the topics. Table 6 shows the three topics extracted with the NMF algorithm.

5. Topic assignment and dataset compilation: Documents were systematically assigned to the extracted topics and a comprehensive dataset encapsulating these associations was compiled.

6. Output documentation: We meticulously documented the results of the analysis for subsequent review and reference, ensuring that findings were accessible and well organized.

Table 5. First row of our database.

| Year	Cited	Title	Abstract
2022	20	algorithm bias machine learning-based market model	articl introduc algorithm bias machin learn ml base market model dramat growth algorithm decis make continu gain momentum market research stream inadequ despit devast asymmetr oppress impact algorithm bias various custom group void studi present framework identifi sourc algorithm bias market draw microfound dynam capabl use systemat literatur review indepth interview ml profession studi primari dimens design bias contextu bias applic bias correspond subdimens model data method cultur social person product price place promot synthes divers perspect use theori practic propos framework build dynam algorithm manag capabl tackl algorithm bias ml-base market decis make
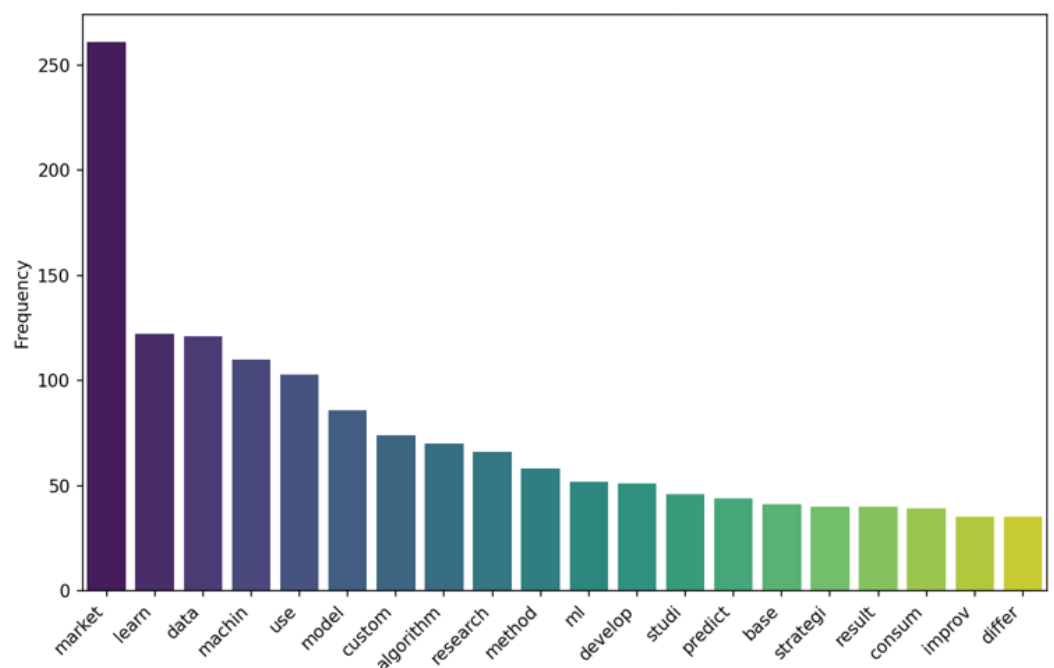
Figure 4. Top 20 most frequent words in abstracts of ML in marketing papers.

Listing 3. Python code for TF-IDF vectorization of text data.

```
1  texts = df[field_name]
2  texts = texts.str.split()
3
```

```
4   tfidf_vectorizer = TfidfVectorizer(
5       min_df=6,
6       max_df=0.85,
7       max_features=750,
8       ngram_range=(1,2),
9       preprocessor=' '.join
10  )
11
12  tfidf = tfidf_vectorizer.fit_transform(texts)
```

Listing 4. Python code for fitting NMF model.

```
1   nmf = NMF(
2       n_components=3,
3       init='nndsvd'
4   ).fit(tfidf)
```

Table 6. Extraction of topics using the NMF algorithm.

No.	Topics
1	data brand media market strategi custom develop onlin social busi base analysi environ new behavior use compani analyz platform inform
2	ml learn research review tool intellig applic method analyt machin mak decision focus manag artifici adopt potenti
3	model algorithm data use consum predict method respons custom research machin learn accuraci databas direct campaign network classif decis perform

As shown in Table 7, the documents were mapped into topics and a dataset with these associations was compiled to support the analysis.

Table 7. Topic assignment and dataset compilation.

Abstract	Topic No.	Topic
The widespread impacts of artificial intelligence (AI) and machine learning (ML) in many segments of society have not yet been felt strongly in the marketing field. Despite such shortfall, ML offers a variety of potential benefits, ...	2	ml learn research review tool intellig applic method analyt machin mak decision focus manag artifici adopt potenti
This article introduces algorithmic bias in machine learning (ML) based marketing models. Although the dramatic growth of algorithmic decision making continues to gain momentum in marketing, research in this stream is still inadequate despite the devastating, asymmetric and oppressive impacts of algorithmic bias on various customer groups...	3	model algorithm data use consum predict method respons custom research machin learn accuraci databas direct campaign network classif decis perform

2.6. Comparative Analysis Using K-Means Clustering

In order to extend the previous analysis and verify the findings derived from the NMF algorithm, we employed k-means clustering on the vector embeddings. This method facilitated the identification of natural groupings within our data. The k-means algorithm was executed through the following procedural steps:

1. Data retrieval and preparation: Initially, we accessed the vector embeddings stored in the Supabase database, preparing them for subsequent analysis (Listing 5).
2. Matrix formation and k-means clustering: A matrix was constructed from these embeddings, upon which k-means clustering was applied. In order to construct the matrix, string embeddings were transformed into numpy arrays (Listing 6).

3. Cluster assignment and analysis output: Documents were systematically assigned to their respective clusters. The clustering results were then summarized to reflect the natural groupings and their characteristics (Listing 7).

4. Data export: The dataset, inclusive of cluster assignments, was exported for extended analysis and future reference (Listing 8).

Listing 5. Python code for creating a Supabase client and loading data.

```
1  # Create client supabase
2  supabase = create_supabase_client()
3  df = get_data_to_supabase(supabase=supabase,
4                            name_table='web_embeddings')
```

Listing 6. Python code for converting string embeddings to numpy arrays.

```
1  # Field to work
2  field_name = 'embedding_abstract'
3
4  # Convert string to numpy array
5  df[field_name] = df[field_name].apply(literal_eval).apply(np.array)
6  matrix = np.vstack(df[field_name].values)
```

Listing 7. Python code for KMeans clustering.

```
1  n_clusters = 3
2
3  kmeans = KMeans(n_clusters=n_clusters,
4                  init="k-means++",
5                  random_state=42,
6                  verbose=0,
7                  n_init="auto")
8  kmeans.fit(matrix)
9  labels = kmeans.labels_
10 df["cluster"] = labels
```

Listing 8. Python code for printing KMeans clustering results and saving to CSV.

```
1  print(f"Number of clusters: {n_clusters} ,
2          Inertia: {kmeans.inertia_}\n")
3
4  df.to_csv("csv_Results_Clustering_KMEANS3.csv",
5            sep="#",
6            columns=['id', 'cluster'])
```

This methodological step allowed for an additional layer of insight into the dataset's thematic organization. The application of k-means clustering provided a comparative perspective that validated and enhanced the thematic structures identified by the NMF. Using the cleaned abstract data for an example involving three clusters, we obtained the following results: number of clusters = 3, inertia = 7.4452. Then, we assigned each document to a cluster.

2.7. Evaluation of Topic Modeling and Clustering Results

As a way to conclude the analysis, a comprehensive evaluation of the outcomes derived from the NMF algorithm and the clustering of k-means was conducted. The evaluation process included the following steps: (i) comparative framework setup and iterative trials—we established a comparative framework to systematically evaluate the two methodologies; (ii) optimal category determination and qualitative and quantitative analysis; and (iii) consolidation of findings, which allowed us to confirm the thematic coherence across the dataset (Figure 5).
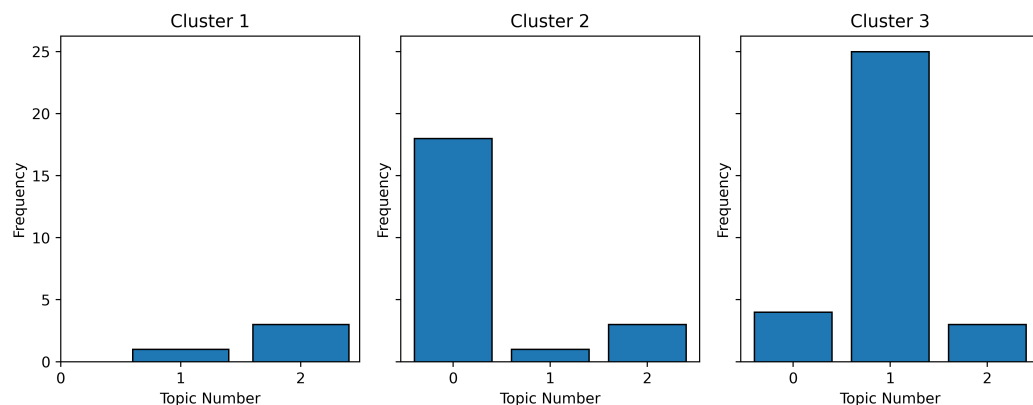
Figure 5. The best result is obtained with 3 clusters and 3 topics.

3. Cluster 1: Machine Learning Algorithms in Marketing

The first cluster includes four research articles focusing on the application of ML algorithms within marketing. In the first article, El Koufi et al. [17] explores precision marketing through big data and ML, with a specific focus on Morocco. This research introduces an algorithm designed to forecast potential clients by employing big data analysis and ML techniques. The study comprises five key steps: (i) data selection and understanding, using a dataset from Attawfiq Micro-Finance, a Moroccan bank—the dataset comprises 6000 records and 14 variables, including customer demographic and behavioral information spanning from 2018 to 2021; (ii) data cleaning and filtering procedures are conducted to eliminate missing values, outliers, and noise, thereby favoring dataset dimensionality and computational efficiency; (iii) feature selection becomes imperative for extracting critical features and eliminating non-significant attributes from the dataset—univariate selection followed by the grid search method is employed to determine optimal parameters; (iv) various ML algorithms, including XGBoost, random forest, gradient boosting, support vector machines, logistic regression, and decision trees, are implemented; and (v) evaluation of results, which employs a confusion matrix to assess predictive model performance—metrics such as precision, recall, accuracy, and f1 are utilized, alongside a detailed description of true positives, true negatives, false positives, and false negatives. The study concludes that gradient boosting outperforms other ML methods in terms of accuracy, f1, recall, precision, and cross-validation score.

In the second article, Saengthongrattanachot et al. [18] investigate the performance of ML techniques on bank marketing datasets. This study encompasses four primary steps: (i) the dataset characteristics are understood—the dataset details direct marketing campaigns of a Portuguese banking institution and comprises 36,548 records, with twenty input attributes and one output attribute (yes/no) indicating term deposit subscription; (ii) data preparation, which involves addressing class imbalance through subsampling of the majority class, resulting in a balanced dataset of 9280 records; (iii) modeling, which employs k-fold cross-validation to construct predictive models, utilizing algorithms such as J48, random forest, naive Bayes, and k-nearest neighbor—the models are initially built with all twenty input attributes, followed by models constructed with the five most relevant attributes identified by J48; and (iv) model evaluation, which employs criteria such as accuracy, precision, recall, and f1. The study concludes that reducing the attribute count to five enhances predictive model simplicity without significant accuracy loss.

In the third article, Zhang [19] focuses on predicting consumer loan processing utilizing ML algorithms and extracted characteristic data. The next steps are followed. (i) The study treats loan prediction as a classification problem, utilizing a neural network, a support vector machines algorithm, and a random forest algorithm; (ii) the support vector machine, known for its powerful classification method, seeks to find the most appropriate hyperplane as a decision surface by training sample data—the study highlights the use of kernel functions to map to a high-dimensional feature space; and (iii) the random forest

algorithm constructs a forest of decision trees independently, with the final classification determined by voting. The study's results, based on a dataset of 400 groups, reveal that random forest exhibits the highest accuracy, averaging over 94% and identifying elderly customers as more responsive to subscribing to term deposits.

Lastly, Sarker [20] integrates data mining algorithms into a bank's database system to predict customer behavior and improve marketing strategies. This study utilizes unsupervised ML algorithms—such as k-means clustering—as well as supervised algorithms—including multi-class decision forest, multi-class logistic regression, etc. The study emphasizes the successful prediction of customer behavior regarding term deposit subscriptions, particularly highlighting the efficacy of the multi-class decision jungle algorithm. These algorithms enable targeted marketing efforts towards customers likely to accept term deposits, thereby enhancing subscription rates. All in all, these articles collectively show the effectiveness of ML algorithms in enhancing marketing strategies within the banking sector.

4. Cluster 2: Application of Big Data and ML in Marketing and Consumer Behavior

The second cluster consists of twenty-two articles that analyze the use of big data and ML in marketing and their impact on consumer behavior. The main objective of this study is to determine the potential of these technologies in marketing processes, as well as the challenges and opportunities that marketers face in connecting with consumers. Marketing is a dynamic and constantly changing business function. To achieve its goal of delivering the right product to the right customers at the right time, it must manage information systematically, accurately, and in a timely manner in regards to the market, products, and customers, as well as the business environment [21]. The adoption of these technologies in marketing shows their full potential in different aspects:

- Social networks: Big data helps gather real-time consumer info. ML is used to analyze reviews and spot fake ones or unusual patterns.
- Understanding consumer behavior: Big data and fuzzy support vector machines show what consumers prefer when searching for info or comparing products. This is handy for launching new products where people need to pick what suits them best. Precision marketing, with big data, targets customers accurately with tailored offers [21].
- Advertising: Big data suggests TV/Internet content and predicts what people will watch and buy, making ads more accurate. It also helps in creative decision making and predicts sales from text, audio, and video data.
- Targeted digital promotions: Companies boost email campaigns by setting clear goals, personalizing content, and tracking engagement [22].
- Direct marketing: Predicting customer behavior helps tailor messages and improve campaign effectiveness [23]. ML helps companies understand customers better and run more targeted campaigns.
- Customer experience: Technology enhances marketing value and creates better customer experiences. Meeting customer needs and expectations requires using technology to understand behavior and assess marketing impact. Big data aids in decision making and understanding consumer behavior in real time.
- E-commerce: By using machine learning models to search for similar items on competitive online platforms, sellers can easily compare prices and offers to optimize their own marketing strategy. This can help sellers attract new customers and increase sales by providing better value propositions [24].
- Dynamic prices: The potential of automated pricing information and targeting marketing messages [25].
- Brand personality: Applied to consumer data, firms can assess how consumers perceive brand personality and study the effects of brand–consumer congruence in personality on social media [26].

All in all, the combination of big data and ML allows for enhancing customer relationship management, and for developing effective marketing strategies. In that way,

businesses can understand customer needs, segment audiences, predict behavior, and personalize marketing efforts. Big data also plays a key role in identifying target audiences for advertising campaigns, ensuring maximum impact and engagement. Table 8 displays several cases of companies that are applying ML and AI in their marketing and customer management processes.

Table 8. Examples of use of ML and AI for enhancing customer relationship management.

Company Name	Industry	Objective	Strategy	Results	Source
Starbucks	Retail	Increase customer loyalty and frequency of store visits	Use ML in their Starbucks Rewards mobile app to analyze purchasing history and customer preferences in order to provide tailored marketing offers and product recommendations	The company achieved a 16% year-over-year growth in its loyalty program participation	Kim and Ahn [27]
Cruzcampo	Brewing	Strengthen Cruzcampo's identity and connect with more people by tapping into the nostalgia and national pride linked to iconic Spanish artist Lola Flores	Use advanced AI and ML to digitally recreate Lola Flores' voice and image for the 'Acento' campaign	Throughout 2021, the campaign generated 614 million impressions, representing the number of times it reached viewers either on television or through social media	Palomo-Domínguez [28]
Netflix	Entertainment	Maximize viewer engagement and reduce churn by personalizing content recommendations	Use ML to analyze viewer data and preferences to enhance recommendation algorithms based on user viewing patterns and habits	Higher viewer satisfaction rate and retention, reducing churn by 75% and saving Netflix approximately USD 1 billion annually in customer retention	Ip [29]
Nike	Sportswear	Increase brand loyalty and customer engagement through innovative marketing campaigns	Implement ML and AI to analyze consumer behavior, tailor marketing content, and boost campaign effectiveness with innovative personalized designs	Customer engagement increased by 20% and social media followers by 30% across all platforms through precise targeting and personalized interactions	Sankar [30]

5. Cluster 3: Machine Learning Applied to Digital Marketing

The third cluster encompasses a collection of thirty-seven articles. These articles are focused on two primary themes: general applications of ML in digital marketing and innovation and technological advancements [31]. One significant application of ML in digital marketing is the identification and recommendation of key indicators within brand communities and public behavioral data. Techniques such as random forest, AdaBoost, and XGboost are utilized to distinguish between active and passive behaviors on Facebook pages of digital product brands and platforms. Additionally, text mining and ML methods such as content and sentiment analysis are employed to identify patterns within public and community behavior data of brands. The integration of AI further strengthens the processing of unstructured information, enhancing data quality through sampling methods like oversampling and undersampling. Furthermore, ML aids in analyzing consumer demand for information and uncovering hidden unstructured elements within data available on social networks, thereby improving marketing effectiveness. AI-based data analysis also helps in understanding the impact of community content on public engagement and how participatory posts enhance interactive experiences.

In digital marketing, ML finds many applications in analyzing the vast amount of data generated by social media and online platforms. NLP techniques enable AI systems to analyze human language across various mediums such as blogs, product reviews, and social media posts. This analytical capability is key in understanding and responding to customer perceptions online. Moreover, ML techniques aid in specific applications such as fraud detection, stock market prediction, customer relationship management, and automated generation of content summaries. Integration of ML tools with marketing strategies enhances performance by leveraging statistics and AI to make informed decisions. Arasu et al. [10] propose a framework comprising three key steps: (i) text mining, which involves the examination of unstructured information to identify meaningful patterns; (ii) use of ML tools to better understand and interpret data; and (iii) performing a data analysis that provides efficient marketing insights. Beyond social media, ML applications extend to various domains within digital marketing, including chatbots, advertisement refinement, email marketing, customer behavioral targeting, lead generation, and automated content

creation. These applications enable enhanced customization, deeper understanding of user behavior, and optimization of marketing strategies for improved efficiency [32].

Innovation and technological development in digital marketing are underscored by intelligent marketing strategies based on simulation with ML and mobile computing. Studies reveal significant improvements in company performance metrics, including efficiency gains of approximately 20% and revenue increases exceeding 30% compared to traditional strategies [33]. Additionally, advancements in ML facilitate automatic content classification in marketing, with neural networks demonstrating superior performance in multi-label classification of news articles compared to other algorithms like random forest and k-nearest neighbors [34]. Their ability to handle high-dimensional data without dimensionality reduction techniques highlights their efficacy in enhancing marketing strategies.

6. Future Trends

Marketing, as can be understood from the the integration of AI and ML, is reshaping innovation [35], moving towards a future dominated by personalized customer experiences that dynamically adapt in real time based on customers' interactions, contexts, and evolving preferences [36]. This is driven by the convergence of data proliferation, computational power, and algorithmic sophistication, allowing marketers to understand consumer behaviors and preferences with remarkable precision. On the consumer-facing side, it allows for (i) improvement of shopping fundamentals and (ii) enhancement of the consumption experience. On the business-facing side, it enables (i) better decision making and (ii) improved financial applications [37]. For example, in digital contexts, AI-driven features like "recommended for you" sections on platforms such as Amazon, Alibaba, and Netflix increase user engagement through tailored suggestions [38]. Another example is Airbnb, which uses ML algorithms to analyze user preferences and behaviors. This enables the platform to provide personalized accommodation choices, adjusting the results to the individual preferences of the clients [39]. In non-digital contexts, such as Starbucks' use of predictive analytics, the technology is applied to tailor marketing offers and product recommendations at an individual level [40]. By analyzing customer data, Starbucks can predict what products customers might be interested in, even before they themselves know, optimizing marketing efforts and increasing the likelihood of sales [41].

With the continuous refinement of ML models, there is not only a strong emphasis on real-time personalization of services and experiences. Predictive analytics are also experiencing a significant boost, enabling marketers to forecast market trends [42], customer behaviors [43], and campaign outcomes with greater accuracy [44]. This foresight will also facilitate more strategic decision making, resource allocation, and timely adjustments to marketing campaigns [45]. The integration of AI in marketing tools is expected to generate systems capable of autonomous operation, where marketing campaigns are automatically created, managed, and optimized without extensive human intervention [46]. These systems will identify opportunities, execute A/B testing, and refine strategies based on real-time data analysis. The proliferation of voice-assisted devices and conversational AI will transform search marketing [47] and customer service [48]. Marketing strategies will need to adapt to voice search optimization [49], and conversational AI will play a central role in customer interactions [50], providing personalized assistance and enhancing engagement. Blockchain technology will also allow the introduction of transparency into marketing practices, particularly in digital advertising. By facilitating secure and transparent transactions, blockchain can help mitigate fraud, ensure fair compensation in ad networks, and build trust among stakeholders [51]. Extended and virtual reality technologies are set to redefine experiential marketing, offering immersive experiences that engage customers in novel ways [52]. The Internet of things will provide marketers with a wealth of data from connected devices, enabling a better understanding of consumer habits and preferences [53]. This will open up new possibilities for personalized marketing, predictive maintenance, and enhanced customer experiences [54].

7. Conclusions

This paper combines the NMF and k-means algorithms to perform an analysis of the existing literature on the use of machine learning in marketing. The application of these two algorithms has facilitated the segmentation of the literature into three distinct clusters. Cluster 1 focuses on the integration of ML in the finance sector, highlighting advancements in customer segmentation and targeted marketing campaigns. Cluster 2 explores the synergy between big data and ML, emphasizing their role in enhancing direct marketing and customer relationship management. Cluster 3 focuses on the application of ML in digital marketing, particularly in optimizing content delivery and personalizing user experiences on digital platforms. The results derived from the analysis are based on the possibility that companies use ML tools in their marketing processes, as their applicability demonstrates their relevance in (i) gathering data from various marketing activities; (ii) processing, recording, and identifying patterns in the data; (iii) transforming data into valuable insights and opportunities; and (iv) predicting customer behavior and aiding in marketing decisions. The pursuit of more precise and adaptable marketing strategies has led to significant changes through the integration of AI, ML, and business analytics. This shift is driven by several factors. Firstly, today's consumers are more demanding and seek personalized products and services. Secondly, in our data-abundant world, businesses face both numerous opportunities and challenges. When used together, AI, ML, and data analytics provide the capabilities to handle large datasets, uncover patterns, and predict customer behavior with remarkable precision.

Furthermore, the use of ML significantly improves the marketing mix model. ML is used in product development, pricing, distribution, and promotion. Common applications include personalizing products and services, enhancing customer service, dynamic pricing, optimizing e-commerce, predicting conversion success, forecasting advance purchases, managing social media engagement, automating image analysis, optimizing search engines, and improving customer experiences, among others. Also, ML techniques are key for understanding customer feedback, enhancing products, and customer experiences. The importance of product reviews in e-commerce is also highlighted, underscoring their role in assisting customers with their buying decisions and guiding companies in their marketing strategies. All in all, the application of ML in marketing ensures that decisions are data-driven and aligned with business objectives. While our analysis provides valuable insights, it is based on existing literature and may be subject to publication bias. Additionally, the rapid evolution of ML technologies and marketing practices means that our findings may need to be updated periodically. Finally, although the study focused on the global analysis of marketing processes, we must highlight the opening of new lines of research focused on each of the elements of the marketing mix model. Therefore, future research should include specific variables to provide more details on the processes that apply to product development, pricing strategies, distribution channels, and brand communication.

Author Contributions: Conceptualization, A.F. and A.A.J.; methodology, A.F. and J.L.; software, A.F.; investigation, V.G., J.L. and C.O.; writing—original draft preparation, V.G. and J.L.; writing—review and editing, A.A.J. and C.O. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All required data are contained in the article.

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

References

1. Eid, R.; El-Gohary, H. The impact of E-marketing use on small business enterprises' marketing success. *Serv. Ind. J.* **2013**, *33*, 31–50. [CrossRef]
2. Ye, G.; Hudders, L.; De Jans, S.; De Veirman, M. The value of influencer marketing for business: A bibliometric analysis and managerial implications. *J. Advert.* **2021**, *50*, 160–178. [CrossRef]
3. Ngai, E.W.; Wu, Y. Machine learning in marketing: A literature review, conceptual framework, and research agenda. *J. Bus. Res.* **2022**, *145*, 35–48. [CrossRef]
4. Herhausen, D.; Bernritter, S.F.; Ngai, E.W.; Kumar, A.; Delen, D. Machine learning in marketing: Recent progress and future research directions. *J. Bus. Res.* **2024**, *170*, 114254. [CrossRef]
5. Akter, S.; Dwivedi, Y.K.; Sajib, S.; Biswas, K.; Bandara, R.J.; Michael, K. Algorithmic bias in machine learning-based marketing models. *J. Bus. Res.* **2022**, *144*, 201–216. [CrossRef]
6. Volkmar, G.; Fischer, P.M.; Reinecke, S. Artificial intelligence and machine learning: Exploring drivers, barriers, and future developments in marketing management. *J. Bus. Res.* **2022**, *149*, 599–614. [CrossRef]
7. Yazıcı, G.; Ozansoy Çadırcı, T. Creating meaningful insights from customer reviews: A methodological comparison of topic modeling algorithms and their use in marketing research. *J. Mark. Anal.* **2023**, 1–23. [CrossRef]
8. Sterne, J. *Artificial Intelligence for Marketing: Practical Applications*; John Wiley & Sons: Weinheim, Germany, 2017.
9. Zhang, L.; Gao, Q.; Li, T. Dynamic adaptation method of business process based on hierarchical feature model. *Information* **2021**, *12*, 362. [CrossRef]
10. Arasu, B.S.; Seelan, B.J.B.; Thamaraiselvan, N. A machine learning-based approach to enhancing social media marketing. *Comput. Electr. Eng.* **2020**, *86*, 106723. [CrossRef]
11. Bonab, H.; Aliannejadi, M.; Vardasbi, A.; Kanoulas, E.; Allan, J. Cross-market product recommendation. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Queensland, Australia, 1–5 November 2021; pp. 110–119.
12. Tirunillai, S.; Tellis, G.J. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *J. Mark. Res.* **2014**, *51*, 463–479. [CrossRef]
13. Kalinić, Z.; Marinković, V.; Kalinić, L.; Liébana-Cabanillas, F. Neural network modeling of consumer satisfaction in mobile commerce: An empirical analysis. *Expert Syst. Appl.* **2021**, *175*, 114803. [CrossRef]
14. Brei, V.A. Machine learning in marketing: Overview, learning strategies, applications, and future developments. *Found. Trends® Mark.* **2020**, *14*, 173–236. [CrossRef]
15. Guo, Y.T.; Li, Q.Q.; Liang, C.S. The rise of nonnegative matrix factorization: Algorithms and applications. *Inf. Syst.* **2024**, *123*, 102379. [CrossRef]
16. Kanakogi, K.; Washizaki, H.; Fukazawa, Y.; Ogata, S.; Okubo, T.; Kato, T.; Kanuka, H.; Hazeyama, A.; Yoshioka, N. Tracing cve vulnerability information to capec attack patterns using natural language processing techniques. *Information* **2021**, *12*, 298. [CrossRef]
17. El Koufi, N.; Belangour, A.; Sdiq, M. Research on precision marketing based on big data analysis and machine learning: Case study of Morocco. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 58–63.
18. Saengthongrattanachot, W.; Na-udom, A.; Rungrattanaubol, J. A comparison of machine learning techniques for classification in bank marketing data. *Thai J. Math.* **2022**, *20*, 157–168.
19. Zhang, M. Research on precision marketing based on consumer portrait from the perspective of machine learning. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 9408690. [CrossRef]
20. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef]
21. Miklosik, A.; Evans, N. Impact of big data and machine learning on digital transformation in marketing: A literature review. *IEEE Access* **2020**, *8*, 101284–101292. [CrossRef]
22. Ellickson, P.B.; Kar, W.; Reeder III, J.C. Estimating marketing component effects: Double machine learning from targeted digital promotions. *Mark. Sci.* **2023**, *42*, 704–728. [CrossRef]
23. Choi, Y.; Choi, J.W. Assessing the predictive performance of machine learning in direct marketing response. *Int. J.-Bus. Res. (IJEBR)* **2023**, *19*, 1–12.
24. Cherednichenko, O.; Ivashchenko, O.; Cibák, L.; Lincenyi, M. Item Matching Model in E-Commerce: How Users Benefit. *Econ. Cult.* **2023**, *20*, 77–90. [CrossRef]
25. Thirumuruganathan, S.; Al Emadi, N.; Jung, S.g.; Salminen, J.; Robillos, D.R.; Jansen, B.J. Will they take this offer? A machine learning price elasticity model for predicting upselling acceptance of premium airline seating. *Inf. Manag.* **2023**, *60*, 103759. [CrossRef]
26. Pamuksuz, U.; Yun, J.T.; Humphreys, A. A brand-new look at you: Predicting brand personality in social media networks with machine learning. *J. Interact. Mark.* **2021**, *56*, 1–15. [CrossRef]
27. Kim, K.; Ahn, S.J. Rewards that undermine customer loyalty? A motivational approach to loyalty programs. *Psychol. Mark.* **2017**, *34*, 842–852. [CrossRef]
28. Palomo-Domínguez, I. Del mito a la viralidad. El caso de la campaña de Cruzcampo que resucitó a Lola Flores. *aDResearch Rev. Int. Investig. Comun.* **2021**, *26*, 38–58.

29. Ip, K. Revolutionising content recommendation: The impact of AI in marketing. *J. AI Robot. Workplace Autom.* **2023**, *2*, 382–389.
30. Sankar, J.G. AI-driven marketing success stories: A case note of industry pioneers. In *AI-Driven Marketing Research and Data Analytics*; IGI Global: Hershey, PA, USA, 2024; pp. 48–66.
31. Chen, Y. Comparing content marketing strategies of digital brands using machine learning. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 57. [CrossRef]
32. Boddu, R.S.K.; Santoki, A.A.; Khurana, S.; Koli, P.V.; Rai, R.; Agrawal, A. An analysis to understand the role of machine learning, robotics and artificial intelligence in digital marketing. *Mater. Today Proc.* **2022**, *56*, 2288–2292. [CrossRef]
33. Mao, S.; Huang, R. Complexity construction of intelligent marketing strategy based on mobile computing and machine learning simulation environment. *Complexity* **2021**, *2021*, 9910834. [CrossRef]
34. Salminen, J.; Yoganathan, V.; Corporan, J.; Jansen, B.J.; Jung, S.G. Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type. *J. Bus. Res.* **2019**, *101*, 203–217. [CrossRef]
35. Chintalapati, S.; Pandey, S.K. Artificial intelligence in marketing: A systematic literature review. *Int. J. Mark. Res.* **2022**, *64*, 38–68. [CrossRef]
36. Haleem, A.; Javaid, M.; Qadri, M.A.; Singh, R.P.; Suman, R. Artificial intelligence (AI) applications for marketing: A literature-based study. *Int. J. Intell. Netw.* **2022**, *3*, 119–132. [CrossRef]
37. De Mauro, A.; Sestino, A.; Bacconi, A. Machine learning and artificial intelligence use in marketing: A general taxonomy. *Ital. J. Mark.* **2022**, *2022*, 439–457. [CrossRef]
38. Guidoum, S.; Saadi, E. Digital marketing in the age of AI: View global experiences. *Econ. Stud.* **2024**, *18*, 141–151.
39. Kaushal, S.; Mishra, D. Strategic implications of AI in contemporary business and Society. In Proceedings of the 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 15–16 March 2024; pp. 1469–1474.
40. Kumar, V.; Rajan, B.; Venkatesan, R.; Lecinski, J. Understanding the role of artificial intelligence in personalized engagement marketing. *Calif. Manag. Rev.* **2019**, *61*, 135–155. [CrossRef]
41. Kumar, D. The Use of Artificial Intelligence in Data Analysis and Business Intelligence. In *AI and Emotional Intelligence for Modern Business Management*; IGI Global: Hershey, PA, USA, 2023; pp. 311–320.
42. Vidhya, V.; Donthu, S.; Veeran, L.; Lakshmi, Y.S.; Yadav, B. The intersection of AI and consumer behavior: Predictive models in modern marketing. *Remit. Rev.* **2023**, *8*, 2410–2424.
43. Surendro, K. Predictive analytics for predicting customer behavior. In Proceedings of the 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, 13–15 March 2019; pp. 230–233.
44. Soni, V. Adopting generative AI in digital marketing campaigns: An empirical Study of drivers and barriers. *Sage Sci. Rev. Appl. Mach. Learn.* **2023**, *6*, 1–15.
45. Zhang, S.; Liao, P.; Ye, H.Q.; Zhou, Z. Dynamic marketing resource allocation with two-stage decisions. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 327–344. [CrossRef]
46. Gołąb-Andrzejak, E. The impact of generative AI and ChatGPT on creating digital advertising campaigns. *Cybern. Syst.* **2023**, 1–14. [CrossRef]
47. Klaus, P.; Zaichkowsky, J. AI voice bots: A services marketing research agenda. *J. Serv. Mark.* **2020**, *34*, 389–398. [CrossRef]
48. Skrebeca, J.; Kalniete, P.; Goldbergs, J.; Pitkevica, L.; Tihomirova, D.; Romanovs, A. Modern development trends of chatbots using artificial intelligence (AI). In Proceedings of the 2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), Riga, Latvia, 14–15 October 2021; pp. 1–6.
49. Runaite, D. How Will Voice Search Optimisation Aid or Limit Digital Marketing? An End-User Perspective. Ph.D. Thesis, National College of Ireland, Dublin, Ireland, 2021.
50. Israfilzade, K.; Sadili, N. Beyond interaction: Generative AI in conversational marketing-foundations, developments, and future directions. *J. Life Econ.* **2024**, *11*, 13–29. [CrossRef]
51. Peres, R.; Schreier, M.; Schweidel, D.A.; Sorescu, A. Blockchain meets marketing: Opportunities, threats, and avenues for future research. *Int. J. Res. Mark.* **2023**, *40*, 1–11. [CrossRef]
52. Eng, S. The new realities in digital marketing: Virtual reality, augmented reality, and extended reality. In *The SAGE Handbook of Digital Marketing*; SAGE Publications: New York, NY, USA, 2022; p. 276.
53. Sehgal, P.; Kumar, B.; Sharma, M.; Salameh, A.A.; Kumar, S.; Asha, P. Role of IoT in transformation of marketing: A quantitative study of opportunities and challenges. *Webology* **2022**, *18*, 5838–5849.
54. Baykal, B. Connected Consumers: Unleashing the Power of IoT for Personalized Marketing. In *Academic Research and Evaluations in Social Sciences—IV*; Bozgeyik, Y., Ed.; Özgür Publications: Gaziantep, Turkey, 2023; p. 69.