AI Marketing Tool

Cynthia Bachor

Abstract

Artificial intelligence (AI) has proven useful in a variety of applications ranging from facial recognition to automated customer service responses. However, while many businesses want to use AI to improve marketing, they lack the resources to implement AI in their marketing business process. The first stage in creating a marketing plan is to understand the target audience and the competition. In recent times large brands and/or companies have started to use the available competitive intelligence softwares' available to analyse their competitors, identify gaps in their strategies and discover growth opportunities for their business; Smaller businesses and start-ups with insufficient funding, however, do not have access to these resources. These businesses struggle to pinpoint their target market and decide how to strategically position themselves in the market using data due to this constraint.

This project aims to develop an AI marketing tool that helps SMEs and start-ups understand their specific businesses through competitive intelligence. For the scope of this project we will be focusing on the nootropics market using select brands and influencers to prove this concept, scrapping their digital footprints from posts on their websites, social pages like Instagram and twitter, Processing and further clustering and segmentation analysis for brand identification using pre-trained models. [1]

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Businesses typically have great internal visibility into their product metrics, brand positioning, sales, marketing, and strategies, but they still lack clear visibility on external factors, such as what their competitors are doing to help channel their resources in the right directions based on analysed data. Competitive intelligence (CI), a branch of knowledge management, aims to observe a company's external environment in order to obtain information helpful to its decision-making process. To identify and seize opportunities for creating a persistent competitive edge, many large organisations deploy software platforms for competitive intelligence. This technology is used by these organisations to make data-driven decisions regarding crucial business matters like marketing and strategic planning. In the past, CI relied on written materials like business reports that were accessible to the general public. The Internet has rapidly transformed in recent years into a very trustworthy source of information on the business environment.[1]

In today's business scenario, artificial intelligence finds applications in a variety of contexts. According to experts and professors, artificial intelligence will shape our civilization in the future. With the development of technology, the world has evolved into a web of interconnected networks. AI technology is a well-liked developing technology that enables businesses to monitor real-time data in order to analyse and swiftly address clients â needs. AI offers consumer insight into behaviour, which is essential for reeling in new customers and retaining them. AI inspires the next move of the customer and redefines the overall experience. In marketing, AI tools can be used;

- To deduce customer expectations,
- In chat-bots for customer service,
- For tools that model the potential outcomes of a new marketing campaign,
- As a recommender systems that assist managers in choosing content for online marketing, or Models that predict customer behaviour. [2] [3]

Artificial intelligence has several potential applications in competitive intelligence. Intelligent algorithms, for instance, can be trained to take information from thousands of sources, filter it for insights that are pertinent, look for patterns in the data, and forecast future outcomes. Since most data gathered from the web is unstructured, it can be slow for computers to understand. It takes a lot of human effort to read, extract, organise, and synthesise pertinent information from the digital footprints, websites, blogs, and social media platforms used by competitors, including Instagram and Twitter. The variability of the sources and the unconventional presentation schema contribute to the task's complexity. The uncertainty of the procedure is further influenced by variations in this data's applicability and dependability. Before employing web content, noise in the content must be handled as a serious problem. Numerous html tags, hashtags, emoticons, stopwords (like are, I, and they), and extraneous characters are all examples of noise in this context.

The proposed project is to develop an AI marketing tool that would help marketing departments at SMEs and start-up companies spot trends in their industry to better position themselves competitively. The following sections include a review of the literature in the use of AI for marketing, details of the methodology used to achieve this, the comparative analysis, results and a conclusion.

CHAPTER 2

Literature Review

In recent times, Artificial intelligence (AI) has been used in the development of software. We see AI applications in software development in a variety of use cases, including personalised shopping on ecommerce platforms, AI-powered assistants, administrative task automations, creating smart content, and personalised content on social media platforms. The use of artificial intelligence powered by machine learning technology allows these applications to become more efficient over time because this technology can learn continuously, analyse information at a rate that humans cannot, and potentially solve problems as they arise.

In contrast to artificial intelligence, which is about automating tasks by relying on "intelligent machines," software engineering is a creative and knowledge-intensive activity that typically involves human experts. As a result, using artificial intelligence in software engineering initially seems to be a contradiction in terms. However, upon deeper examination, it is possible for machines to support creative software engineering processes, provided that they are backed by self-optimizing algorithms.[4].

In contrast to the 20th century, which saw a dearth of intelligent system applications in marketing[5], recent years have seen tremendous improvements in the use of artificial intelligence in the sales and marketing sector. AI is currently being used in a variety of applications, including automatic fact-checking in journalism, powering chat-bots that engage with customers on websites, market for casting and automation.

Table 2.1: A literature review on artificial intelligence in marketing

Author	Title	Overview and findings
Sanjeev Verma et al.[6]	Artificial intelligence in marketing: Systematic review and future research direction	The study adds to the literature by highlighting the key theoretical issues and variables that explore AI approaches and the most popular marketing techniques, yielding three major clusters (brand role, components of interaction, and results of interaction).
Shuliang Li [7]	AgentStra: An Internet-based multi-agent intelligent system for strategic decision-making	This study looks at how an Internet-based multi-agent intelligent system can enhance the processes of creating marketing strategies, creating competitive strategies, and formulating relevant IT/IS/e-commerce strategies.
Francisco J et al.[8]	Artificial intelligence-based systems applied in industrial marketing: a historical overview, current, and future insights	This is a literature review of artificial intelligence-based systems applied to marketing, covering a time period of several decades (from the 1970s to the present day), with special focus on applications to industrial marketing.
Niladri Syama et al. [9]	Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice	The authors discuss the impact of machine learning and AI on sales processes such as prospecting, pre-approach, approach, presentation, objection handling, close, and follow-up.
Jose Ramon Saura et al.[10]	Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research	The primary uses of AI-based CRMs in B2B digital marketing strategies are addressed in this study, and the findings define the various CRM types and their typologies as well as examine the key strategies and applications of AI-based CRMs in B2B digital marketing.

2.1 Trends in the AI-driven marketing industry

2.1.1 The use of AI in Strategic Planning

Traditionally, strategic teams in organisations frequently make assumptions about what will happen in the future based on data that is out of date, incomplete, inaccurate, or biased, resulting in a reliance on gut feeling and experience, which doesn't always count for much in a changing world with no precedent. This results in strategic plans that are incorrect and potentially harmful to the organisation. Furthermore, strategic planning is typically an annual, if not twice-yearly, event that does not always account for market changes that can occur unexpectedly (like Covid-19). Events that have an impact on strategy occur all the time, even if teams are unaware of them and most times it is too late to incorporate them the following year. The continuous evolution of AI technology affects the future of strategic planning, [11] for example, some of the most relevant problems, such as the alignment of strategic orientation with market potential [12], are solved nowadays using AI solutions. In this way, implementers of AI-based marketing solutions have noted improvements in business model decisions [13], as organizations have access to information on time and can make quick decisions on new product development [14], sales management [15], advertising [16] etc. as the market changes.

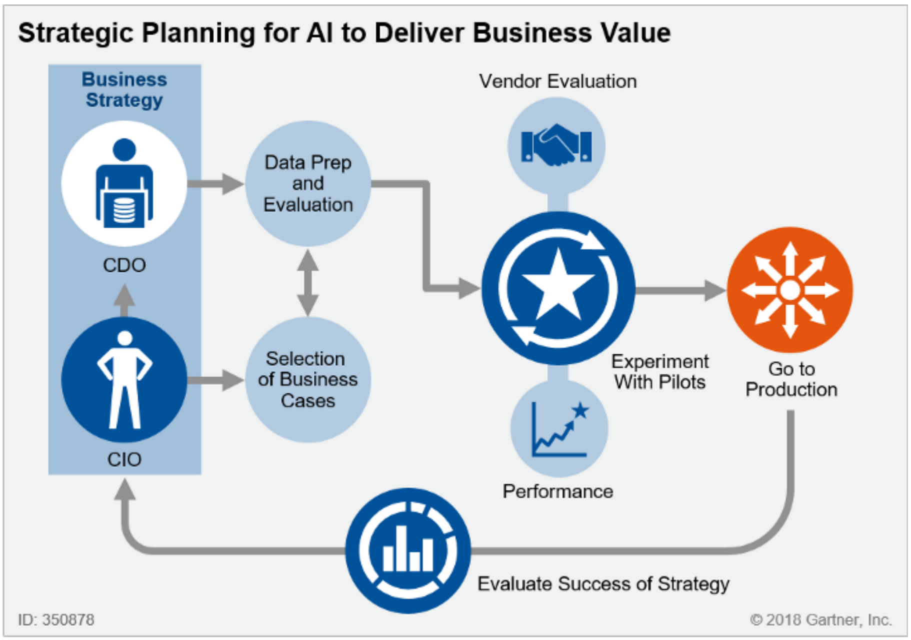

Figure 2.1: Some key elements of a strategic plan for AI

Companies are also been strategic about their service offerings, utilizing AI to drive innovation and enable increased efficiency, changing the nature of the workplace and how tasks are distributed, for instance with tasks that rely on experiential learning (e.g., engagement with customers, types of cases resolved etc), IT companies like Microsoft and Google use Artificial Intelligence in their support ticket routing systems to increase efficiency with managing support tickets.

2.1.2 The use of AI in Interactive media

With information stored in media formats including text, image, and video, internet social media and mobile devices have dramatically expanded interactions between businesses and their customers. Based on this media information, businesses must comprehend the demands, preferences, and brand positioning insights of their customers. A significant area of competitiveness is in the creation of interesting and educational material to promote product uptake. In these interactive environments, machine learning-powered AI technologies are used because of their superior ability to provide insights and prescribe remedies.[17]

Although human communication is naturally multi-modal, incorporating speech, gesture, and body movement, digital media create new and emergent conditions, structures, and grammars for mediated expression and communication (syntactic, aural, visual, kinetic and haptic both in 2D and in 3D environments). As such, the algorithmic turn in contemporary communication represents a new frontier in theorization, production, and evaluation of interactive media.

The ecosystem of smart speakers, smart TVs, and their computational assistants (such as Siri, Alexa and Google Assistant) today offers a window into the world of cutting-edge developing technologies and the progressive removal of mediation. Currently, non-linear aural narrative capabilities are supported by artificial intelligence speakers (e.g. Select a Story, Amazon Alexa).[18]

2.1.3 The use of AI in Product Management

Product managers can benefit from artificial intelligence in a number of ways to speed up the product development process, enhance consumer experience and increase product adoption. The use of AI is now been used for automating some stages of the product development life cycle process like competitor and market analysis, creating a feedback pipeline to automatically gather and analysis product feedback to gain product insights and validate ideas so product managers may have more time to devote to strategic planning and other elements of their work.

The ability of a product manager to understand the full user journey is another crucial component of product development. Digital technology has made it possible for businesses to gather fine-grained data from client touchpoints, enabling them to see continuous decision journeys with feedback loops from the perspective of specific customers. Businesses gain a lot from putting together a complete picture of a customer’s journey since it enables them to monitor and direct the customer while providing the appropriate information, service, and promotion at the appropriate time and place. When a lead is created, a thorough understanding of the customer helps with the creation of a programme to lead the customer through the purchasing process while modifying strategies to boost conversion. Following conversion, the interaction continues to ensure a positive customer experience and maximum lifetime value. Machine learning methods are crucial in managing consumer decision journeys, bringing marketing effectiveness to a new level.[17]

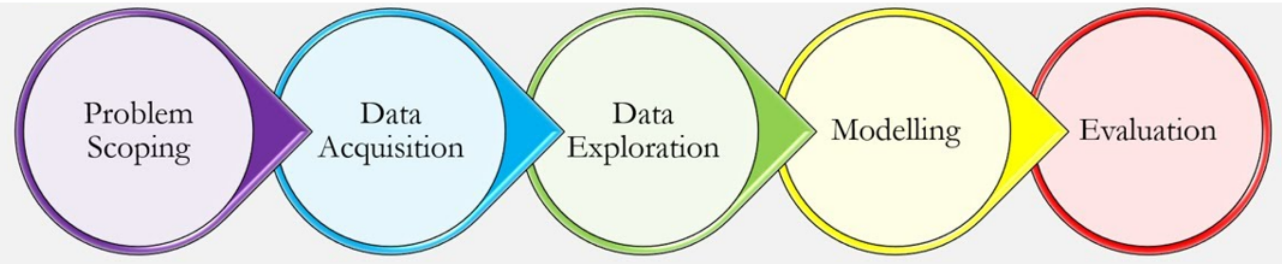

Figure 2.2: Managing AI Products lifecycle

2.1.4 The use of AI in Price management

In reality, companies frequently base their pricing decisions on those of their competitors. Such a straightforward approach will frequently result in pricing that are not optimal in terms of profit and, in certain situations, even in terms of revenue (e.g., for price-insensitive products a larger profit might be better). A common issue with competition-based pricing is that it can lead to "price wars," in which competing brands match one other’s prices, resulting in either a cycling pattern or a downward spiral if both competitors keep on pricing below the other vendor[19]

Pricing entails the consideration of numerous factors in the finalisation of a price, and it is a calculation-intensive job. The complexity of pricing is increased by real-time price variation based on fluctuating demand. In a real-time scenario, an artificial intelligence-based multiarmed bandit algorithm can dynamically adjust the price [20]. In their paper Optimal pricing in e-commerce based on sparse and noisy data, Josef Bauer and Dietmar Jannach [21] proposed a novel machine-learning based algorithm for estimating optimal prices based on Bayesian inference combined with bootstrap-based confidence estimation and kernel regression.

2.1.5 The use of AI for Personalization and targeting

Marketing is becoming more personalised. Machine learning methods are propelling large-scale context-dependent personalization and targeting to a new level in markets where rich data is commonplace and digital channels make personalised offerings easy to deliver. [22] Segmentation is becoming finer-grained, with hundreds of precisely carved-out microsegments replacing a few coarsely defined large ones. With continued refinement, each consumer becomes a distinct segment, receiving tailored offers based on her individual profile. Taking it a step further, preferences and needs are fluid, and opportunities are transient. Thus, effective targeting necessitates not only matching the right offer to the right consumer, but also delivering it at the right time and in the right context. A large portion of personalization and context-dependent targeting is powered by powerful machine learning algorithms and these practices also drive the methods' rapid evolution.[17]

Businesses like Instagram currently employs machine learning algorithms to identify the types of material that customers are interested in and predict similar content (e.g; when you search, engage, share or save contents on restaurants in London, Instagram shows you more posts of restaurants in London in the days that follow). Giving customers personalized content .

2.2 Work done so far

The Project team began the first phase of this project, where we chose to focus the use-case for this project on Nootropics, a health supplement market and the main companies that has offerings are mentioned in this article.[23]. In the first phase of the project we worked on the data gathering using tools and libraries like Instaloader, Facebook Scrapper and Tweepy for data extraction from Twitter, Facebook and Instagram and also designed a custom made scrapper to get data from websites, Pre-processed the data by removing html tags, hashtags, emoji, and irrelevant characters and conducted further analysis by extracting keywords, checking for similarities between Nootropics brands and influencers as well as visualization using Heatmaps and Wordcloud.

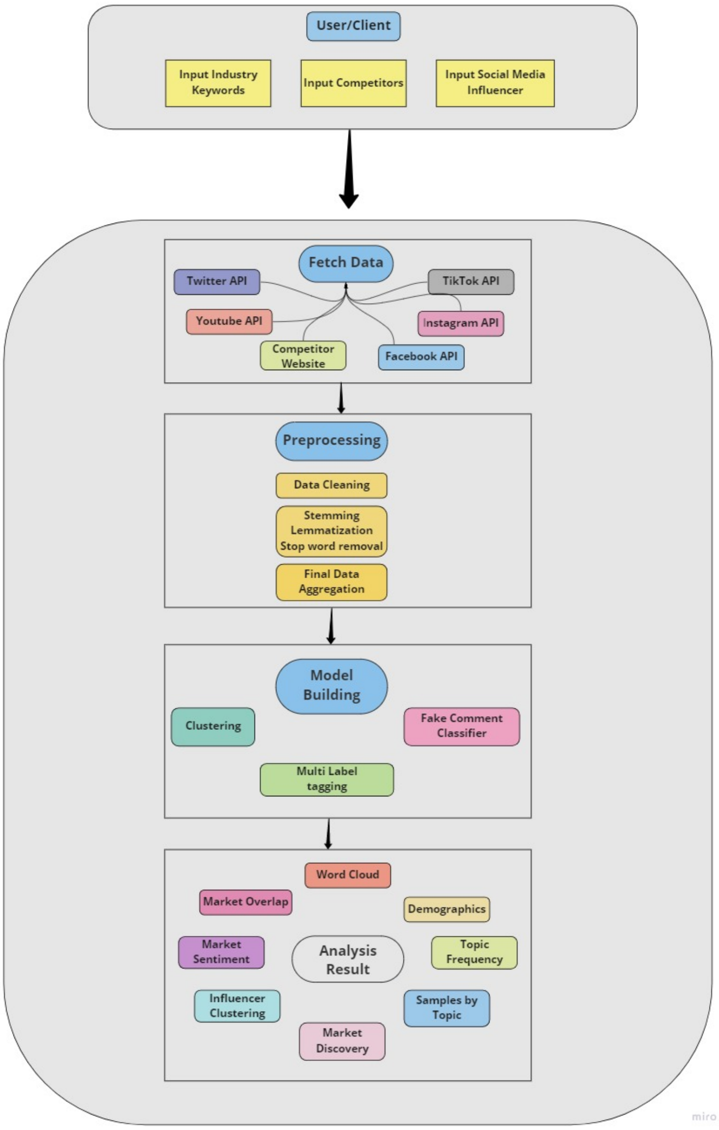

Figure 2.3: System Architecture used

CHAPTER 3

Methodology

The approach we have taken for this project is to look at brand data to get as much information as possible. We are considering the possibilities of identifying a brand's position by looking at data obtained from their digital footprints, and we are considering clustering as a way to identify market segments using clustering. This is in light of the various approaches used so far by individual academics to implement artificial intelligence in marketing.

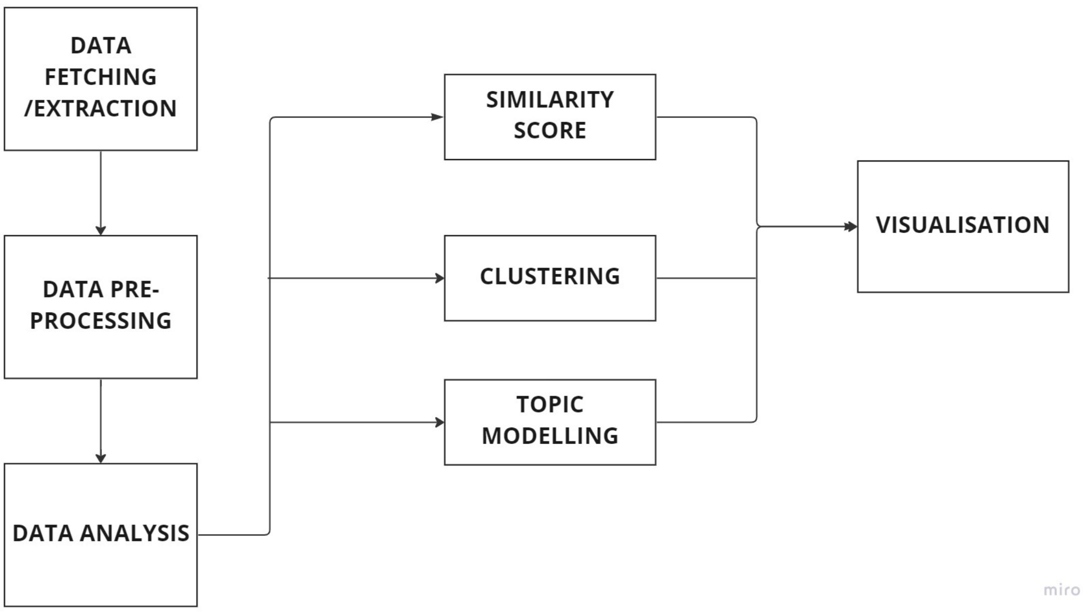

Figure 3.1: Implementation Process Flow

The project's work thus far has been based on the process flow depicted in Figure 3.1 above. Although no website, front-end design, or app has been created, the steps have been implemented in the backend.

3.1 Data Fetching

[24] To gather data for this project, we chose the nootropics market, which is known for products that improve cognition, memory, and learning. We have a set of brands in this industry that we have chosen for the purpose of this project, as well as a list of influencers, some of which focus on health and similar sectors, as well as other neutral influencers to test our hypothesis.

Name	Website	Instagram	Twitter	Facebook
Neat Nutrition	neat-nutrition.com	neat_nutrition	neat_nutrition	neatnutrition
Bulk	bulk.com	Bulk	Bulkofficial	Bulk
Live Innermost	liveinnermost.com	liveinnermost	liveinnermost	innermost
The Nue co	thenueco.com	thenue_co	thenue_co	thenueco
Indisupplement	indisupplements.com	indisupplements		indisupplements
Form Nutrition	formnutrition.com	formnutrition	formnutrition	formnutrition
Puresport	puresport.co	puresport	puresportcbd	puresportclubs
Medterra	medterracbd.co.uk	medterra.international	medterracbd	medterra.international
Neurohacker	neurohacker.com	neurohacker	theneurohacker	neurohacker collective
Motion Nutrition	motionnutrition.com	motionnutrition	MotionNutrition	motionnutrition

Table 3.1: List of Select Influencers and their social handles

S/N	Instagram Handle
1	thebodycoach
2	aaroncgshore
3	lucymeck1
4	danosborneofficial
5	charlottedawsy
6	slimmingworld
7	chessieking
8	thefitnesschef_
9	aliceliveing
10	ohpolly
11	oliviadbowen
12	chloe.khan
13	ini.helen
14	jesswright77
15	_jackfowler_
16	brown.elle
17	jessica_rose_uk
18	mac_griffiths_
19	gabbydawnallen
20	jesshunt2
21	katiepiper_
22	korisampson
23	itsalwayshana
24	sylvijaa
25	adamcollard
26	jamesgshore
27	mattdoesfitness
28	rogersnipes
29	courtneydblack
30	jamessmithpt

3.1.1 Trafilatura Web Scrapper

We have developed a web scraping function utilising the Trafilatura module, [25]which functions as a text exploration and retrieval tool that downloads, parses, and scrapes web page data in real time, to extract data from websites of each brand. It scans a website for text, finds it, and analyses it. The extractor keeps some text formatting and page structure but focuses on the metadata, major body text, and comments. The URLs of the websites whose text we need to extract are parsed as inputs, producing plain text outputs with all the textual information from the websites.

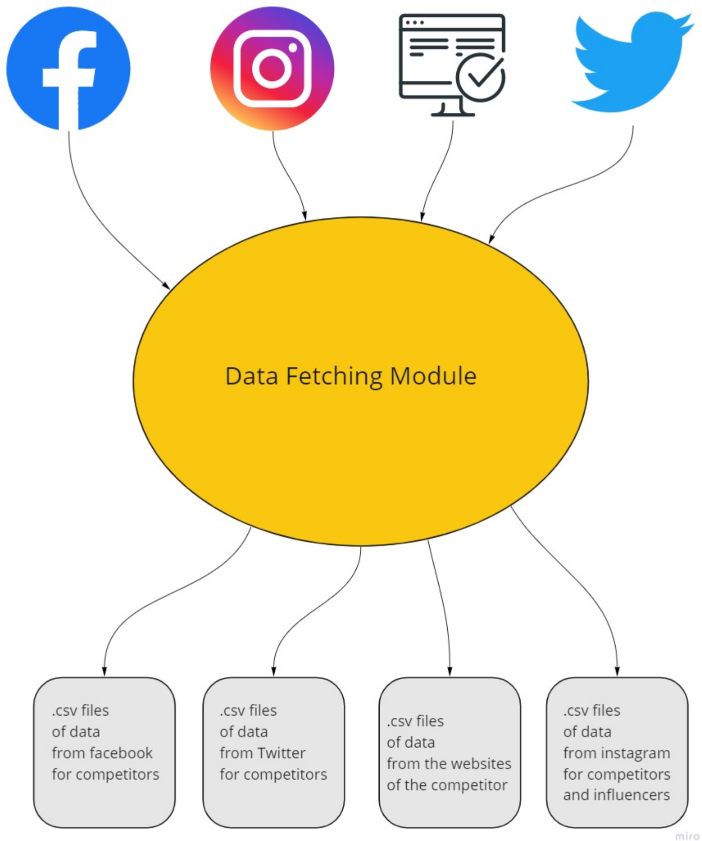

Figure 3.2: Module for Data Fetching

3.1.2 Instaloader API

A free and open-source terminal tool called Instaloader acts as a scraper for posts from both public and private accounts, stories, IGTV, comments on posts, profile information, and story highlights. For this project, we used this module to build functions that would pull captions, comments, and the total amount of likes from each brand's and influencer's Instagram posts. To complete this process, the usernames of these brands and influencers are essential.

3.1.3 Tweepy API

Furthermore, we have gathered information from tweets posted by brands to their tweeter accounts and by social influencers. We utilised the open source Python package Tweepy to develop an extractor with quick access to the Twitter API. With the help of this module, we have created a function to pull plain text information from brand and social influencer twitter profiles. To complete this process, the usernames of these brands and influencers are needed.

3.1.4 FacebookScrapper API

From each competitor's and influencer's Facebook page, plain text data was extracted using the Facebook scraper. For this purpose, the scraper functions as an open source API.

3.2 Data Pre-processing

For the pre-processing module ([26]), a data integrator and data cleaner functions were written.

3.2.1 Data Integrator

All of the information gathered during the previous data collecting step was combined by the data integrator into two csv files. While the information gathered from influencers is put into a file, the information gathered from competitors' platforms is all put into a.csv competitor file. This data integrator's output is a csv file with the following format.

The code also changes the names of the initial files' columns to [competitor_name] and [Text] and swaps out the lengthy generic names for just the competitor's names.

Table 3.2: Competitor file output format after merging

Competitor_name	**Platform**	**Text**
CompetitorA	Facebook	Sample text.
CompetitorB	Twitter	Sample text.
CompetitorC	Twitter	Sample text.
CompetitorD	Facebook	Sample text.
CompetitorE	Facebook	Sample text.

Table 3.3: Influencer file output format after merging

Influencer_name	**Platform**	**Text**
InfluencerA	Instagram	This is an example text.
InfluencerB	Instagram	Sample text.
InfluencerC	Instagram	Sample text.
InfluencerD	Instagram	Sample text.
InfluencerE	Instagram	Sample text.

3.2.2 Data Cleaner

The text that was imported from the data integrator has all special characters and emojis removed by the functions in this file. Spaces have been used in place of characters like?,!,,@, and other superfluous ones from the imported texts thanks to the regex (re) function. Since hashtags are commonly used in text data, they have been eliminated and will be treated as a different type of data for subsequent analysis in order to prevent repeated tags from showing up as words in the data.
There are four outputs from this phase,

- CSV file of competitors excluding special characters, emojis, HTML tags, and stopwords
- CSV file of influencers excluding special characters, emojis, HTML tags, stopwords
- CSV file of competitors excluding emojis, HTML tags, stopwords, special characters except fullstops and question marks to keep the context of the sentences.
- CSV file of influencers excluding emojis, HTML tags, stopwords, special characters except fullstops and question marks to keep the context of the sentences.

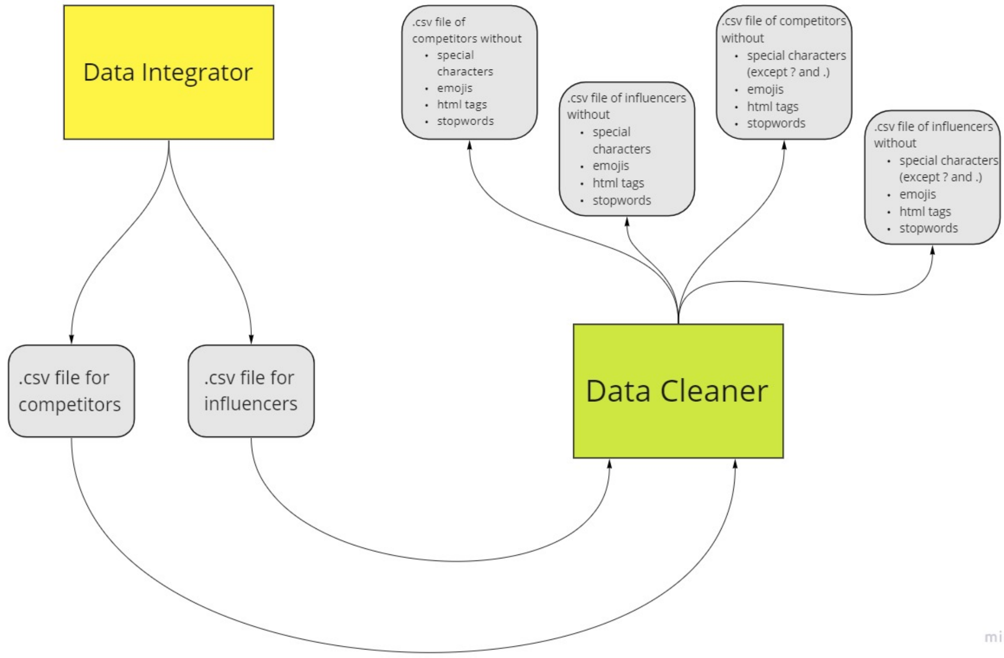

Figure 3.3: Flow of the Data Integrator and Cleaning Module

3.3 Analysis

After processing and cleaning up the data, we began the analysis on this data to help draw conclusions in line with the goal of this project which is to identify brand positioning and keywords, customer segments using influencers, and similarities between each brands and each brand and influencer to determine which influencer is better suited for the brand. All the above mentioned have been approached using different modules in trying to see whats best suited for this usecase.

3.3.1 keyword Analysis

The capability of this instrument to extract the keywords employed by the brand for their digital marketing is one of its requirements. We use the.csv file containing the data excluding special characters, emojis, HTML tags, and stopwords from the data obtained once cleaning is completed in the previous module. We explored three distinct keyword extractors to see which produces the best results in order to analyse various approaches and compare results.:-

keywords_with_frequency: The countvectorizer is used in this function to tokenize, count,

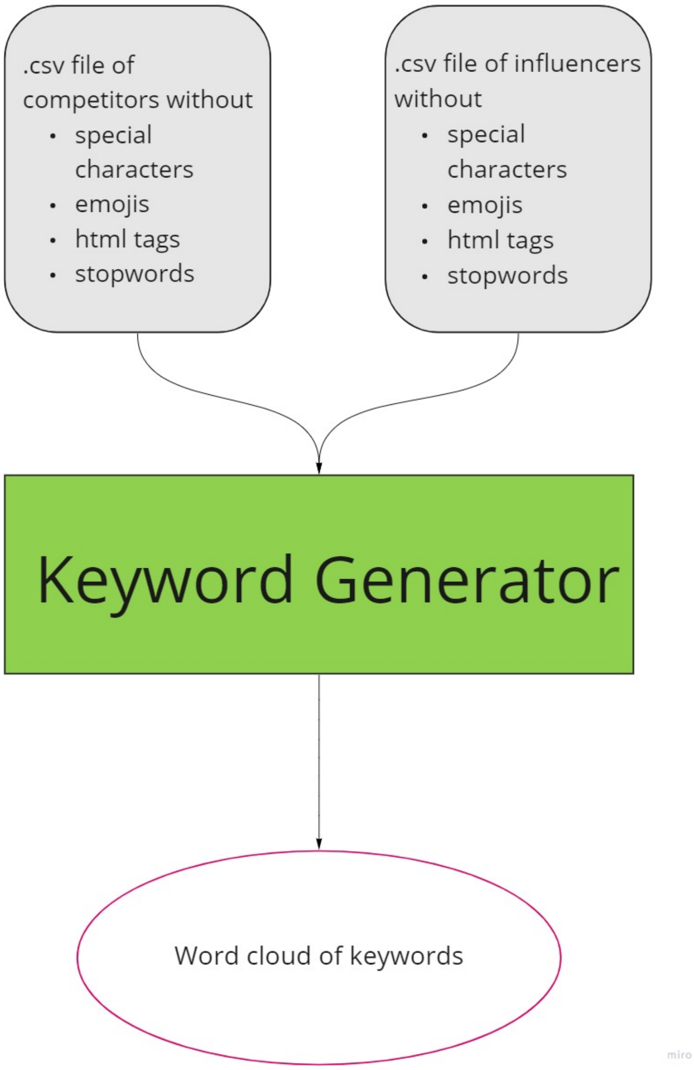

Figure 3.4: Keyword Generator

and return keywords based on their frequency.

keywords_with_tfidf: Using the TfidfVectorizer, the approach here performs the same function as the countvectorizer while also normalizing the data

keywords_with_rake: The final approach we looked at is using Rapid Automatic Keyword Extraction algorithm which uses a list of stopwords and phrase delimiters to detect the most relevant words or phrases in a piece of text

3.3.2 Similarity Score

The Similarity Score, which ranges from 0% to 100%, illustrates how distinctive any given segment's overlap with the audience being studied (such as brand affinity, influencers, media channels, etc.) is. In comparison to all other online audiences, the audience overlap increases as similarity increases.

Use cases for Similarity Score in Marketing

- **Influencer identification**: For instance in analysing a brands audience and the kind of influencer that suits that audience, if we have influencers with a similarity score of 99.9% with the specified while others have lower percentagesâwhich means that these influencers reside in the 99.9th percentile for how similar the brands audience. In essence, there is no better fit possible. Alternately, you can discover an overlap between the said influencer and your current audience. If so, it could be time to expand and connect with new markets by teaming up with another social media influencer that is not in the high percentile bracket.
- **Audience relevance:** In various situations, brands can seek to assess how similar or different their own patrons and/or social followings are to those of their ideal target market, a major competitor, or even a brand they want to rival. This can provide you some ideas about how to more effectively market your goods and services in order to get the results you want. For instance, you might discover that there is little overlap between your present customers and a competitor's audience and decide that it is time to engage in new collaborations and channels, particularly those where your rival sees success in capturing market share.[27]

In this research, we also conducted analysis to look at how various brands compare to one another as well as how various social influencers' digital footprints compare to the brands. To do this, we employed a module that tells us which texts are statistically more similar to one another than others, or less similar. The Logic behind this is to;

- Take a sentence and make it into a vector.
- Convert a large number of other sentences into vectors.

- Find the sentences that have the shortest distance (Euclidean similarity) or the smallest angle (cosine similarity) between them

As demonstrated below, two approaches were employed for this project and contrasted in order to decide which was the best.

The BERT Module: BERT (Bidirectional Encoder Representations from Transformers) is a state-of-the-art language processing model developed by Google trained to consider a word's role in a sentence as a whole rather than just its constituent components. As a result, BERT excels in a variety of natural language tasks, including text classification and language translation. After removing the emojis and other special characters from the data, the text data is encoded using the BERT module's encoder method. To ensure that the sentences are preserved, full stops and question marks are also included in the data sets parsed using this module. We also examined two approaches when determining the similarity score.

- By averaging all the similarity ratings: This method calculates the cosine angle between the texts in order to determine how similar they are, measures that similarity, and then averages all the scores to determine the similarity score between competitors and/or competitors and influencers.
- Clustering the embedding in the n-dim hyperspace and checking for overlapping. With this approach, we have similarity by calculating the overlap between both the clusters.

Universal Sentence Encoder: In the second method, we encoded the text data into high-dimensional vectors using the Universal Sentence Encoder module in order to compare similarity scores. This method involves calculating the cosine angle between the texts in order to determine how similar they are, then taking the average of all the scores to determine how similar the texts are between competitors and/or competitors and influencers.

Using this model, we have further investigated alternative methods, such as using the separation between cluster centroids in n-dimensional hyperspace. Since there is a higher degree of similarity between the clusters if they are closer to one another, we have taken the inverse of this distance to get the similarity measure from this distance.

As a final approach we have calculated the degree of similarity as a ratio between the distance of the centroids of the clusters and the theoretical maximum possible distance in the

n-dimensional embedding space. And all of these distances are euclidean and calculated only after transforming the embeddings data using *StandardScaler* from the sci-kit learn python package.
Using this approach we have computed the following similarity score metrics;

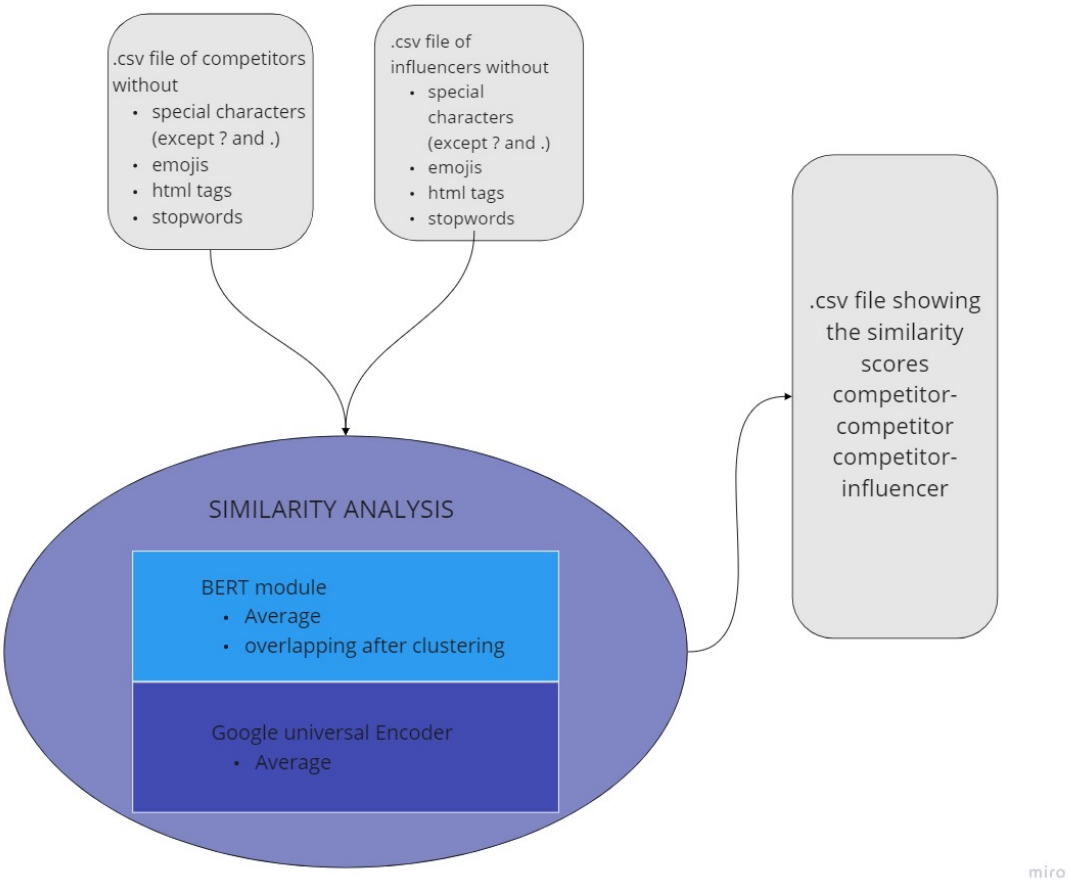

Figure 3.5: Similarity Analysis

- **competitor_competitor_similarity:** The findings of this function's comparison of all possible competitor combinations were saved in a.csv file. The output of the BERT Module's computation of the similarity score is shown below.

Table 3.4: Competitor versus Competitor Similarity score analysis

Competitor 1	**Competitor 2**	**Similarity**

bulk	formnutrition	0.09571436
bulk	indisupplements	0.10826271
bulk	medterra.international	0.08594484
bulk	motionnutrition	0.08755383
bulk	neat_nutrition	0.11315563
bulk	neurohacker	0.09602132
bulk	thenue_co	0.09773716
bulk	liveinnermost	0.10293566
bulk	puresport	0.10506216
formnutrition	indisupplements	0.1216798
formnutrition	medterra.international	0.12784594
formnutrition	motionnutrition	0.11198979
formnutrition	neat_nutrition	0.15380442
formnutrition	neurohacker	0.16190937
formnutrition	thenue_co	0.14106414
formnutrition	liveinnermost	0.14144558
formnutrition	puresport	0.13224193
indisupplements	medterra.international	0.11125488
indisupplements	motionnutrition	0.08331801
indisupplements	neat_nutrition	0.11285597
indisupplements	neurohacker	0.14618888
indisupplements	thenue_co	0.14529008
indisupplements	liveinnermost	0.11678053
indisupplements	puresport	0.15259959
medterra.international	motionnutrition	0.11768612
medterra.international	neat_nutrition	0.146678
medterra.international	neurohacker	0.16251624
medterra.international	thenue_co	0.14407927
medterra.international	liveinnermost	0.13569838
medterra.international	puresport	0.1544892

motionnutrition	neat_nutrition	0.14052732
motionnutrition	neurohacker	0.13346273
motionnutrition	thenue_co	0.11666483
motionnutrition	liveinnermost	0.1258248
motionnutrition	puresport	0.10700307
neat_nutrition	neurohacker	0.18085147
neat_nutrition	thenue_co	0.1604027
neat_nutrition	liveinnermost	0.19494691
neat_nutrition	puresport	0.14683163
neurohacker	thenue_co	0.18064076
neurohacker	liveinnermost	0.18775654
neurohacker	puresport	0.17501606
thenue_co	liveinnermost	0.1574012
thenue_co	puresport	0.15309742
liveinnermost	puresport	0.14664814

- **competitor_influencer_similarity:** This function computed the similarity between all the combinations of competitors and influencers and saved the results in .csv file. Below is the result of this similarity score computation using the BERT Module *competitor_name*, *influencer_name* and *similarity_score*.

Table 3.5: Competitor versus Influencer Similarity score analysis

Competitor 1	**Competitor 2**	**Similarity**
bulk	aaroncgshore	0.32539356
bulk	adamcollard	0.47380814
bulk	aliceliveing	0.4681129
bulk	brown.elle	0.22809556
bulk	charlottedawsy	0.4667351

bulk	chessieking	0.44980028
bulk	chloe.khan	0.3000237
bulk	courtneydblack	0.46914703
bulk	danosborneofficial	0.40538093
bulk	gabbydawnallen	0.45292887
bulk	ini.helen	0.22002934
bulk	itsalwayshana	0.41031796
bulk	jamesgshore	0.3650703
bulk	jamessmithpt	0.33164653
bulk	jesshunt2	0.27885893
bulk	jessica_rose_uk	0.3157875
bulk	jesswright77	0.34945253
bulk	katiepiper_	0.41391551
bulk	korisampson	0.32937753
bulk	lucymeck1	0.42466816
bulk	mac_griffiths_	0.35877553
bulk	mattdoesfitness	0.34154332
bulk	ohpolly	0.44624043
bulk	oliviadbowen	0.39806923
bulk	rogersnipes	0.43603134
bulk	slimmingworld	0.541391
bulk	sylvijaa	0.28485703
bulk	thebodycoach	0.45938155
bulk	thefitnesschef	0.49822393
bulk	_jackfowler_	0.35116935
formnutrition	aaroncgshore	0.29518047
formnutrition	adamcollard	0.4857793
formnutrition	aliceliveing	0.48658204
formnutrition	brown.elle	0.18278557
formnutrition	charlottedawsy	0.48484513

formnutrition	chessieking	0.48244077
formnutrition	chloe.khan	0.26817
formnutrition	courtneydblack	0.49299434
formnutrition	danosborneofficial	0.39671567
formnutrition	gabbydawnallen	0.4632332
formnutrition	ini.helen	0.16907531
formnutrition	itsalwayshana	0.42918476
formnutrition	jamesgshore	0.34203368
formnutrition	jamessmithpt	0.31778997
formnutrition	jesshunt2	0.23070864
formnutrition	jessica_rose_uk	0.2929776
formnutrition	jesswright77	0.33288115
formnutrition	katiepiper_	0.41498232
formnutrition	korisampson	0.30091184
formnutrition	lucymeck1	0.4356749
formnutrition	mac_griffiths_	0.3400807
formnutrition	mattdoesfitness	0.3189781
formnutrition	ohpolly	0.43374878
formnutrition	oliviadbowen	0.39573297
formnutrition	rogersnipes	0.4613913
formnutrition	slimmingworld	0.59632075
formnutrition	sylvijaa	0.25556728
formnutrition	thebodycoach	0.4689621
formnutrition	thefitnesschef	0.57600474
formnutrition	_jackfowler_	0.3324965
indisupplements	aaroncgshore	0.32121944
indisupplements	adamcollard	0.45450404
indisupplements	aliceliveing	0.463152
indisupplements	brown.elle	0.25234556
indisupplements	charlottedawsy	0.4681031

indisupplements	chessieking	0.45892665
indisupplements	chloe.khan	0.31494308
indisupplements	courtneydblack	0.45874742
indisupplements	danosborneofficial	0.39384943
indisupplements	gabbydawnallen	0.44654414
indisupplements	ini.helen	0.24466743
indisupplements	itsalwayshana	0.4184334
indisupplements	jamesgshore	0.36765125
indisupplements	jamessmithpt	0.35131207
indisupplements	jesshunt2	0.29614082
indisupplements	jessica_rose_uk	0.33087814
indisupplements	jesswright77	0.35859567
indisupplements	katiepiper_	0.4205119
indisupplements	korisampson	0.34043574
indisupplements	lucymeck1	0.43104
indisupplements	mac_griffiths_	0.36724356
indisupplements	mattdoesfitness	0.34279847
indisupplements	ohpolly	0.4322944
indisupplements	oliviadbowen	0.40558285
indisupplements	rogersnipes	0.45302948
indisupplements	slimmingworld	0.5522501
indisupplements	sylvijaa	0.30380186
indisupplements	thebodycoach	0.44118452
indisupplements	thefitnesschef	0.533696
indisupplements	_jackfowler_	0.3574721
medterra.international	aaroncgshore	0.35117263
medterra.international	adamcollard	0.45130825
medterra.international	aliceliveing	0.4649686
medterra.international	brown.elle	0.28642625
medterra.international	charlottedawsy	0.48343784

medterra.international	chessieking	0.46421987
medterra.international	chloe.khan	0.3481281
medterra.international	courtneydblack	0.44738057
medterra.international	danosborneofficial	0.420986
medterra.international	gabbydawnallen	0.45660326
medterra.international	ini.helen	0.26963925
medterra.international	itsalwayshana	0.43721962
medterra.international	jamesgshore	0.39248687
medterra.international	jamessmithpt	0.34658995
medterra.international	jesshunt2	0.32505593
medterra.international	jessica_rose_uk	0.34802848
medterra.international	jesswright77	0.39345193
medterra.international	katiepiper_	0.44671607
medterra.international	korisampson	0.3630308
medterra.international	lucymeck1	0.4467209
medterra.international	mac_griffiths_	0.37453258
medterra.international	mattdoesfitness	0.35680085
medterra.international	ohpolly	0.47766525
medterra.international	oliviadbowen	0.42453554
medterra.international	rogersnipes	0.4489515
medterra.international	slimmingworld	0.51799524
medterra.international	sylvijaa	0.32238606
medterra.international	thebodycoach	0.44516134
medterra.international	thefitnesschef	0.46299618
medterra.international	_jackfowler_	0.38810536
motionnutrition	aaroncgshore	0.3412237
motionnutrition	adamcollard	0.44388047
motionnutrition	aliceliveing	0.4547305
motionnutrition	brown.elle	0.25545874
motionnutrition	charlottedawsy	0.4789258

motionnutrition	chessieking	0.47928965
motionnutrition	chloe.khan	0.31601802
motionnutrition	courtneydblack	0.4368497
motionnutrition	danosborneofficial	0.41095155
motionnutrition	gabbydawnallen	0.44455504
motionnutrition	ini.helen	0.25357357
motionnutrition	itsalwayshana	0.42765737
motionnutrition	jamesgshore	0.37369922
motionnutrition	jamessmithpt	0.34741738
motionnutrition	jesshunt2	0.29218832
motionnutrition	jessica_rose_uk	0.33267355
motionnutrition	jesswright77	0.36882278
motionnutrition	katiepiper_	0.42268428
motionnutrition	korisampson	0.34662414
motionnutrition	lucymeck1	0.42927432
motionnutrition	mac_griffiths_	0.36159366
motionnutrition	mattdoesfitness	0.35414675
motionnutrition	ohpolly	0.44886777
motionnutrition	oliviadbowen	0.40938184
motionnutrition	rogersnipes	0.4482324
motionnutrition	slimmingworld	0.50059414
motionnutrition	sylvijaa	0.2934278
motionnutrition	thebodycoach	0.4217173
motionnutrition	thefitnesschef	0.48121044
motionnutrition	_jackfowler_	0.37698063
neat_nutrition	aaroncgshore	0.33143663
neat_nutrition	adamcollard	0.38992584
neat_nutrition	aliceliveing	0.40181008
neat_nutrition	brown.elle	0.29726747
neat_nutrition	charlottedawsy	0.40875062

neat_nutrition	chessieking	0.38630942
neat_nutrition	chloe.khan	0.3388635
neat_nutrition	courtneydblack	0.3882701
neat_nutrition	danosborneofficial	0.3687001
neat_nutrition	gabbydawnallen	0.39377323
neat_nutrition	ini.helen	0.29439145
neat_nutrition	itsalwayshana	0.37001377
neat_nutrition	jamesgshore	0.3610642
neat_nutrition	jamessmithpt	0.341093
neat_nutrition	jesshunt2	0.33964175
neat_nutrition	jessica_rose_uk	0.3431094
neat_nutrition	jesswright77	0.3577106
neat_nutrition	katiepiper_	0.383996
neat_nutrition	korisampson	0.34856343
neat_nutrition	lucymeck1	0.3924847
neat_nutrition	mac_griffiths_	0.3621187
neat_nutrition	mattdoesfitness	0.33888444
neat_nutrition	ohpolly	0.40226078
neat_nutrition	oliviadbowen	0.37858826
neat_nutrition	rogersnipes	0.3917638
neat_nutrition	slimmingworld	0.43051308
neat_nutrition	sylvijaa	0.3293084
neat_nutrition	thebodycoach	0.38078767
neat_nutrition	thefitnesschef	0.41245386
neat_nutrition	_jackfowler_	0.35431755
neurohacker	aaroncgshore	0.3355669
neurohacker	adamcollard	0.45648438
neurohacker	aliceliveing	0.4691215
neurohacker	brown.elle	0.26356423
neurohacker	charlottedawsy	0.45533428

neurohacker	chessieking	0.44420356
neurohacker	chloe.khan	0.32630694
neurohacker	courtneydblack	0.439316
neurohacker	danosborneofficial	0.39943722
neurohacker	gabbydawnallen	0.4507146
neurohacker	ini.helen	0.25538793
neurohacker	itsalwayshana	0.43116173
neurohacker	jamesgshore	0.38914725
neurohacker	jamessmithpt	0.36436895
neurohacker	jesshunt2	0.30185416
neurohacker	jessica_rose_uk	0.3283642
neurohacker	jesswright77	0.35763827
neurohacker	katiepiper_	0.43932265
neurohacker	korisampson	0.35730892
neurohacker	lucymeck1	0.41804898
neurohacker	mac_griffiths_	0.3849016
neurohacker	mattdoesfitness	0.36580756
neurohacker	ohpolly	0.45086768
neurohacker	oliviadbowen	0.4033935
neurohacker	rogersnipes	0.4675332
neurohacker	slimmingworld	0.5053455
neurohacker	sylvijaa	0.2974768
neurohacker	thebodycoach	0.43044248
neurohacker	thefitnesschef_	0.48230654
neurohacker	_jackfowler_	0.38070044
thenue_co	aaroncgshore	0.30835906
thenue_co	adamcollard	0.46649444
thenue_co	aliceliveing	0.47469467
thenue_co	brown.elle	0.23597103
thenue_co	charlottedawsy	0.46704042

thenue_co	chessieking	0.46354297
thenue_co	chloe.khan	0.30448124
thenue_co	courtneydblack	0.46049765
thenue_co	danosborneofficial	0.38738805
thenue_co	gabbydawnallen	0.45343995
thenue_co	ini.helen	0.22136483
thenue_co	itsalwayshana	0.43323994
thenue_co	jamesgshore	0.36801296
thenue_co	jamessmithpt	0.34381998
thenue_co	jesshunt2	0.28154132
thenue_co	jessica_rose_uk	0.31480753
thenue_co	jesswright77	0.3495744
thenue_co	katiepiper_	0.43377477
thenue_co	korisampson	0.33739257
thenue_co	lucymeck1	0.4282668
thenue_co	mac_griffiths_	0.36707616
thenue_co	mattdoesfitness	0.3443993
thenue_co	ohpolly	0.45169505
thenue_co	oliviadbowen	0.40540507
thenue_co	rogersnipes	0.46563804
thenue_co	slimmingworld	0.5468346
thenue_co	sylvijaa	0.28318784
thenue_co	thebodycoach	0.4356077
thenue_co	thefitnesschef	0.52267337
thenue_co	_jackfowler_	0.35835257
liveinnermost	aaroncgshore	0.33758366
liveinnermost	adamcollard	0.48744375
liveinnermost	aliceliveing	0.5014976
liveinnermost	brown.elle	0.2562783
liveinnermost	charlottedawsy	0.48941192

liveinnermost	chessieking	0.47828734
liveinnermost	chloe.khan	0.32749623
liveinnermost	courtneydblack	0.48497206
liveinnermost	danosborneofficial	0.41689062
liveinnermost	gabbydawnallen	0.47897378
liveinnermost	ini.helen	0.24851365
liveinnermost	itsalwayshana	0.45220083
liveinnermost	jamesgshore	0.39721513
liveinnermost	jamessmithpt	0.36736414
liveinnermost	jesshunt2	0.30191386
liveinnermost	jessica_rose_uk	0.33575028
liveinnermost	jesswright77	0.37227833
liveinnermost	katiepiper_	0.4523603
liveinnermost	korisampson	0.36202294
liveinnermost	lucymeck1	0.4466439
liveinnermost	mac_griffiths_	0.39401442
liveinnermost	mattdoesfitness	0.36799836
liveinnermost	ohpolly	0.47350147
liveinnermost	oliviadbowen	0.42648202
liveinnermost	rogersnipes	0.48511398
liveinnermost	slimmingworld	0.57279533
liveinnermost	sylvijaa	0.30686706
liveinnermost	thebodycoach	0.46355963
liveinnermost	thefitnesschef	0.5423825
liveinnermost	_jackfowler_	0.3822813
puresport	aaroncgshore	0.32776156
puresport	adamcollard	0.4721535
puresport	aliceliveing	0.47813934
puresport	brown.elle	0.2443737
puresport	charlottedawsy	0.4675665

puresport	chessieking	0.45273575
puresport	chloe.khan	0.31470585
puresport	courtneydblack	0.45742455
puresport	danosborneofficial	0.41583857
puresport	gabbydawnallen	0.4596398
puresport	ini.helen	0.22914216
puresport	itsalwayshana	0.42250407
puresport	jamesgshore	0.38463473
puresport	jamessmithpt	0.33753705
puresport	jesshunt2	0.2861293
puresport	jessica_rose_uk	0.3203051
puresport	jesswright77	0.36182135
puresport	katiepiper_	0.44218072
puresport	korisampson	0.34500146
puresport	lucymeck1	0.42887607
puresport	mac_griffiths_	0.3662688
puresport	mattdoesfitness	0.34866637
puresport	ohpolly	0.46484667
puresport	oliviadbowen	0.40779525
puresport	rogersnipes	0.45360768
puresport	slimmingworld	0.53019494
puresport	sylvijaa	0.2901702
puresport	thebodycoach	0.44318193
puresport	thefitnesschef	0.4725133
puresport	_jackfowler_	0.36930153

3.3.3 Clustering for Customer Segmentation

The clustering method helps with customer segmentation, which is the process of putting comparable customers into the same segment, in a business context. The clustering approach helps with both static and dynamic customer behaviour analysis. Businesses can benefit from this method by creating segment-specific marketing strategies because customers with similar characteristics usually interact with businesses in similar ways.[28] In this project, where it is important to segment customers based on their preferences as seen in their digital footprints, we would cluster influencers based on their dynamic behaviours using text data extracted and processed from their digital footprints (Social media pages; Instagram, Twitter, Facebook). Numerous clustering algorithms exist, including k-means, hierarchical clustering, DBSCAN clustering, and others. In this project, K-means clustering has been the main focus.

K-means Clustering

The unsupervised learning class of machine learning models includes the commonly used K-means technique for data grouping. It's perfect for a number of client segmentation strategies. In contrast to supervised learning models, unsupervised learning models find previously hidden patterns in data without the aid of a label to instruct them on what to do. Unsupervised learning models are just given a set of data and instructed to sort the data according to how similar the data in each group are. This could involve grouping customers according to their frequency, financial value, the products they buy, or a variety of other factors for a clustering method on a customer data set. Based on a variety of different variables, the model develops distinct client segments and uses them to assign a cluster name to each group. [29]

Unlike supervised algorithms, which have many metrics to measure their goodness of fit, such as accuracy, r-square value, sensitivity, specificity, and so on, clustering techniques use the Silhouette Coefficient or Silhouette score, which has a value ranging from -1 to 1.
1: Means cluster are distinct and separated by a large distance.
0: This implies that there are no statistically significant differences across clusters, or that clusters are not differentiable.
-1: TThis suggests that the clusters were assigned in a wrong manner.
In working on this project we have compared different ways of clustering to try to determine

the best way to segment the influencers with respect to each brand

Clustering the influencers for each brand based on similarity score

In trying this method we have focused on clustering based on the similarity scores of the competitor and influencer following the steps below. The code is saved in the file named *SimilarityClustering.ipynb*

- Imported the competitor_Influencer similarity analysis
- Set a filter to select a brand
- Use silhouette coefficient to find the optimal number of cluster and
- Use KMeans to cluster the influencers

The brands can be filtered to get their individual cluster output.

Clustering the influencers based on text data and finding the similarity score

With this method we have focused on clustering the influencers using the data extracted and cleaned and checking the similarity score between each cluster and the individual brands, the process followed is listed below.The code is saved in the file named *SimilarityClustering_influencer.ipynb*

- Import the influencer pre-processed text data
- Use TfidfVectorizer to transform the text to feature vectors
- Use silhouette coefficient to find the optimal number of cluster
- Use KMeans to cluster the influencers
- Import the competitor data
- Calculate the similarity between the competitor data and the clusters

Using all the pre-processed texts of the influencers, we notice that the recommended number of K is 3, however to see some more clusters options we used the next best number which was 7. Below is a table showing the clusters

Table 3.6: Influencer Clustering Table using all Text

Influencer Name	**Cluster**
gabbydawnallen	0
gabbydawnallen	0
rogersnipes	0
mac_griffiths_	0
adamcollard	0
aliceliveing	0
thebodycoach	1
slimmingworld	1
courtneydblack	1
thefitnesschef_	1
jamessmithpt	1
itsalwayshana	2
mattdoesfitness	3
jackfowler	3
korisampson	3
jamesgshore	3
sylvijaa	4
brown.elle	4
ini.helen	5
oliviadbowen	6
ohpolly	6
aaroncgshore	6
katiepiper_	6
jesswright77	6
jessica_rose_uk	6
danosborneofficial	6
chloe.khan	6
chessieking	6

charlottedawsy	6
lucymeck1	6
jesshunt2	6

Clustering the influencers based on keywords extracted and finding the similarity score
With this method we have focused on clustering the influencers using the keywords extracted from the influencers and then checking the similarity score between each cluster and the individual brands, the process followed is listed below. The code is saved in the file named *SimilarityClustering_influencerbyKeyword_1.ipynb*

- Import the influencer keyword data
- Use TfidfVectorizer to transform the text to feature vectors
- Use silhouette coefficient to find the optimal number of cluster
- Use KMeans to cluster the influencers
- Import the competitor data
- Calculate the similarity between the competitor data and the clusters

Below is the table that shows the cluster of influencers using the keywords data

Table 3.7: Influencer Clustering Table using keywords

Influencer Name	**Cluster**
aliceliveing	0
courtneydblack	0
ohpolly	0
thebodycoach	0
chessieking	1
chloe.khan	1
katiepiper_	1
jackfowler	1

aaroncgshore	2
charlottedawsy	2
danosborneofficial	2
ini.helen	2
jesswright77	2
lucymeck1	2
oliviadbowen	2
jamessmithpt	3
mac_griffiths_	3
rogersnipes	3
thefitnesschef	3
jesshunt2	4
adamcollard	5
gabbydawnallen	5
itsalwayshana	5
jamesgshore	6
korisampson	6
mattdoesfitness	6
jessica_rose_uk	7
brown.elle	8
sylvijaa	8
slimmingworld	9

CHAPTER 4

Comparative Analysis

Its normal for companies to want to identify with segments that are the best fit, so logically based on the clustering done above brands will go with the influencer segments with the highest similarity score. However there also is the possibility of an untapped market with a slightly closer similarity or more advantages than the others. For reason we are exploring two methods to be considered for recommending a cluster as a segment.

Influencer Cluster Recommendation

With this recommendation method, we've suggested a formula that takes into account the average of the second to fourth highest similarity scores under the assumption that some clusters, despite not having the highest similarity scores, may have the potential to be a much better target segment because the highest similarity score may simply represent a market that is saturated or has been fully penetrated with a significant retention rate.

- Select a Competitor
- Filter the similarity scores in descending order
- Calculate the average of the second, third and fourth similarity scores
- Select the similarity score closest to the average
- Recommend the cluster

For Example if we look at the similarity score of the competitor Neurohacker below, we see that the best cluster fit is **cluster 6** as it has the highest similarity, however the proposition is to find another cluster with a relationship that can be harnessed and used to posibly identify a new target segment.

Table 4.1: Similarity score table of neurohacker

Competitor	**Cluster**	**Similarity Score**
neurohacker	6	0.318582
neurohacker	9	0.31388992
neurohacker	0	0.2536287
neurohacker	5	0.23512466
neurohacker	3	0.22808231
neurohacker	1	0.21927465
neurohacker	7	0.18651901
neurohacker	4	0.17052253
neurohacker	8	0.15630308
neurohacker	2	0.14542738

Calculating for a recommended cluster for this competitor based on the proposed algorithm shows;
Average of the next top 3: (0.3138892 + 0.2536287 + 0.23512466)/3 = 0.26754776
The closest similarity on the table to this average is 0.2536287, so we can make the recommendation for **cluster 0** to be a segment to be considered by Neurohacker.

Competitive Advantage

For each brand, it's critical to consider each cluster and how they relate to other competitor in order to determine whether one has an advantage over the others. To achieve this, we have devised a method to determine a cluster's competitive advantage over another; To Calculate the competitive advantage of a brand;
Competitive Advantage = Similarity Score - Competitive advantage score;
Where the **competitive advantage score** is the average of the 'similarity_score' for each cluster group excluding each row; For Example

Table 4.2: Sample Cluster and Similarity score table

Competitor	**Cluster**	**Similarity Score**
Bulk	2	0.037
Bulk	5	0.256
Formnutrition	2	0.094
Formnutrition	5	0.288
Indisupplements	2	0.041
Indisupplements	5	0.346

Using our proposed formula to calculate competitive advantage scores and work out a recommendation;

For Bulk

- For cluster 2, the competitor average is (0.094 + 0.041)/2 = 0.068. Bulks competitive advantage score is thus 0.037 - 0.068 = -0.031.
- For cluster 5, the competitor average is (0.288 + 0.346)/2 = 0.317.
- Bulks competitive advantage score is thus 0.256 - 0.317 = -0.061.

Since -0.031 > -0.061, Bulk could consider making changes to target cluster 2 even though its similarity with cluster 5 is currently higher.

For Formnutrition:

- For cluster 2, the competitor average is (0.037 + 0.041)/2 = 0.039. Formnutritions competitive advantage score is thus 0.094 - 0.039 = 0.055.
- For cluster 5, the competitor average is (0.256 + 0.346)/2 =0.301.
- Formnutritions competitive advantage score is thus 0.288 - 0.301 = -0.013

Since 0.055 > -0.013, Formnutrition could consider making changes to target cluster 2 even though its similarity with cluster 5 is currently higher.

For Indisupplements:

- For cluster 2, the competitor average is (0.037 + 0.094)/2 = 0.066. Indisupplemets competitive advantage score is thus 0.041 - 0.066 = -0.025.

- For cluster 5, the competitor average is (0.256 + 0.288)/2 = 0.272.
- Indisupplemets competitive advantage score is thus 0.346 - 0.272 = 0.074.

Since -0.025 < 0.074, Indisupplements is wellplaced to target cluster 5.
Using this method, below is the resulting table for the competitive advantage analysis showing the advantage score for each brand in reference to a cluster.

Table 4.3: Competitive Advantage score Table

Competitor Name	Cluster	Similarity Score	Competitive Advantage Score
bulk	0	0.21394557	-0.083469433
formnutrition	0	0.2843158	-0.005280289
indisupplements	0	0.3186174	0.0328326
medterra.international	0	0.20233245	-0.0963729
motionnutrition	0	0.31016856	0.023445
neat_nutrition	0	0.42383572	0.149741844
neurohacker	0	0.2536287	-0.039377067
thenue_co	0	0.2760549	-0.014459067
liveinnermost	0	0.36951104	0.089381089
puresport	0	0.23827046	-0.056441778
bulk	1	0.08358265	-0.067993416
formnutrition	1	0.120223254	-0.027281634
indisupplements	1	0.08896716	-0.062010628
medterra.international	1	0.17527696	0.03388915
motionnutrition	1	0.2603973	0.128467306
neat_nutrition	1	0.21468815	0.077679361
neurohacker	1	0.21927465	0.082775472
thenue_co	1	0.16868961	0.026569872
liveinnermost	1	0.06722558	-0.086167939
puresport	1	0.049441934	-0.105927545
bulk	2	0.037148982	-0.069634581
formnutrition	2	0.093580544	-0.006932845

indisupplements	2	0.041387	-0.064925672
medterra.international	2	0.10197611	0.002395562
motionnutrition	2	0.16706789	0.074719762
neat_nutrition	2	0.21263593	0.125350917
neurohacker	2	0.14542738	0.050674751
thenue_co	2	0.094961345	-0.005398622
liveinnermost	2	0.091623425	-0.009107422
puresport	2	0.012392439	-0.097141851
bulk	3	0.276119	-0.006152818
formnutrition	3	0.2799387	-0.001908707
indisupplements	3	0.3100599	0.031559293
medterra.international	3	0.17025042	-0.123784573
motionnutrition	3	0.4032563	0.135110849
neat_nutrition	3	0.38339654	0.113044449
neurohacker	3	0.22808231	-0.059526918
thenue_co	3	0.24601313	-0.039603784
liveinnermost	3	0.36658219	0.094361838
puresport	3	0.15286687	-0.143099629
bulk	4	0.14374363	-0.038732469
formnutrition	4	0.1549145	-0.026320391
indisupplements	4	0.16675055	-0.013169224
medterra.international	4	0.14742175	-0.034645669
motionnutrition	4	0.22746633	0.054292753
neat_nutrition	4	0.22605625	0.052725998
neurohacker	4	0.17052253	-0.008978136
thenue_co	4	0.19921231	0.022899398
liveinnermost	4	0.22064115	0.04670922
puresport	4	0.12929952	-0.05478148
bulk	5	0.25633186	-0.03242608
formnutrition	5	0.28804517	0.002810931

indisupplements	5	0.34637356	0.067620253
medterra.international	5	0.16742086	-0.13121608
motionnutrition	5	0.28488916	-0.000695747
neat_nutrition	5	0.4009598	0.128271631
neurohacker	5	0.23512466	-0.055989636
thenue_co	5	0.21442108	-0.078993613
liveinnermost	5	0.44462317	0.176786487
puresport	5	0.216964	-0.076168147
bulk	6	0.20217113	-0.037731756
formnutrition	6	0.26152807	0.0282204
indisupplements	6	0.24262172	0.007213344
medterra.international	6	0.16912326	-0.074451611
motionnutrition	6	0.3739013	0.153079544
neat_nutrition	6	0.27579415	0.0440716
neurohacker	6	0.318582	0.091613656
thenue_co	6	0.17223114	-0.070998411
liveinnermost	6	0.2761231	0.0444371
puresport	6	0.06922123	-0.185453867
bulk	7	0.051743202	-0.037108231
formnutrition	7	-0.016878208	-0.113354242
indisupplements	7	0.057859942	-0.030311853
medterra.international	7	0.08377047	-0.001522378
motionnutrition	7	0.17302911	0.097653889
neat_nutrition	7	0.12651783	0.045974689
neurohacker	7	0.18651901	0.112642667
thenue_co	7	0.116667196	0.03502954
liveinnermost	7	0.08275756	-0.002647833
puresport	7	-0.010580014	-0.106356249
bulk	8	0.15742004	0.008949047
formnutrition	8	0.1381765	-0.012432664

indisupplements	8	0.1315069	-0.01984333
medterra.international	8	0.09628095	-0.058983275
motionnutrition	8	0.20240007	0.058926859
neat_nutrition	8	0.23007338	0.089674981
neurohacker	8	0.15630308	0.007707981
thenue_co	8	0.17833498	0.03218787
liveinnermost	8	0.1504412	0.001194781
puresport	8	0.052721873	-0.107382249
bulk	9	0.37342316	0.085267317
formnutrition	9	0.3484019	0.057465917
indisupplements	9	0.40981308	0.125700562
medterra.international	9	0.19790158	-0.10975666
motionnutrition	9	0.33282858	0.040162228
neat_nutrition	9	0.32934767	0.036294551
neurohacker	9	0.31388992	0.019119273
thenue_co	9	0.27549097	-0.023546227
liveinnermost	9	0.30862182	0.013265828
puresport	9	0.077107064	-0.243972789

CHAPTER 5

Results and Conclusion

With the similarity scores between competitors and competitors and influencers, we can tell how similar one is with the other. However with marketing trends, asides similarity it is a important to identify segments which we have done with clustering influencers. We then looked at each cluster as a representation of a customer segment and checked for the similarity between the competitors and each cluster. Seeing that we are representing each cluster as a target segment, we can conclude that for the scope of this study, the 10 brands have 10 target segments.

Another angle we have looked at in this project is a preferred cluster recommendations for brands not just based on the highest similarity scores as seen in table 4.1 but based on a proposed algorithm. Selecting the brand Neurohacker as an example which has its highest similarity score pointing at cluster 6 , the recommended cluster using this method which selects the closest cluster to the similarity score of the average of the second to fourth scores is cluster 0. After looking deeply into the individual influencers in the recommended cluster 0 from the calculations of the proposed algorithm and that of cluster 6 (see table 3.6 for influencer details) we can deduce the following;

- Individually, the influencer with the most similar score to the Neurohacker brand is slimming world
- From the recommendation method, cluster 0 which consists of aliceliveing, courtneydblack, ohpolly and thebodycoach has a great potential as individually we see each influencer having a relatively high similarity score with the brand as well, so we can

say that this might be a great target market fit for the brand

- From the competitive advantage score method, we see that the cluster identifies influencers with a more average similarity score compared to that of the highest score per individual influencer and the recommended cluster using the proposed method.

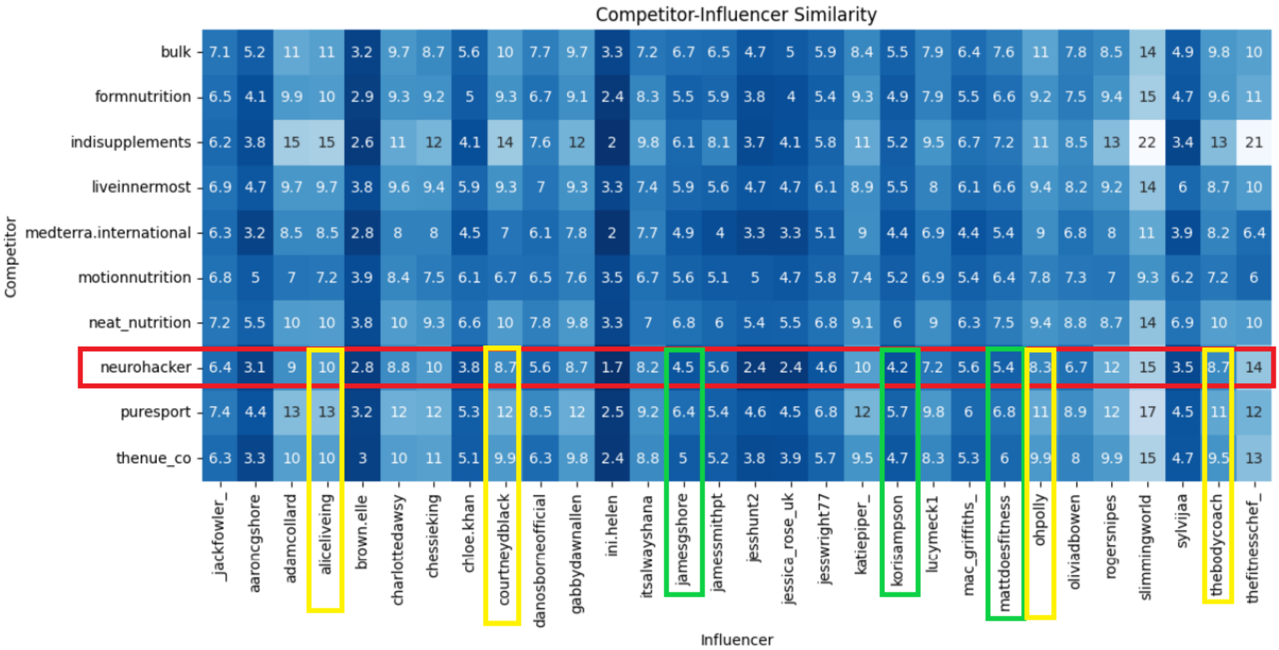

Figure 5.1: Heatmap for Neurohacker Analysis

So, from this we can conclude that using AI can definitely give better insights to drive decision making, whilst a marketing strategy can be to go with the most similar target market, another strategy can be to identify a segment within the sector that is not fully saturated but with enough similarity with to the company, and another can be to go with a competitive advantage strategy, creating an avenue to move into new markets as this compares the segments with other key players in the industry who are the competing brands.

Bibliography

[1] C. Bachor, "Ai tool for marketing project proposal(ce902)," 2022, university of Essex.

[2] J. Shabbir and T. Anwer, "Artificial intelligence and its role in near future," *CoRR*, vol. abs/1804.01396, 2018. [Online]. Available: http://arxiv.org/abs/1804.01396

[3] D. G. T. B. Thomas Davenport, Abhijit Guha. How artificial intelligence will change the future of marketing. [Online]. Available: https://rdcu.be/cZjy8

[4] K. H. Shankari and R. Thirumalaiselvi, "A survey on using artificial intelligence techniques in the software development process," *Int. J. Eng. Res. Appl*, vol. 4, no. 12, pp. 24–33, 2014.

[5] T. G. Gill, "Early expert systems: Where are they now?" *MIS quarterly*, pp. 51–81, 1995.

[6] S. Verma, R. Sharma, S. Deb, and D. Maitra, "Artificial intelligence in marketing: Systematic review and future research direction," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100002, 2021.

[7] S. Li, "Agentstra: an internet-based multi-agent intelligent system for strategic decision-making," *Expert Systems with Applications*, vol. 33, no. 3, pp. 565–571, 2007.

[8] F. J. Martínez-López and J. Casillas, "Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights," *Industrial Marketing Management*, vol. 42, no. 4, pp. 489–495, 2013.

[9] N. Syam and A. Sharma, "Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice," *Industrial marketing management*, vol. 69, pp. 135–146, 2018.

[10] J. R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marqués, "Setting b2b digital marketing in artificial intelligence-based crms: A review and directions for future research," *Industrial Marketing Management*, vol. 98, pp. 161–178, 2021.

[11] N. Capon and R. Glazer, "Marketing and technology: a strategic coalignment," *Journal of marketing*, vol. 51, no. 3, pp. 1–14, 1987.

[12] B. Vlačić, L. Corbo, S. C. e Silva, and M. Dabić, "The evolving role of artificial intelligence in marketing: A review and research agenda," *Journal of Business Research*, vol. 128, pp. 187–203, 2021.

[13] G. von Joerg and J. Carlos, "Design framework for the implementation of ai-based (service) business models for small and medium-sized manufacturing enterprises," *Journal of the Knowledge Economy*, pp. 1–19, 2022.

[14] S. S. Rao, A. Nahm, Z. Shi, X. Deng, and A. Syamil, "Artificial intelligence and expert systems applications in new product developmentâa survey," *Journal of Intelligent Manufacturing*, vol. 10, no. 3, pp. 231–244, 1999.

[15] K. M. Vladimirovich, "Future marketing in b2b segment: Integrating artificial intelligence into sales management," *International Journal of Innovative Technologies in Economy*, no. 4 (31), 2020.

[16] J. Kietzmann, J. Paschen, and E. Treen, "Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey," *Journal of Advertising Research*, vol. 58, no. 3, pp. 263–267, 2018.

[17] L. Ma and B. Sun, "Machine learning and ai in marketing–connecting computing power to human insights," *International Journal of Research in Marketing*, vol. 37, no. 3, pp. 481–504, 2020.

[18] P. N. Vicente, "Interactive media," *The SAGE International Encyclopedia of Mass Media and Society, SAGE*, 2020.

[19] M. Xie and J. Chen, "Pricing and ordering strategies of e-retailers in electronic business environments," *Tsinghua Science and Technology*, vol. 10, no. S1, pp. 778–789, 2005.

[20] K. Misra, E. M. Schwartz, and J. Abernethy, "Dynamic online pricing with incomplete information using multiarmed bandit experiments," *Marketing Science*, vol. 38, no. 2, pp. 226–252, 2019.

[21] J. Bauer and D. Jannach, "Optimal pricing in e-commerce based on sparse and noisy data," *Decision support systems*, vol. 106, pp. 53–63, 2018.

[22] T. S. Chung, R. T. Rust, and M. Wedel, "My mobile music: An adaptive personalization system for digital audio players," *Marketing Science*, vol. 28, no. 1, pp. 52–68, 2009.

[23] A. Toumazou. Best nootropic supplements to improve work-from-home productivity. british gq, apr 2020. [Online]. Available: https://www.gq-magazine.co.uk/lifestyle/gallery/best-nootropics-for-productivity

[24] P. G. . C. B. et al., "Ai marketing planning tool," 2022, university of Essex.

[25] J. Phoenix, "How to extract the text from multiple webpages in python," *Analytics Vidhya*, August 2021. [Online]. Available: https://understandingdata.com/python-for-seo/how-to-extract-text-from-multiple-webpages-in-python/

[26] R. Agrawal, "Must known techniques for text preprocessing in nlp," *Analytics Vidhya*, June 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/06/must-known-techniques-for-text-preprocessing-in-nlp/

[27] A. Intelligence. Statsocial launches similarity score for a new approach to audience measurement. [Online]. Available: https://www.statsocial.com/statsocial-launches-similarity-score-for-a-new-approach-to-audience-measurement

[28] D. Gong. Clustering Algorithm for Customer Segmentation. [Online]. Available: https://towardsdatascience.com/clustering-algorithm-for-customer-segmentation-e2d79e28cbc3

[29] P. D. Science. How to use k means clustering for customer segmentation. [Online]. Available: https://practicaldatascience.co.uk/machine-learning/how-to-use-k-means-clustering-for-customer-segmentation