BRAND MANAGEMENT ON SOCIAL MEDIA:
A MACHINE LEARNING-BASED FRAMEWORK USING TEXT AND IMAGES

A doctoral dissertation submitted in partial fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

IN

MANAGEMENT

by

V. ANAND

Regn. No. 18MBPH18

Under the supervision of

Dr. D. V. SRINIVAS KUMAR

Associate Professor

SCHOOL OF MANAGEMENT STUDIES

UNIVERSITY OF HYDERABAD

HYDERABAD – 500046, TS, INDIA

DECLARATION

I, V. Anand, hereby declare that this thesis entitled **“Brand Management on Social Media: A Machine Learning-based Framework using Text and Images”** in fulfilment of the requirements for the award of Degree of Doctor of Philosophy in Management Studies, is the outcome of original study, free of plagiarism, undertaken by me under the supervision of **Dr. D. V. Srinivas Kumar.**

This thesis is free from plagiarism and has not been submitted in part or full earlier to any other University or Institution for the award of any Degree or Diploma. I hereby agree that my thesis can be deposited in Shodhganga / INFLIBNET. A report of plagiarism statistics from the University Librarian is enclosed.

Place: Hyderabad — V. Anand

Date: December 07, 2023 — Reg. No: 18MBPH18

CERTIFICATE

This is to certify that the thesis entitled **"Brand Management on Social Media: A Machine Learning-based Framework using Text and Images"** submitted by **V. Anand** bearing registration number **18MBPH18** in partial fulfilment of the requirements for award of Doctor of Philosophy in the **School of Management Studies, University of Hyderabad** is a bonafide work carried out by him under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Parts of this thesis have been published in the following publications:
1. International Journal of Internet Marketing and Advertising (Indexing: Scopus, ABDC-C)
2. Asia Pacific Journal of Information Systems (Indexing: Scopus, ABDC-C)

The student had presented papers in the following conferences:
1. 7^{th} International Conference on Business Analytics and Intelligence, IIM Bangalore, December 2019
2. International Conference on Data Analytics for Business and Industry (Online), University of Bahrain, October 2020

Further, the student has passed the following courses towards fulfilment of coursework requirement for Ph.D.:

Course Code	**Name**	**Credits**	**Pass / Fail**
1. PH-101	Statistics for Research	3	Pass
2. EG-825	Academic Writing for Doctoral Students	2	Pass
3. MB-207	Research Methodology for Managers	3	Pass
4. BA-305	Big Data Analytics	3	Pass
5. PH-104	Research Methodology – II	4	Pass
6. BA-207	Machine Learning and Data Mining	3	Pass

Supervisor

Dean, School of Management Studies

ACKNOWLEDGEMENTS

My supervisor, Dr. D. V. Srinivas Kumar, was involved in every aspect of the Ph.D. journey: suggesting good journals for finding research topics and research gaps; keeping track of the important papers; suggesting unstructured data; helping me with the research proposal; guiding me on how to write articles and respond to reviewers' comments; offering relevant insights about the marketing domain; communicating with Google LLC about the $1000 grant on Google Cloud; planning the activities with an eye on submitting the thesis within a reasonable timeframe; and providing review comments about the thesis. I sincerely thank him for his instrumental role in ensuring that my Ph.D. work was feasible and relevant to the marketing domain.

The Doctoral Committee members – Sr. Prof. B. Rajashekhar, Prof. G. V. R. K. Acharyulu and Dr. Pramod Kumar Mishra – advised me in countless ways about refining the research proposal, findings and presentations. Without their constructive criticism and suggestions, this thesis would have been poorer in content and rigor.

I am thankful to the current dean, Prof. V. Mary Jessica, and the former dean, Prof. P. Jyoti, for their timely help with various administrative matters.

Prof. Irala Lokanandha Reddy (Statistics), Prof. Vijaya Bhaskar Marisetty (Research Methodology II), Dr. Jasti Appa Swami (Academic Writing), Dr. P. Murugan (Research Methodology I), Dr. G. Manoj Kumar (Data Visualization), Dr. Suresh Kandulapati (Data Mining, Machine Learning and Big Data Analytics) and Dr. Naresh Manwani (Data Mining and Machine Learning) taught me several useful aspects of research during the coursework. I gratefully acknowledge the influence of their expertise on my thesis.

Our school's Ph.D. coordinator, Dr. P. Murugan, arranged talks by researchers from India and other countries. In this thesis, I utilized Golder et al. (2023) and Faff (2013) material from the Ph.D. WhatsApp group managed by him. I thank Dr. P. Murugan for his untiring efforts to improve the quality of research output of the school.

It was a pleasant experience to attend the Ph.D. coursework with my batchmates Smt. Nishchala, Smt. Indira, Mr. Rahul, Mr. Sourya and Mr. Rushabh. I am thankful to my co-scholars Mr. Ajay and Mr. Sunil for helping me with some of the processes and presentations.

The school's non-teaching staff, Mr. Naganna, Mrs. Parimala and Mr. Parasuram, helped me navigate the administrative processes pertaining to the PhD program. Mr. Rajasekhar, IT systems administrator, helped me with the software, hardware, audio and video issues. I am obliged to them for their prompt and efficient support.

I appreciate the staff of Indira Gandhi Memorial Library, University of Hyderabad for providing me access to various books, theses, journals and anti-plagiarism services. Also, they never disturbed me during my frequent naps in the spacious reading halls of the library.

University of Hyderabad's non-NET fellowship allowed me to defray some of the living expenses. I am indebted to the university for their financial assistance.

Dr. Joseph T. Yun, University of Pittsburgh, USA and Dr. Chen Wang, University of Illinois, Urbana-Champaign, USA shared their expertise on how to use the Social Media Macroscope. Dr. Anne-Marie Nanne, University of Twente, Netherlands helped me with image analytics. I appreciate their scholarly advice on using deep learning in marketing analytics.

Google's grant of $1000 via grant ID 209475195 enabled me to use the Google Cloud Vision in my research. I am thankful to Google LLC for their financial assistance.

Finally, and most importantly, I thank my mother and father.

ABSTRACT

Effective brand management on the social media networks has been a challenge for the brand managers. This thesis focuses on monitoring some of the brand management aspects using text data from Twitter and image data from Instagram.

Social media increased the expressiveness of users and decreased the firms' control over their brand personalities. Brand managers feed massive amounts of firm-generated and user-generated content to the business information systems to understand the online perceptions of brands. By analyzing 83,264 tweets of 13 prominent Indian brands in four sectors using a machine learning model, we established that some of the newer brands are savvier in managing their brand personalities than the relatively older ones. The public sector brands have poorer alignment between communicated and perceived brand personalities than the private sector brands on Twitter. We provided evidence of positive association between the communicated brand personality and brand position. A counterintuitive finding is that the brand sentiment changes in value but not in category (positive, neutral or negative), after removing the bots and suspicious Twitter accounts.

Maintaining a brand identity via images on social media is a relatively new territory for the brand managers. The brand identity-brand image alignment on the social media is an important yet mostly-overlooked phenomenon. We proposed a scalable Google Cloud Vision-based approach for measuring the alignment between brand identity and brand image, and understanding the brand positions. We analyzed 3247 images of 13 leading Indian brands on Instagram. Images containing wordy announcements by the firms are in stark contrast with the relatively more emotive images by the users. It leads to a noticeable disconnect between the brand identity and brand image. Also, the private sector brands do not always outperform the public sector brands in branding efforts. By offering practical guidance on how to measure and

reduce the misalignment, this study paved a feasible path towards better visual branding on Instagram.

Our major contribution is an integrated framework which processes unstructured text and image data to provide useful brand-related insights.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

S. No.	Abbreviation / Acronym	Full form
1.	API	Application Programming Interface
2.	BAE	Brand Analytics Environment
3.	BERT	Bidirectional Encoder Representations from Transformers
4.	BPS	Brand Personality Scale
5.	BSNL	Bharat Sanchar Nigam Limited
6.	CEO	Chief Executive Officer
7.	CNN	Convolutional Neural Network
8.	Doc2Vec	Document-to-Vector
9.	FGC	Firm-Generated Content
10.	FGI	Firm-Generated Images
11.	GCV	Google Cloud Vision
12.	HDFC	Housing Development Finance Corporation
13.	IDDI	Index of Distance in Distance Image
14.	IT	Information Technology
15.	kNN	k-Nearest Neighbors
16.	LASSO	Least Absolute Shrinkage and Selection Operator
17.	LDA	Latent Dirichlet Allocation
18.	LDA2Vec	LDA-to-Vector
19.	LIWC	Linguistic Inquiry and Word Count
20.	ML	Machine Learning
21.	NRC	National Research Council
22.	OBIM	Online Brand Image

S. No.	Abbreviation / Acronym	Full form
23.	RAM	Random Access Memory
24.	RoBERTa	Robustly Optimized BERT Pretraining Approach
25.	SBI	State Bank of India
26.	SMILE	Social Media Intelligent Learning Environment
27.	SMM	Social Media Macroscope
28.	TCS	Tata Consultancy Services
29.	TF-IDF	Term Frequency – Inverse Document Frequency
30.	Top2Vec	Topic-to-Vector
31.	TPR	True Positive Rate
32.	UGC	User-Generated Content
33.	UGI	User-Generated Images
34.	USA	United States of America
35.	VADER	Valence Aware Dictionary for sEntiment Reasoning
36.	VIF	Variance Inflation Factor
37.	Word2Vec	Word-to-Vector
38.	YOLO	You Only Look Once

Chapter 1
INTRODUCTION

“A brand is no longer what we tell the consumer it is — it is what consumers tell each other it is.” – Scott Cook, Intuit co-founder.

/***/

1.1. Analytics

Analytics is “a set of statistical and operational research techniques, artificial intelligence, information technology and management strategies used for framing a business problem, collecting data, and analyzing the data to create value to organizations” (Kumar, 2017). Ever since the advent of e-commerce sites such as Amazon.com in 1994, there has been an abundance of data on the internet. With the Facebook’s incorporation in 2005, organizations started using text, audio, image and video to position themselves, and shape the consumers’ perceptions on social media networks. Social media is classified into six types: collective projects, microblogs, content communities, social networks, massively multiplayer online games and social virtual worlds (Kaplan & Haenlein, 2010).

Twitter(X) and Instagram are in the category of social networks. More than 500 million tweets on Twitter(X) (Internetlivestats.com, 2023) and more than 95 million images on Instagram (Statista, 2023) are being posted daily. Most of the data being posted on the social media is unstructured data, i.e., information presented without a pre-defined organization.

Social media analytics involves the collection, analysis and visualization of social media data using informatics tools. The specific analytics approach is driven by the business requirements. The desired outcome is the extraction of useful patterns and actionable insights from vast amounts of business data.

Marketing analytics, using the definition of Wedel and Kannan (2016), "involves collection, management, and analysis — descriptive, diagnostic, predictive, and prescriptive— of data to obtain insights into marketing performance, maximize the effectiveness of instruments of marketing control, and optimize firms' return on investment (ROI)". The diagnostic breadth of the marketing analytics is shown in Figure 1.1. It shows an arrow representing the diagnostic breadth, which is a function of mostly structured data generated by the firm and mostly unstructured data generated by the external sources such as users. As the arrow breadth increases, the volume of useful insights would supposedly help the companies' strategic goals.

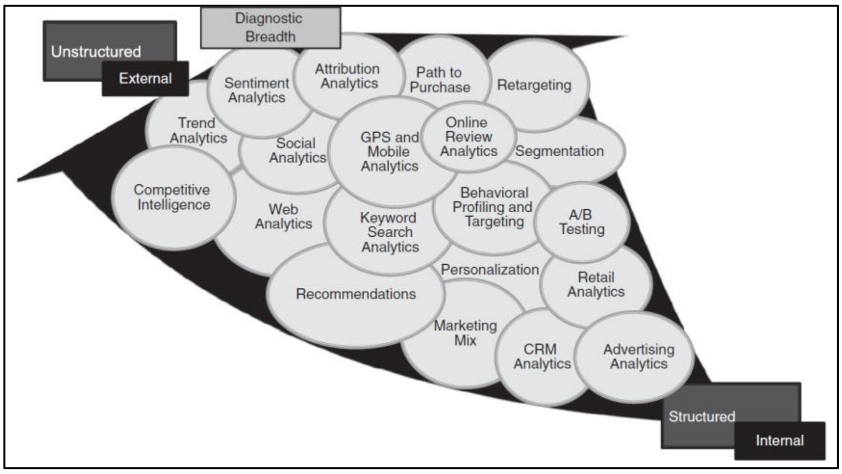

Figure 1.1. Diagnostic breadth of marketing analytics (Source: Wedel & Kannan, 2016)

1.2. Brand management

A brand is a name, term, design, symbol or any other feature that identifies one seller's good or service as distinct from those of other sellers (Dibb et al., 2019). The strategic brand management process, as laid out by Keller and Swaminathan (2019), consists of four main steps shown in Figure 1.2. We focus on some of the aspects related to Step 2 – suggesting how to improve the branding communications on Twitter and Instagram, and Step 3 – measuring and interpreting the branding communications and perceptions on those social media platforms.

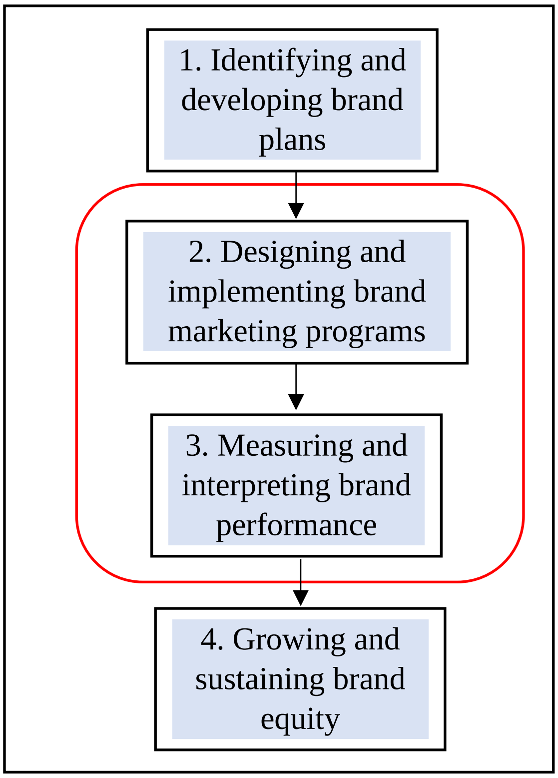

Figure 1.2. Strategic brand management process

Brand management involves several facets; our study is limited to a few aspects of brand identity and brand image shown in Figure 1.3.

BRAND IDENTITY	**BRAND IMAGE**
• How the strategists want the brand to be perceived (Aaker, 2012)	• How the brand is now perceived (Aaker, 2012)
• Communicated brand personality: Sincerity, Competence, Excitement, Ruggedness, Sophistication (Aaker, 1997)	• Perceived brand personality: Sincerity, Competence, Excitement, Ruggedness, Sophistication (Aaker, 1997)
• Brand position: The part of the brand identity and value proposition to be actively communicated to a target audience (Aaker, 2012)	• Brand sentiment: Measure of opinions and emotions (Moussa, 2019; Mostafa, 2013; Smith et al., 2012)

Figure 1.3. Brand identity and brand image

Brand management on social media requires newer ways of collecting and analyzing data. Text analytics has been used in the marketing context for the last 25 years. Image analytics is being seen in marketing studies for less than a decade.

1.3. Text analytics

Text data can be analyzed for understanding and predicting brand personality. Swift analysis of massive amounts of data requires newer methods from machine learning. Machine learning provides the scalability that is nearly impossible with methods that have been traditionally used in social sciences. Berger et al. (2020) catalogued at least 20 different ways of deriving marketing insights from text data. Transformers such as BERT (Devlin et al., 2018) from Google and RoBERTa (Liu et al., 2019) from Facebook revolutionized the way text data is processed in business analytics. A scalable way of analyzing Twitter data is available in the form of Pamuksuz, Yun and Humphreys model (Pamuksuz et al., 2021). It was made possible with the advent of transformers technology. The open-source software, openness to new brand personality dimensions and new data sources makes it an excellent choice for the current study.

1.4. Image analytics

Images can be entertaining or eye-catching, and strike an emotional chord with their audiences. With more than 2 billion users across the world, Instagram is the most important social media platform for posting images. Its largest userbase, around 230 million users, is in India (Statista, 2022). It is popular among the 18-29 age group in India. Brands post many images each week to promote their products and services. These images are known as Firm-Generated Content (FGC). FGC is predominantly positive content that attempts to nurture the brand image. Consumers prefer clear and detailed images instead of textual descriptions (Kane & Pear, 2016). Instagram posts result in more impactful engagement than the Twitter posts and

Facebook posts (Arora et al., 2019). Social media marketing also leads to better brand awareness (Park et al., 2016).

Due to the availability of smart phones and ease of image creation, the consumers frequently post photographs and doctored images about their favorite brands. The images posted by the consumers are known as User-Generated Content (UGC). UGC can convey positive impressions of a brand or dissatisfaction with a brand. UGC also provides valuable information about which consumer segments are using the services. This information in turn allows fine-tuning of the segmentation and targeting activities (Wedel & Kannan, 2016). The UGC on various social media networks has an enormous influence on consumers' pre-purchase thought process (Meppurath & Varghese, 2022; Sangwan & Sharma, 2022). It also positively affects the consumers' intention to buy through social commerce (Shoheib & Abu-Shanab, 2023).

1.5. Brand identity

Brand identity is how the strategists want a brand to be perceived. It is a "unique set of brand associations that the brand strategist aspires to create or maintain" (Aaker, 2012). Brand identity is based on one or more of the four perspectives: brand as a person, brand as a product, brand as an organization and brand as a symbol. One of this study's objectives is to explore the brand-as-a-person perspective.

Brand identity constitutes the following elements (Harris & de Chernatony, 2001; de Chernatony, 1999):

1. Brand vision
2. Brand culture
3. Positioning
4. Personality

5. Relationships
6. Presentations

Signal theory (Spence, 1973) posits that an organization should design messaging that enhances its brand value or equity. Text messages and images posted by a firm to promote its brand are part of the brand identity.

1.6. Brand image

Brand image is “the understanding consumers derive from the total set of brand-related activities engaged by the firm” (Park et al., 1986). Brand image represents how the consumers perceive a brand (Aaker, 2012). It is a “set of brand associations in the minds of consumers” (Mowen & Minor, 2001). Twitter users’ messages and Instagram users’ brand-related images are part of the brand image. The perceptions expressed in the social media posts could be subjective or objective assessments of the brand. Kotler (2000) mentioned that a good brand image could also improve the parent company’s image.

In August 2022, the online food delivery app Zomato launched a new advertisement on Twitter featuring a popular movie star from the Hindi film industry. Zomato’s ambiguous reference to Mahakal restaurant in Ujjain, India angered many Twitter users (Mint, 2022). Thousands of unpleasant tweets and a “#Boycott_Zomato” trend emerged within 24 hours. Zomato noticed that the advertisement is adversely impacting its brand image. Zomato issued an apology and withdrew the advertisement around 48 hours after its launch. This incident illustrates the necessity of monitoring tweets in real-time to manage a brand’s image and sentiment not just on Twitter but on any social media network.

1.7. Need for brand identity – brand image alignment

Maintaining a reasonably good alignment between brand identity and brand image has been one of the top priorities of market strategists. Researchers often use questionnaire-based primary data to quantify the brand identity and brand image. The results from such approaches quickly become outdated, given the large volumes of new data appearing on social media every week.

The comparison of communicated brand personalities and perceived brand personalities on Twitter is an under-researched area, although a text-based personality detection model is now available.

Drawing useful conclusions from image data requires the usage of image processing and computer vision techniques. Computer vision models detect objects, logos, people and emotions in an image with reasonably good accuracy rates. These models need vast amounts of training data during the learning phase. They need to be tuned, i.e., various hyperparameter values need to be tweaked to obtain the optimum performance. The models also need an amount of RAM that is not available on a typical personal computer. Other than Liu et al. (2020), there is little evidence of marketing researchers training neural networks with images.

Many researchers in social sciences and management studies possess neither the technical expertise nor the computational platforms for training their own computer vision models. Nevertheless, they appreciate the need to use sophisticated technology to monitor their visual branding activities. The pre-trained deep learning models such as Google Cloud Vision and Clarifai opened up new possibilities in marketing analytics using brand-related images.

Over the last seven years, a moderate amount of research has been done with images pertaining to brands in developed countries. But the existing literature lacks any specific research on the Indian brands using the images from social media. According to an April 2023 UN report, India has become the most populous country on the planet with a population of 1.425 billion (UN, 2023). Market researchers have immense intense in gaining better insights into what branding strategies work well in the Indian context.

1.8. Service brands

American Marketing Association defined service as “activities, benefits or satisfactions which are offered for sale, or are provided in connection with the sale of goods" (AMA, 1960). Social media marketing should be done differently by different types of brands (Park et al., 2016). Several aspects of services tend to have an intangible nature; they cannot be physically viewed or felt (Keller & Swaminathan, 2019). There has been a debate for long about managing the service brands differently from the product brands (Blankson & Kalafatis, 1999). Bakri et al. (2020) provides sufficient impetus for intensifying the research on service brands using image data. Constructs like service brand identity continue to warrant separate scales of measurement (Pareek & Harrison, 2020).

Service brands are dominating the economy of not just the developed countries (Huang and Dev, 2020) but also that of a developing country like India. In order to cater to the needs of consumers, the customization of services becomes necessary in maintaining the brand identity. What customers tell each other about their experience with the service brands is a part of the brand image. Managers understand the need to align service brand identity and brand image on the social media networks.

1.9. Research Gaps

The following are the research gaps addressed in this thesis. The literature review which resulted in the identification of these research gaps is presented in Chapter 2.

1. As the brand-related communication proliferates through the social media, the existing methods of survey-based brand-related research do not scale with the digital communication. There is a need to develop a framework to manage the brands based on large volumes of diverse social media data in order to complement the primary data-based research. As far as we know, there is no integrated framework which provides

brand-related insights based on text data from Twitter(X) and image data from Instagram.

2. Very little secondary-data research beyond sentiment analysis exists in the area of brand management. In addition to sentiment, there is a need for studying the Indian brands' personality, brand position, identity and image using unstructured data.
3. In the marketing literature, there has been a disproportionate amount of focus on user-generated content. The insights from firm-generated content are also needed.
4. To the best of our knowledge, there is no method for measuring the gap between brand identity and brand image using images.

1.10. Research questions

Based on the research gaps, the following research questions are formulated for further exploration.

1. What is the congruence between communicated brand personality and perceived brand personality of major Indian brands on Twitter(X) ?
2. Does the brand personality impact brand sentiment and brand position on Twitter(X) ?
3. How should the managers measure the extent of brand identity-brand image alignment using unstructured secondary data from Instagram ?
4. Is an integrated machine learning-based framework available for managing the brands on Twitter (X) and Instagram ?

1.11. Research objectives

The research questions are mapped to the following research objectives.

1. To investigate the congruence between communicated and perceived brand personalities of major Indian brands using Twitter(X) data.
2. To study the impact of brand personality on brand sentiment and brand position on Twitter(X).
3. To develop an approach to measure the alignment between brand identity and brand image using Instagram data.
4. To craft a brand management framework based on Twitter(X) and Instagram data.

1.12. Contributions

To address the research objectives, we made a four-fold contribution. First, this is the first study to investigate the gap between the Indian service brands' communicated personality and perceived personality. It revealed a counterintuitive finding that the bots and suspicious accounts on Twitter(X) change the value of sentiment but not the category (positive, neutral or negative) of the sentiment.

Second, we found a weak positive association between the communicated brand personality and brand position.

Third, we proposed a Google Cloud Vision-based approach for understanding the brand positioning, and the brand identity-brand image alignment. We also investigated the contrast between the same-coloured logos and wordy announcements by the brands with the emotive images posted by the users. These differences revealed why a given brand is differently positioned in the clusters based on FGI images and UGI images.

Fourth, we created an integrated brand management framework that currently handles text data from Twitter(X) and image data from Instagram.

The usage of the term "framework" is in line with the contribution types proposed by Samtani et al. (2023). We used an empirics-first approach (Golder et al., 2023). It involves identifying a real-world problem or phenomenon, collecting data and producing useful conclusions for the real-world marketing issues.

1.13. Organization of the thesis

This thesis consists of six chapters with the following contents:

Chapter 1 begins with an introduction of analytics, marketing analytics and social media analytics. It lists the research gaps, research questions and research objectives. It briefly explains our contributions and provides an overview of the organization of the thesis.

Chapter 2 explores the existing literature in the area of marketing and marketing analytics. It explains several newer developments in research using secondary data and unstructured data. It sets the research context and scope for novelty in the areas identified as the main focus of the thesis. It includes a few theories relevant to the current study.

Chapter 3 explains the data sources, the data collected for investigation, concise descriptions of the various data mining, rule-based learning, machine learning and deep learning methods employed in the current study. It offers a rationale for using publicly available pre-trained models for researchers dealing with limited computational resources.

Chapter 4 contains the results of the data analysis in tabular and pictorial form. It provides a detailed explanation of how the brands compare with each other on Twitter and Instagram.

Chapter 5 discusses the methodological and managerial implications of the study. A certain amount of impact of our study would be readily obvious. With minor changes to our work, the brand managers of leading Indian firms could derive additional benefit from our work, as suggested in this chapter.

Chapter 6 concludes the thesis by summarizing the limitations and future research directions. The limitations arose due to the API-related constraints and the technological constraints such as the lack of a full-fledged technology stack for processing the text data in Indian languages. The future research directions necessitate major improvements which would, hopefully, result in proportionately substantial insights for the brand managers.

Though Twitter was rebranded as X sometime in the year 2023, we continue to refer to the social media platform as Twitter in the remaining chapters of this thesis.

Chapter 2

REVIEW OF THE LITERATURE

In this chapter, we provide the background information and relevant literature in the marketing analytics using unstructured data. After an overview of the recent literature, we identify the scope for novel contributions and explain the specific research gaps.

2.1. Recent brand analytics literature using text data

Text data can be obtained from sources such as company websites, company pages on social media, e-commerce websites and the messages posted by the users anywhere on the internet. We review the relevant literature from the years 2014 to 2023 in Table 2.1 and Table 2.2. The progression of research from simple word counts to natural language processing to Bayesian probability models to word embedding models would be evident in our review. It serves as a segue for the research gaps.

Brand sentiment in the form of positive and negative UGC was studied by Tang et al. (2014). They collected the users' comments on 39 brands. They categorized neutral comments as mixed-neutral – containing positive and negative remarks – and indifferent-neutral – containing neither positive nor negative remarks. They hypothesized that mixed-neutral UGC strengthens the effects of positive and negative UGC, whereas indifferent-neutral UGC weakens the positive and negative UGC.

In one of the early studies using vast amounts of text data, Tirunillai and Tellis (2014) examined the reviews of 15 brands. They concluded that the dimensions of quality such as ease-of-use, performance, receptivity, safety, comfort, etc. change not only across brands but also across markets. They also established that the brand positions along those dimensions may change over a period of time. The dimensions are referred to as topics in topic modeling literature. Their study is often cited as a good example of extracting marketing insights from text data using topic modeling, machine learning and natural language processing techniques.

Using Twitter data related to 200 brands, Culotta and Cutler (2016) proposed a social perception score based on a brand's three perceptual attributes, namely, eco-friendliness, nutrition and luxury. They opined that the existing data mining methods at that time are inadequate and developed their own method to estimate the attribute ratings. The estimates were strongly correlated with those calculated from the survey data. They also used the Twitter metadata associated with each tweet to identify the connections among a given brand's users and influencers.

Hewett et al. (2016) looked into a variety of questions around Twitter communications by the firms and by the users. In addition to brief press releases and advertisements, Twitter is also effective for sending personalized responses.

The volume and the sentiment of tweets received a lot of attention, but there was no other form of classification of tweets for a while. However, Kalampokis et al. (2016) applied the brand equity theory to classify the tweets mainly into three categories: brand image, brand satisfaction and purchase intention. Their natural language processing-based approach uses LingpipeDynamicLMClassifier for classifying the tweets. Some tweets such as "I go to IKEA just to people-watch …" do not fall into any of the three categories. Manually classifying tweets into categories such as brand image, brand satisfaction and purchase intention can be a tedious and error-prone task. By developing an algorithm which can classify tweets into those three categories with acceptable accuracy, this study increased the marketing researchers' interest in Twitter data.

The communicated brand personality and perceived brand personality were the subject of Ranfagni et al.'s (2016) study. They collected data from company websites, blogs and

Facebook pages. By using their newly-developed ratios, namely, consumer-brand personality ratio, inter-brand similarity ratio and perceived inter-brand similarity ratio, they made suggestions about how the companies could change their brand differentiation and communication strategies. Their study was limited to fashion brands and did not use Twitter or Instagram data. This study threw light upon a major concern about how companies are losing control over their brands' perceptions on the internet.

Table 2.1. Selected studies in brand analytics using text data (2014-2018)

S. No.	**Authors and Year**	**Method**	**Data**
01.	Tang et al., 2014	Sentiment analysis, Multiple linear regression	Facebook and YouTube comments on 39 brands collected from Facebook and YouTube. Sales data from AutoNews Data Center. Ad spending data from Kantar.
02.	Tirunillai & Tellis, 2014	LDA, MDS, Jaccard Coefficient	Amazon, Yahoo, Epinions. 5 categories. 15 brands. 350,000 reviews over 4 years.
03.	Culotta & Cutler, 2016	Social Perception Score by the authors	Twitter. 239 brands. Connections among 14.6M accounts.
04.	Hewett et al., 2016	Vector autoregression	News articles, Tweets, Consumer sentiment
05.	Kalampokis et al., 2016	LingpipeDynamicLMClassifier, Decision tree	Twitter. 2 brands. 12,122 tweets.
06.	Ranfagni et al., 2016	Consumer-brand personality ratio, Inter-brand similarity ratio, Perceived inter-brand similarity ratio	Style.com, company websites, company Facebook pages. 113 fashion brands.
07.	Liu et al., 2017	LDA and sentiment analysis	Twitter. 5 categories. 20 brands. 1.7M tweets.
08.	Ordenes et al., 2017	OLS regression, Logistic regression, Authors' formulae for sentiments	Consumer reviews and ratings from Amazon.com, TripAdvisor.com, BN.com

Confirming that many tweets are about product, service and promotions, Liu et al. (2017) further justified additional research on Twitter data. They suggested that the analysis of positive and negative tweets should be done separately for each brand in order to obtain a more accurate understanding how consumer sentiments vary across brands within an industry. However, they did not compare the UGC with the FGC.

The impact of tentative language, commissive language and discourse patterns on overall sentiment was studied by Ordenes et al. (2017). Their detailed explanation of sentiment was initially based on consumer reviews and later verified on tweets.

Table 2.2. Selected studies in brand analytics using text data (2019-2023)

S. No.	**Authors and Year**	**Method**	**Data**
01.	Hu et al., 2019	LIWC, ElasticNet regression analysis	Twitter, Glassdoor. 1.9M employee descriptions on Twitter. 312k reviews on Glassdoor. 680k tweets from the organization.
02.	Ibrahim & Wang, 2019	LDA, Sentiment analysis, Network analysis	Tweets of Amazon UK, Argos, Asda, John Lewis, and Tesco. 386000 tweets from a three-month period.
03.	Srivastava & Kalro, 2019	Tobit regression, LDA	Reviews from tripadvisor.com and yelp.com
04.	Yun et al., 2020; Yun et al., 2019; Yun et al., 2018	IBM Personality Insights-based text analytics	Twitter. 3 brands. 622,631 tweets over 7 years.
05.	Mitra & Jenamani, 2020	Sentiment analysis, Co-word network analysis	Amazon. 5 brands. 2500 reviews.
06.	Wu et al., 2023	Chi-squared test, Logistic regression, Naïve Bayes classifier, Support Vector Machine (SVM), SVM with stochastic gradient descent	125,887 tweets from 21 brands.

Hu et al. (2019) made a notable attempt to understand the link between brand personality and consumer personality using firm content, employee content and consumer content. They used tweets and Glassdoor reviews to build a text analytics framework which predicts the perceived brand personality. However, they did not explore the gaps between communicated brand personality and perceived brand personality.

The tweets of five online retailers in UK revealed that most tweets are about delivery, product and customer service (Ibrahim & Wang, 2019). Topic modeling and sentiment analysis suggests that most negative tweets are about delivery and customer service. Network analysis shows the connections between the topics discussed by the customers. This study was especially useful in understanding the reasons for negative sentiments; it shows a network analysis of the negative themes and the reasons that led to the negativity.

Srivastava and Kalro (2019) made use of Elaboration Likelihood Model to understand the helpfulness of online reviews. They were able to examine the impact of content factors, reviewer factors and the latent factors such as clarity and comprehensiveness on the helpfulness of a given review. This study systematically reinforced the notion that not all reviews are created equal, and some reviews are more helpful than others.

In a series of studies on brand personality, Yun et al. (2020), Yun et al. (2019) and Yun et al. (2018) established that a brand's Twitter personality will be more congruent to their followers' personalities compared to non-followers' personalities. Their work evolved into a brand personality detection model (Pamuksuz et al., 2021). It was made available for free as Social Media Macroscope (SMM) by the National Center for Supercomputing Applications at the University of Illinois, Urbana-Champaign, IL, USA. (Please note that the older version of Social Media Macroscope was available at socialmediamacroscope.org until August 31, 2022. The description in this thesis reflects the older version. As of November 13, 2023, the newer version at smm.ncsa.illinois.edu does not offer all the functionality mentioned in this document.)

A new metric – Online Brand Image (OBIM) score – calculated as a product of favorability, strength and uniqueness of brand was developed by Mitra and Jenamani (2020) using the Amazon reviews. They extracted favorability, strength and uniqueness of the brand using text processing techniques such as aspect-based sentiment analysis and co-word network analysis. OBIM's applicability to other text data sources such as Twitter, Facebook and Reddit is yet to be ascertained; however, it is still considered a significant metric for studies where natural language processing is the primary methodology.

Wu et al. (2023) published an interesting investigation of the relationship between message types and customer engagement on Twitter. They found that the customer engagement is higher with socio-emotional tweets and those tweets which followed the dialogic loop principles. This integration of social exchange theory and dialogic theory in the context of anthropomorphic brands involved machine learning methods such as SVM with stochastic gradient descent.

On the whole, the brand-related analysis of text data resulted in contributions in the form of frameworks, methods, metrics and predictive models.

2.2. Brand personality

Brand personality has remained an important construct in the brand management literature (Martineau, 1958). Brand personality is "the act of applying human characteristics or traits to a brand, inducing consumers to think of a brand as if it had human-like qualities" (Aaker, 1997). Brand personality, one of the determinants of brand equity, distinguishes one brand from another (Keller, 1993). Copying product characteristics is relatively easier; copying a brand's personality is harder. A well-crafted brand personality becomes an inimitable asset to an organization. Different brands are well-known for different personality dimensions as shown in Figure 2.1.

Figure 2.1. Brand personality dimensions of well-known brands (Source: Dvornechuck, 2020)

Several scales available to measure the brand personality are listed in Table 2.3. Compared to the other scales proposed by Chen and Rodgers (2006), Geuens et al. (2009), Freling et al. (2011), and Ham and Lee (2015), Aaker's Brand Personality Scale (BPS) is the most dominant scale in the literature. It was also validated in other cultures and contexts (Aaker et al., 2001). Among the 37 brands used by Aaker for developing the BPS, technology, telecom and financial services brand categories were present. Therefore, it is appropriate to use Aaker's BPS in our study.

Table 2.3. Brand personality scales

S.No.	**Scale Name**	**Dimensions**
01.	Aaker's brand personality scale (Aaker, 1997)	Sincerity, Competence, Excitement, Ruggedness, Sophistication
02.	Website personality scale (Chen & Rodgers, 2006)	Intelligent, Fun, Organized, Candid, Sincere
03.	Geuens's brand personality scale (Geuens et al., 2009)	Responsibility, Activity, Aggressiveness, Simplicity, Emotionality
04.	Brand personality appeal scale (Freling et al., 2011)	Favorability, Originality, Clarity
05.	Internet media personality scale (Ham & Lee, 2015)	Intelligent, Amusing, Confusing, Convenient, Sociable

Aaker's BPS defines brand personality using 5 dimensions, 15 facets and 42 traits. The dimensions of sincerity, competence, excitement, ruggedness and sophistication are listed in Table 2.4.

Table 2.4. Aaker's brand personality scale

Dimension →	**Sincerity**	**Excitement**	**Competence**	**Sophistication**	**Ruggedness**
Facets and traits →	*Down-to-earth*: down-to-earth, family-oriented, small-town *Honest*: honest, sincere, real *Wholesome*: wholesome, original *Cheerful*: cheerful, sentimental, friendly	*Daring*: daring, trendy, exciting *Spirited*: spirited, cool, young *Imaginative*: imaginative, unique *Up-to-date*: up-to-date, independent, contemporary	*Reliable*: reliable, hard working, secure *Intelligent*: intelligent, technical, corporate *Successful*: successful, leader, confident	*Upper class*: upper class, glamorous, good looking *Charming*: charming, feminine, smooth	*Outdoorsy*: outdoorsy, masculine *Tough*: tough, rugged

Some brands such as Harley-Davidson personify one dimension, namely ruggedness, more strongly than others. Other brands such as Google emphasize more than one dimension, namely competence and sophistication, in their brand personality.

Upon deciding a brand personality, it is essential to convey the intended personality on various communication channels including the social media such as Twitter, Facebook and Instagram. The communications may involve text, audio, video and presence in virtual worlds such as Metaverse. Our focus is on understanding the brand personality using tweets posted by the brand and by the Twitter users.

The opinions on Twitter quickly influence the others in their network. They in turn affect the brand sentiment and brand personality as perceived by the users.

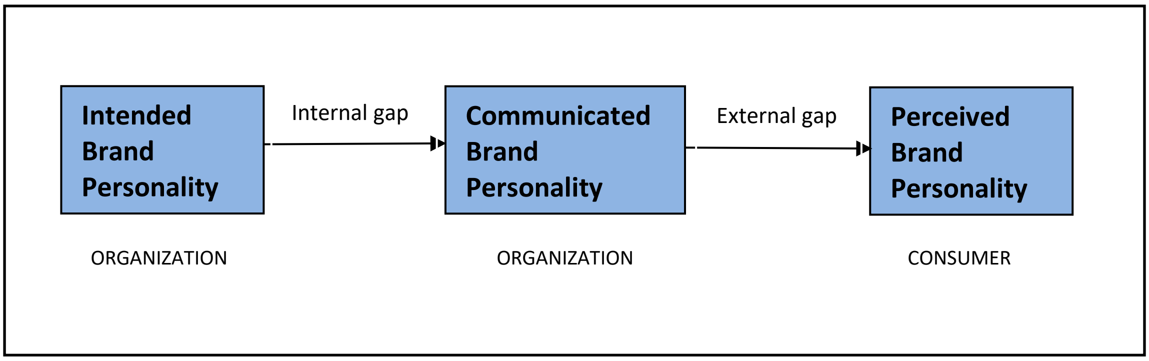

Figure 2.2. Brand personality gaps

Based on the self-congruity theory (Sirgy, 2018; Sirgy, 1982), it is evident that consumers tend to gravitate towards a brand whose personality aligns with their own personality (Arora et al., 2023; Kim et al., 2008). As illustrated in Fig. 2.2, Masiello et al. (2020) facilitated a better understanding of the differences among intended, communicated and perceived brand personalities. A brand management team designs an intended brand personality and communicates it on various media. It might not always succeed in ensuring that the intended personality is the same as the communicated personality. The perceived brand personality depends on the perceptions of the users as well as the conversations among users (Kuksov et al., 2013). The personality communicated on social media and the personality perceived by the users should be measured and monitored continually. Any substantial difference between the communicated personality and perceived personality on the social media would require timely intervention by the brand.

The brand personality maintained on the websites, blogs, microblogs, review websites, discussion boards and social networks has been of interest to the researchers. Ghorbani et al.

(2022) published a meta-analysis of 107 studies that appeared in the literature between the years 2005 and 2021.

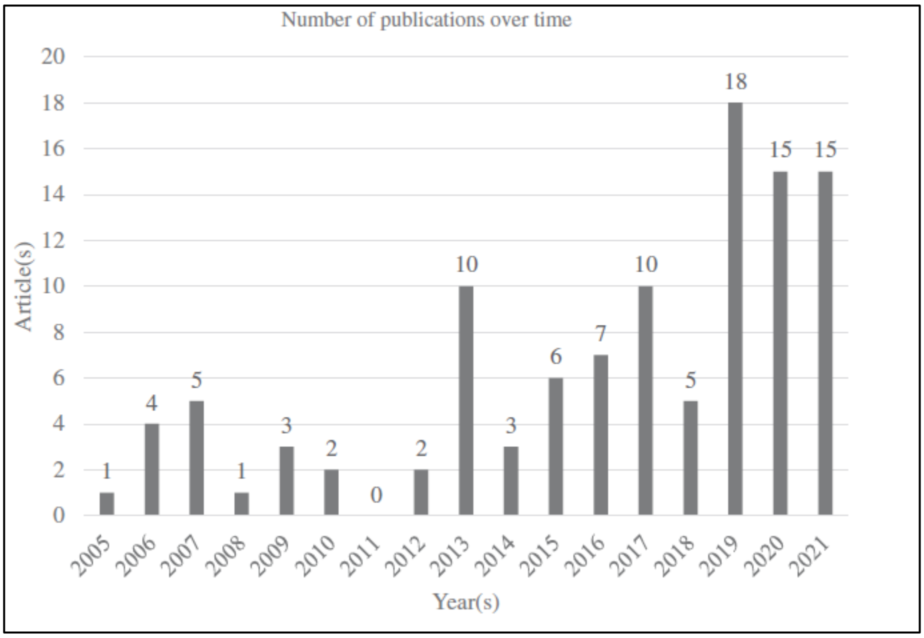

Figure 2.3. Digital brand personality literature (Source: Ghorbani et al., 2022)

Based on Figure 2.3, the sustained interest in investigating the digital brand personality strengthens the need for further research in this area.

2.3. Brand sentiment

Sentiment is a measure of opinion and emotion about a brand (Moussa, 2019). Sentiment tracking on social media has become an integral part of brand management after the advent of Amazon and Facebook. Sentiment is being computed by the brand managers based on the UGC in the form of comments, reviews, tweets and emojis available on social media (Zhang et al., 2022). Sentiment analysis is "the computational treatment of opinion, sentiment, and subjectivity in text" (Pang and Lee, 2008.). Sentiment analysis is also referred to as opinion mining or emotional polarity analysis in the literature (Mostafa, 2013).

Given that a single negative comment could sometimes have major impact on the brand image, companies have been continuously monitoring the social media content. Proactively pre-empting the negative brand sentiment was the subject of Cooper et al. (2019) study. The brand sentiment comments are not exclusively based on loyalty, product failure, customer service dissatisfaction, etc.; consumers might also express negative sentiment if they perceive a violation of ethics, morals or individual rights.

Depending on the method, the sentiment in a given piece of text is expressed as a number or a noun. It could be expressed on a scale of -10 to 10, -1 to 1, etc. or it could be summarized with a single noun such as 'Anger', 'Joy', 'Pleasure', 'Trust', 'Sadness' and 'Disgust'. The best-known study on brand sentiment is a pioneering study by Smith et al. (2012) that used two apparel brands from the USA to classify the consumer emotions as positive, neutral or negative. The choice of method varies with the source of text data and the type of outcome – numerical or categorical – desired by the researcher.

In addition to its marketing relevance, brand sentiment could also serve as a predictor of stock price performance (Schweidel & Moe, 2014).

2.4. Brand position

As part of the brand position, a firm conveys its value proposition and tries to differentiate itself from the competing brands. The value proposition can be based on functional benefits, emotional benefits or self-expressive benefits.

Computing brand position provides valuable information to the firm about whether its messaging is in harmony with its marketing strategy. Despite a firm's intention to maintain a positive brand position in its social media messages, the position could also be unintentionally neutral or negative. If measured periodically and corrected whenever necessary, the brand position could make a brand stand out in a crowded field of players.

Many brand sentiment studies based on UGC are available, but very little research was undertaken on the FGC. We therefore set out to calculate the brand position based on FGC. The brand position values depend on the method chosen to calculate them. With VADER method, the brand position values in the range of -1 to +1 throw light on how the various brands are positioning themselves on Twitter.

2.5. Marketing analytics using images

There have been attempts to classify images with seemingly simple methods such as naïve Bayes classifier and logistic regression. Those methods do not scale well if the training needs to be done with a large number of images. Deep learning approaches fare relatively better, provided they are run on a powerful computational platform. Some of the notable studies in image-based brand analytics are listed in Table 2.5.

An image could contain a brand logo or product or person or all of them. A set of engagement parameters – brand logo, person, product, sentiment, etc. – were defined by Mazloom et al. (2016). Their investigation of UGC images in the fast-food category using GCV revealed that the presence of a product or a person in an image could drive up the popularity of that Instagram post. However, a brand logo does not necessarily make an image more popular. They developed a support vector regression-based model which would predict the popularity of a given image based on the engagement parameters present in it.

With the image contents extracted using Clarifai, Jaakonmäki et al. (2017) trained a model that predicts the potential popularity of the images. They used the Least Absolute Shrinkage and Selection Operator (LASSO) regression to predict the user engagement. User engagement was measured by counting the number of likes and number of comments. User engagement was the dependent variable; image class, emoji type, words in the image, the users' age, posting time, posting day and number of followers were the independent variables. Their data was from German-speaking countries and its generalizability to other countries is arguable.

Table 2.5. Recent studies in brand analytics using image data

S. No.	Authors and Year	Method	Data Sources	Sample Size
01.	Mazloom et al. (2016)	GCV, Support Vector Regression	Instagram	75000 posts
02.	Jaakonmäki et al. (2017)	Clarifai API, LASSO regression	Instagram	13396 images
03.	Ferwerda & Tkalcic (2018)	GCV, Spearman's correlation	Instagram	54962 images
04.	Klostermann et al., 2018	Sentiment analysis and clustering	Instagram	1 brand. 10375 posts containing images, captions and tags.
05.	Moussa, 2019	Emoji-based brand sentiment score by the author	Twitter	18 brands. 720 tweets.
06.	Liu et al., 2020	SVM, Convolutional Neural Network (CNN)	Flickr (training), Instagram (testing)	56 brands. 186,000 images.
07.	Nanne et al., 2020	Computer vision, based on neural networks	Instagram	24 brands. 21,000 images.
08.	Arabadzhyan et al., 2021	GCV, IDDI distance measure	Instagram	860,000 images of European island destinations

Ferwerda and Tkalcic (2018) used GCV in their study which investigated whether a relationship exists between the personality traits of an Instagram user and the contents of the images posted by that user. It appears that people with certain personality traits include certain objects and emotions in an image. It allows the brand managers to create a personalized visual messaging based on the user type.

Klostermann et al. (2018) developed a practical way of using tags, captions and images to analyze a brand's perceptions on Instagram. They studied around 10375 images of a single brand, McDonald's, using GCV and sentiment analysis. They created associative networks of product-related images to demonstrate the strength of associations among the product attributes. Their work is limited to analyzing the user-generated content but the novelty lies in using both images and text to provide information about the consumers' brand perceptions.

Emoji is also a form of an image, albeit with relatively lesser space for expressiveness. By extracting emojis pertaining to 6 product brands on Twitter, Moussa (2019) developed an emoji-based brand sentiment metric. This metric helps the companies assess the users' sentiments towards their respective brands. The newly developed score's positive correlation with American Consumer Satisfaction Index (ACSI) scores served as a reflection of its validity.

A novel "visual listening in" approach to measure how brands are being perceived on social media was proposed by Liu et al. (2020). Classifying the images in the categories of apparel and beverages on the basis of attributes such as glamorous, fun, rugged and healthy, they trained a neural network. With the trained model, they calculate a brand portrayal metric for a given brand's images. They also demonstrated how a brand manager could create brand perceptual maps. This study is very different from other studies in that it did not use a pre-trained neural network model; it involved training a neural network with brand-related images and tuning it to achieve good accuracy levels. Since the trained network is not publicly available, its usage in other studies currently seems to be infeasible.

The emergence of multiple deep learning options necessitated investigation into the appropriateness of each approach. In order to help the brand managers make informed choices in visual branding, Nanne et al. (2020) provided a comparative analysis of Google Cloud Vision (Google, 2022), Clarifai (Clarifai, 2023) and You Only Look Once (YOLO) version 2 (Redmon et al., 2016). While Clarifai generated 20 labels per image, GCV generated an average of 7.87 labels per image. Despite the lower number of average labels per image, the accuracy of GCV was reported to be higher than that of Clarifai (Nanne et al., 2020). YOLO v2 produces fewer labels of lower accuracy than the other two approaches. So, we deemed it appropriate to develop a GCV-based approach for comparing the brand identity with brand image.

In a destination marketing study pertaining to European islands, Arabadzhyan et al. (2021) investigated how close is one destination to another destination in terms of users' perceptions. They developed a distance measure named Index of Distance in Destination Image (IDDI) to compare the similarity between images. IDDI is a good metric for cross-sectional and longitudinal analysis of images. The study also correlated the changes in perceptions to internal events in the organizations and external shocks such as climate changes in those destinations.

All the aforementioned studies involved North American and European brands. An image-based investigation of Indian brands was not undertaken so far.

2.6. Indian brands in the literature

Marketing studies about the country of product brand exist in literature (Srinivasan et al., 2004; Nebenzahl et al., 1997; Samiee, 1994). In our study, we consider a brand which has its origins in India as an Indian brand. A local brand strives to present itself as a "son of the soil" (Cayla & Eckhardt, 2007). Its implicit claim is that it understands the local customer better than a global brand.

The words and actions of consumers in a country are influenced by its culture. The most influential work on national cultures is the study by Geert Hofstede (1980)'s which posits that nations differ based on five dimensions: individualism versus collectivism, power distance, masculinity versus femininity, uncertainty avoidance, and long-term orientation versus short-term orientation. The U.S. scored 91 on individualism when compared to India's score of 48.

India's collectivistic culture was also mentioned by Triandis (1995) study. India's collectivistic attitudes drive the way the consumers respond to the online branding efforts. This makes the alignment of brand identity and brand image in India a vastly different exercise compared to the Western markets such as the U.S. or European Union. Sharma et al. (2022) established the interaction between cultural elements and brand identity factors in the Indian context.

Table 2.6 lists some of the brand-related studies in India. Some of the studies explored the performance of global brands in India. A few other studies have been predominantly about the product brands in India. Systematic investigation of the Indian service brands needs further consideration.

Table 2.6. Recent brand-related literature in the Indian context

Authors and year	**Brands**	**Focus of analysis**
Hsieh (2004)	25 car brands	Global brand equity in 20 countries including India
Lieven & Hildebrand (2016)	20 product brands	Impact of brand gender on brand equity in 10 countries including India
Heinberg et al. (2018)	55 consumer goods brands	Brand equity of local consumer goods in India and China
Zarantonello et al. (2020)	More than 100 consumer goods global and local brands	Relationship between consumer-based brand equity and market share in 29 countries including India
Sengupta et al. (2022)	3 apparel brands	Affinity for global brands over an Indian apparel brand
Yadav et al. (2022)	1 Indian FMCG brand	Aspect-level sentiment analysis of a product brand
Bhatia et al. (2023)	1 Indian product brand	Factors affecting the virality of consumer brand sabotage

There was a call for further research by Klostermann et al. (2018) on the “sets of brands to study similarities and dissimilarities”. While Nanne et al. (2020) investigated a few sets of product brands, they did not study any service brands. Since both those studies used user-generated content, they could not explore the crucial aspect of aligning the brand identity with the brand image.

2.7. Need for an empirics-first study on Indian service brands

An empirics-first approach differs from a theory-first approach by choosing a real-world phenomenon, processing real-world data and producing valid marketing implications (Golder et al., 2023). It does not always refine existing theory or develop new theory. For instance, the findings based on machine learning models tend to be associational instead of causal in nature. A notable example in recent times is the mining of Twitter data by Rust et al. (2021) to monitor how the stakeholders feel about the world's top-100 brands. They created a brand reputation tracking mechanism based on a longitudinal study. An empirics-first approach appears to be an appropriate choice for exploring the image analytics in marketing.

Consumer perceptions depend on the firm's communications and on the consumers' conversations about that brand (Kuksov et al., 2013). There have been no studies that focused exclusively on the Instagram images about Indian brands. Since its largest userbase is in India, Instagram becomes a bigger priority than Facebook and Twitter. It is not sufficient to merely post images on Instagram; Indian brands need a way of empirically measuring the gap between FGI images and UGI images.

The extant literature on product brands revealed useful information about how to analyze images. There does not seem to be an image-based study of the services brand, possibly due to the inherent nature of images posted by services brands; the lack of products in the images makes it difficult for a computer vision model to decipher what is being communicated to the target audience.

Our work is different from Klostermann et al. (2018) and Nanne et al. (2020) that focused on product brands. We not only investigated service brands but also proposed a replicable method for monitoring the alignment between brand identity and brand image.

2.8. Research context and scope for novelty

Faff (2013) suggested an approach for identifying the research context and novelty. In Prof. Faff's suggested approach, circles representing the areas of research are used in Venn diagrams. There is no fixed notion of what these circles should represent; they could denote markets or regulations or data or technology. If the overlapped areas are largely under-investigated, they would present opportunities for new research. If the zone of triple intersection is an area where novel contributions are yet to be made, the researchers could focus on such an area.

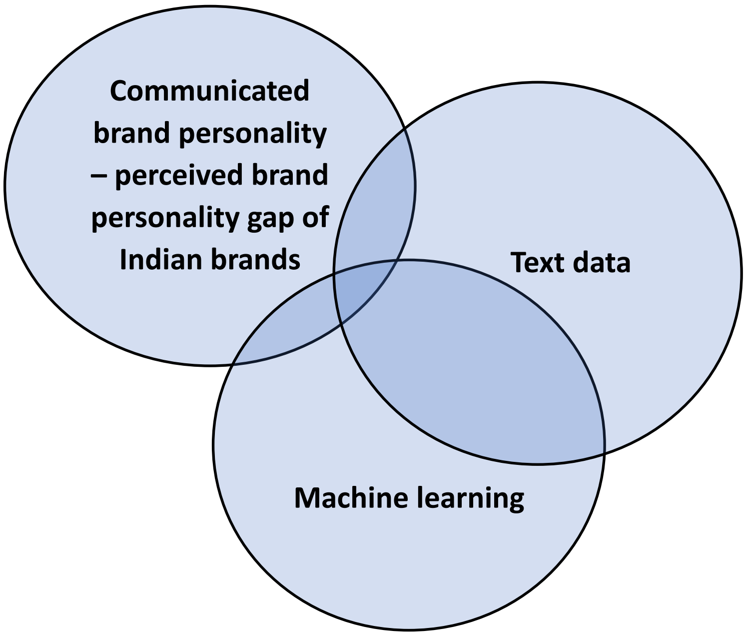

Figure 2.4. Research context for the current study in text analytics

Figure 2.4 illustrates the scope for novel contributions at the intersection of brand personality, text data and machine learning. There is nearly 25 years of research available using text data and machine learning in the marketing context. However, once we add the circle representing the 'communicated brand personality – perceived brand personality gap of Indian brands' to the Venn diagram, the scarcity of research at the triple intersection presents us with an opportunity to address this research gap.

Similarly, Figure 2.5 helps understand the scope for research at the intersection of brand identity – brand image gap, image data and machine learning. The research using image data and machine learning in the marketing context started less than a decade ago. The intersection of those two circles alone has a vast scope for further contributions. After adding the 'brand identity – brand image alignment' circle to the Venn diagram, there does not seem to be any published research at the triple intersection of these circles to the best of our knowledge.

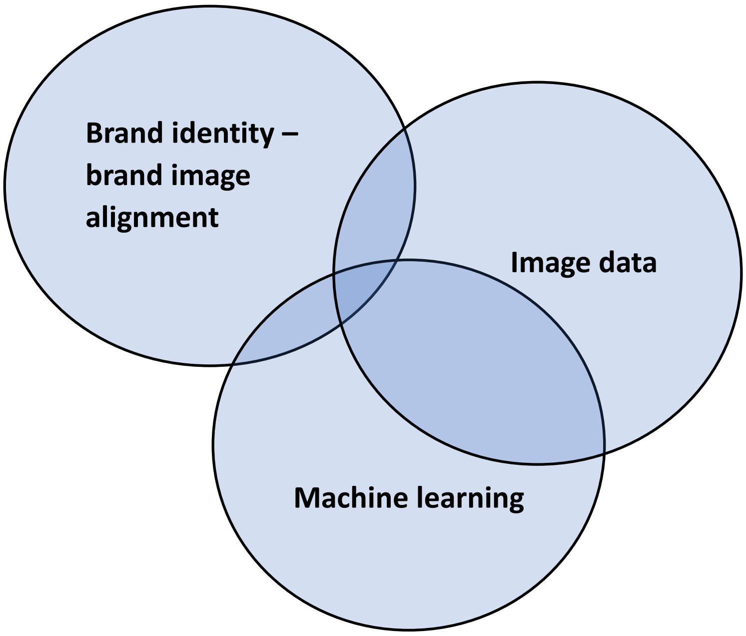

Figure 2.5. Research context for the current study in image analytics

2.9. Positioning of the current study

Though the novelty of our study is fairly clear based on Faff's approach, we perused the studies on the 13 selected Indian brands that appeared in Scopus-indexed and ABDC-indexed journals over the last seven years. Table 2.7 presents a juxtaposition of the existing literature with the current study.

The only study which investigated the brand personality of one of the selected brands is by Srivastava and Sharma (2016). They used survey data to study the brand personality of Airtel. Sentiment analysis of 7 of the selected brands was done by a few studies using website data and Twitter data (Ahmad et al., 2021; Trivedi & Singh, 2021; Sindhani et al., 2019; Ranjan et al., 2018). Brand strength of 3 technology brands was explored using survey data and structural equation modeling by Ray and Sharma (2022).

Table 2.7. Existing literature vis-à-vis current study on the 13 selected brands

Authors and Year	**Brands**	**Data**	**Method / Tool**	**Brand Personality**	**Brand Sentiment**	**Brand Position**	**Brand Identity-Brand Image Alignment**
Srivastava & Sharma, 2016	AirTel	Survey	Confirmatory factor analysis	✓			
Ranjan et al., 2018	Axis Bank, Canara Bank, HDFC Bank, SBI	mmb.moneycontrol.com	Textblob		✓		
Sindhani et al., 2019	Swiggy, Zomato	Twitter	TIBCO Spitfire		✓		
Ahmad et al., 2021	Jio	Twitter	Sentiment analysis		✓		
Trivedi & Singh, 2021	Swiggy, Zomato	Twitter	NRC emotion lexicon		✓		
Ray & Sharma, 2022	TCS, Infosys, Wipro	Survey	SEM				
Current study	13 Indian brands	Twitter, Instagram	Deep learning, rule-based learning	✓	✓	✓	✓

The contrast between those studies and our study is evident. Our study on these 13 brands is novel because we use not only unstructured text but also images to investigate communicated and perceived brand personalities, brand sentiment, brand position and brand identity – brand image alignment.

2.10. Theories used in the context of unstructured data

Appendix A, Table A1 lists 46 theories that were cited in the context of marketing studies using unstructured data (Balducci & Marinova, 2018). Different theories are relevant to different research designs and types of data. We elaborate upon three theories that are connected with our study.

1. Self-congruity theory

Sirgy (1982) interpreted the self-congruity theory in the consumption context. Prof. Sirgy explained the self-concept as a "totality of the individual's thoughts and feelings having reference to him as an object" while discussing the aspects of consumer behavior. The self-congruity theory suggests that the consumers lean towards products or services which are largely aligned with their self-concept. Therefore, the self-congruity theory has become important for the managers who try to make the consumers identify themselves with a brand.

Brand loyalty could be strengthened by finding ways to decrease the incongruence between the consumer personality and brand personality (Batra, Ahuvia & Bagozzi, 2012). This is tied to our first research objective which measures the alignment between communicated brand personality and perceived brand personality.

2. Social influence theory

Social influence theory (Fromkin, 1970) states that people build their own opinions based on the consensus of the groups which they inhabit in social settings. The words of others on the social media can have a substantial impact on a consumer's behavior. The conformity pressure might, at times, be dominated by the uniqueness pressure which makes some consumers deviate from the group consensus.

The effects of social influence theory were investigated by Sridhar and Srinivasan (2012) and Ordenes et al. (2018). This theory is highly relevant to all our research objectives pertaining to Twitter and Instagram.

3. Speech act theory

Speech act theory was proposed by Austin (1962) and further developed by Searle (1969). It delves into how language isn't just about conveying information—it is also about performing actions. Searle categorized speech acts into five types: assertive, commissive, directive, declaratory and expressive.

Speech act theory pertains to the marketing communications using text, image, audio and video data. Ordenes et al. (2017) and Ordenes et al. (2018) tested the hypotheses about how the various categories of speech act theory present in text messages influence their ratings or their valence. This theory is tied to our second research objective.

Chapter 3

RESEARCH METHODOLOGY

We begin this chapter by reiterating our research objectives. We explain the criteria for brand selection, the selected brands, and the methods used for processing the text and images.

3.1. Research objectives

Based on an extensive perusal of brand analytics literature, we identified the following research objectives. Although mentioned in Chapter 1, they are repeated here for ready reference.

1. To investigate the congruence between communicated and perceived brand personalities of major Indian brands using Twitter data.
2. To study the impact of brand personality on brand sentiment and brand position.
3. To develop an approach to measure the alignment between brand identity and brand image using Instagram data.
4. To craft a brand management framework based on Twitter and Instagram data.

3.2. Brand selection

To avoid concerns of biased or arbitrary selection, our research relied on a few reputed lists of the top brands in India. London-based Kantar Group and Brand Finance study the Indian brands every year. They release a list of top brands based on market share, market cap, sustainability and other parameters. For our study, we considered Kantar's *Top 75 Most Valuable Indian Brands 2020* (Kantar, 2020), Kantar's *India's Most Purposeful Brands 2021* (Kantar, 2021) and Brand Finance's *India 100* (Brand Finance, 2021) list. Table 3.1 lists the brands included in our study. We included the years-in-existence in order to see if it directly correlates with superior brand personality or brand sentiment.

Table 3.1. Selected Indian brands

Brand Category	**Brand Name**	**Years in Existence**
Telecom	AirTel	28
	BSNL	23
	Jio	16
Online food delivery	Swiggy	9
	Zomato	15
Banking	Axis Bank	30
	Canara Bank	117
	HDFC Bank	29
	SBI	68
Technology	Infosys	42
	TCS	55
	TechMahindra	37
	Wipro	78

Kantar's lists and Brand Finance's list contain 15 to 100 brands. So, we selected the brands based on five criteria:

1. A mix of public sector and private sector brands
2. Service brands which are hard to analyze but need more exploration
3. Diverse sectors of brands
4. Top-tier, mid-tier and bottom-tier brands in each of the four sectors wherever feasible
5. Old players and relatively newer players in a sector

Our criterion 4 is essential for testing whether the deep learning model works well across the spectrum of brand ranks in a given category. To put it differently, the model should detect variance in the data and must not place all the brands in a single cluster.

We chose 13 brands from 4 categories: AirTel, Bharat Sanchar Nigam Limited (BSNL) and Jio in telecom; Swiggy and Zomato in online food delivery; Axis Bank, Canara Bank, HDFC Bank and State Bank of India (SBI) in banking; Infosys, Tata Consultancy Services

(TCS), TechMahindra and Wipro in information technology (IT) and IT-enabled services. BSNL, Canara Bank and SBI are public sector companies, and the rest are from the private sector. Swiggy and Zomato are similar to GrubHub and DoorDash in the USA. We verified that all 13 brands maintain official brand pages on Twitter and Instagram. There are a considerable number of posts by the brands and conversations among the users about these brands.

3.3. Data

3.3.1. Twitter data

Established in the year 2006, Twitter is one of the oldest and active social media networks. Twitter was rebranded as X in the year 2023. We chose Twitter as one of the data sources for studying Indian brands because of the following reasons. First, Twitter has more than 20 million active users in India (Statista, 2023). Second, the major Indian brands maintain an active presence on Twitter in the form of official brand pages. Third, Twitter provides an Application Programming Interface (API) that allows retrieval of tweets without resorting to scraping or any other questionable means. Fourth, Twitter blocks suspicious accounts and labels dubious messages as 'MISLEADING.' The misleading posts quickly disappear from Twitter, thereby providing a relatively cleaner corpus of data which, of course, could still require further clean-up. The noise and the ambiguity in some of the Twitter messages notwithstanding, the tweets are a valuable source of information about a brand.

Our unit of analysis is a tweet posted by the firm or the user. The data used in our study initially comprised of 85,800 tweets collected using Social Media Macroscope (SMM) between April 03, 2022 and June 01, 2022. There are no scraped tweets in our dataset. We collected 41,600 tweets from the official brand pages to compute the communicated brand personality and brand position. We assumed that the posts on the official brand pages on Twitter do not require any clean-up.

3.3.1.1. Clean-up of the tweets posted by users

To understand the perceived brand personality and brand sentiment, we collected 44,200 tweets from the conversations among Twitter users. The presence of bots on Twitter is an indisputable fact. A bot is a partially or fully automated account which posts messages based on scripted patterns. Some bots are malicious and some are not. For instance, a harmless bot which posts messages on the beneficial effects of retirement planning or physical exercise would not be blocked by Twitter.

The bots and spam accounts are detected based on the following textual and behavioral aspects in their posts (Mukherjee et al., 2013):

- Abnormal text patterns
    - Many positive or many negative messages
    - Word distribution
    - Message length
    - Grammar/spelling patterns
- Atypical and suspicious behavior
    - Deviation from the consensus
    - Many messages per day
    - Content similarity

Among the free tools available for detecting spam accounts and bots, we chose OSoMe Botometer® API (Davis et al., 2016) developed by Indiana University. (It should be noted that OSoMe Botometer is not functional as of June 30, 2023 due to the deprecation of an endpoint at Twitter. It is unclear if a new version of OSoMe Botometer would be available at a later time.)

We removed 1625 bot accounts and 2536 tweets based on abnormal patterns in language and behavior. Among those removed, some are bot accounts and some are considered "suspicious accounts" because their account activity matches that of a bot. (The list of bot accounts and accounts with bot-like behavior are available upon request.) We computed the perceived brand personality and brand sentiment using the remaining 41,664 tweets by 27,909 unique users. Since each of these users posted at least one tweet, our sample defies the oft-stated view that most tweets are from a small number of users (Wu et al., 2011). The number of tweets per brand is in Table 3.2.

Table 3.2. Twitter data of selected Indian brands

Brand Name	**Tweets by the Firm**	**Tweets by the Consumers** (*before* removing suspicious accounts and bots)	**Tweets by the Consumers** (*after* removing suspicious accounts and bots)
AirTel	3200	3400	3289
BSNL	3200	3400	3252
Jio	3200	3400	3227
Swiggy	3200	3400	3242
Zomato	3200	3400	3207
Axis Bank	3200	3400	3182
Canara Bank	3200	3400	3319
HDFC Bank	3200	3400	3268
SBI	3200	3400	3238
Infosys	3200	3400	3143
TCS	3200	3400	3153
TechMahindra	3200	3400	3050
Wipro	3200	3400	3094
TOTAL	**41600**	**44200**	**41664**

Twitter API does not provide demographical information about its users. So, it is not possible to provide the profile of user groups in our sample in terms of age, gender, education and income. The location of a few users is available via Twitter API if those users chose to disclose their location. Although this study is limited to Indian brands, it should be noted that a Twitter user residing anywhere on the planet might have commented on these brands. Our tweets sample is location-agnostic.

3.3.2. Instagram data

Many major Indian brands maintain official pages on Instagram. The brands' followers post images and short videos known as reels to express their opinions about brands. The images are often accompanied by tags and comments. We did not include the reels, tags and captions in our study. Instagram's Application Programming Interface (API) allows legitimate retrieval of images.

Our unit of analysis is an image posted by the firm or the user. To protect the privacy of Instagram users, we did not retrieve any user's demographic information or personally identifiable information. We collected around 6000 images between September 01, 2022 and Nov 14, 2022.

3.3.2.1. Clean-up of Instagram data

Brands routinely post an image repeatedly in order to advertise a new service offering or a new discounted offer. Some images posted by users do not reveal any useful information about the tagged brand. We removed 2753 duplicate images and uninformative images. A couple of deleted images are shown in Figure 3.1 and Figure 3.2.

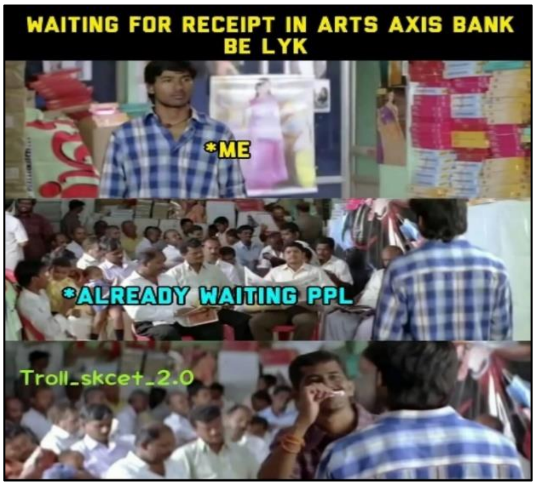

Figure 3.1. Deleted meme image with 'AxisBank' tag

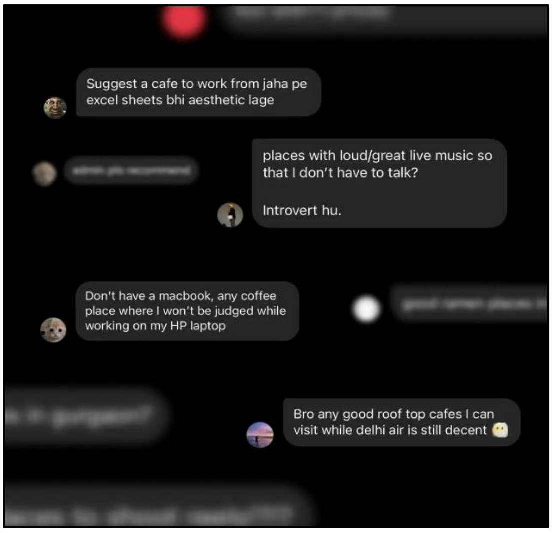

Figure 3.2. Deleted uninformative image with 'Zomato' tag

Our study uses the remaining 3247 images. The corpus includes 655 images from the official brand pages for studying the brand identity and 2592 images posted by the users for studying the brand image. The per-brand image counts are listed in Table 3.3. A few sample images are shown in Figure 3.3.

Table 3.3. Instagram data of selected Indian brands

Brand Category	Brand Name	Years in Existence	Images Posted by the Brand	Images Posted by the Users
Telecom	AirTel	28	52	205
	BSNL	23	59	238

	Jio	16	50	208
Online food delivery	Swiggy	9	52	264
	Zomato	15	50	261
Banking	Axis Bank	30	51	201
	Canara Bank	117	54	208
	HDFC Bank	29	51	203
	SBI	68	54	224
Technology	Infosys	42	46	225
	TCS	55	54	88
	TechMahindra	37	49	48
	Wipro	78	33	219
Total number of images			**655**	**2592**

The images posted by the firms using their respective official accounts are included in the FGI. Other images about a brand posted by Instagram users in their individual capacity are included in the UGI. This delineation of FGI and UGI is in line with the earlier studies by Colicev et al. (2019) and Liu et al. (2020). Another set of images is shown in Figure 3.4.

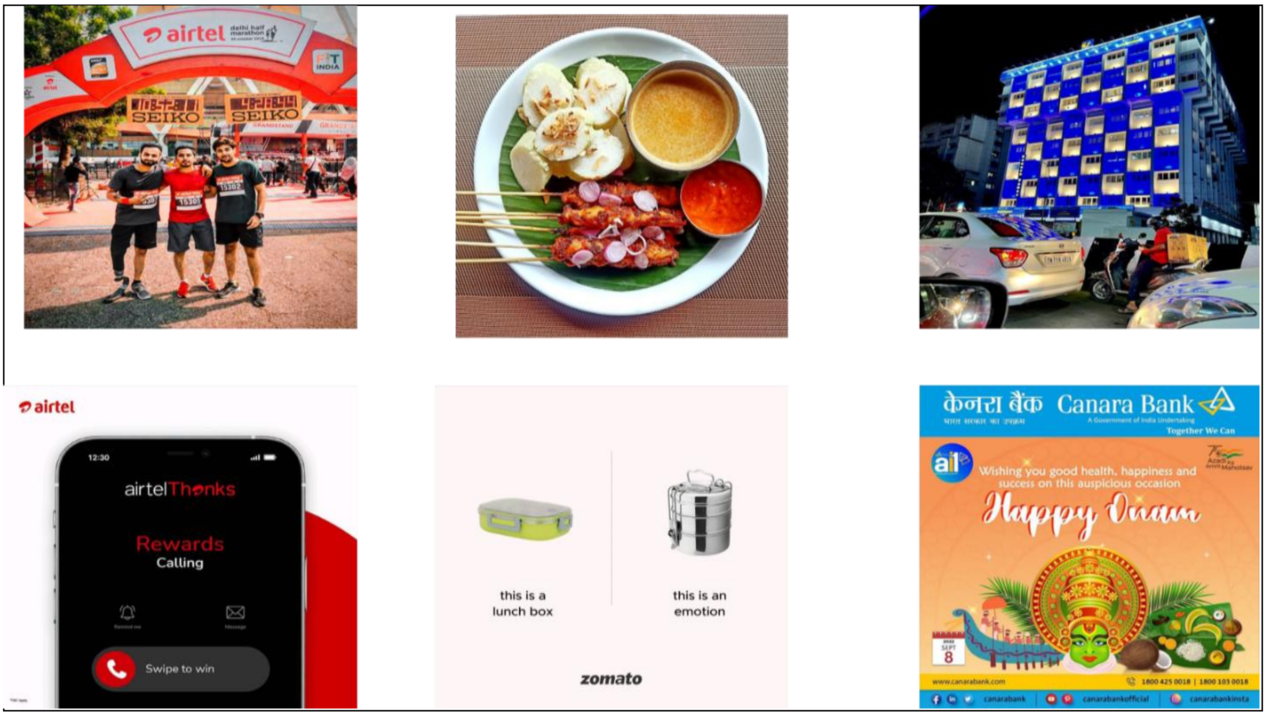

Figure 3.3. Sample images – I [clockwise from the top-left]: AirTel UGI, Zomato UGI, Canara Bank UGI, Canara Bank FGI, Zomato FGI, AirTel FGI

Will UGI contain posts by the employees too ?

There exists a valid concern of employees posting a lot of positive content about a brand using their personal accounts on Instagram. This possibility could potentially blur the distinction between FGI and UGI. As a matter of corporate policy, the employees could be refrained from posting any images that tarnish the reputation of their respective brands. But it is unlikely that a company would stop its employees from posting an occasional harmless message or positive message about the brand using his/her personal account.

Figure 3.4. Sample images – II [clockwise from the top-left]: Jio FGI, Infosys FGI, Infosys UGI, Jio UGI

Arabadzhyan et al. (2021) used an approach of randomly sampling images and manually checking their authors' antecedents. They found around 8% of images to be from the people managing the destinations. Theirs is a non-scalable and somewhat infeasible approach in other settings. Also, the Instagram users are not required to identify themselves as working for any firm at all. The existing literature provides very little evidence to corroborate the unverified presumption that brand image is largely controllable by a firm's employees.

3.4. Methods

RESEARCH OBJECTIVE #1

3.4.1. Pamuksuz, Yun and Humphreys model: Brand personality

University of Illinois at Urbana-Champaign developed the Social Media Macroscope (SMM) (Yun et al., 2020). SMM offers Brand Analytics Environment (BAE) for computing the communicated brand personality (Yun et al., 2019; Yun et al., 2018) and Social Media Intelligent Learning Environment (SMILE) for computing the perceived brand personality (Yun et al., 2018). SMM was trained using 266,105 posts from Facebook and Twitter. It was tested using 53,221 posts. A few recent studies used SMILE to analyse data from Twitter and Reddit (Donelson et al., 2021; Gesing et al., 2021; Nagarkar et al., 2021).

The computation model shown in Figure 3.5 is adapted from Pamuksuz et al. (2021) and enhanced with brand position and brand sentiment computation.

The brand personality computation uses the following steps:

1. SMM accepts a brand name and collects up to 3200 tweets in one attempt from the last two weeks. It can also collect the tweets that mentioned a given brand name. For example, to collect the tweets posted by users about Infosys, we specify @Infosys as the search string. The raw data does not contain any brand personality scores. SMM cleans the raw data by removing stop words and punctuations. It performs stemming on the tweets.

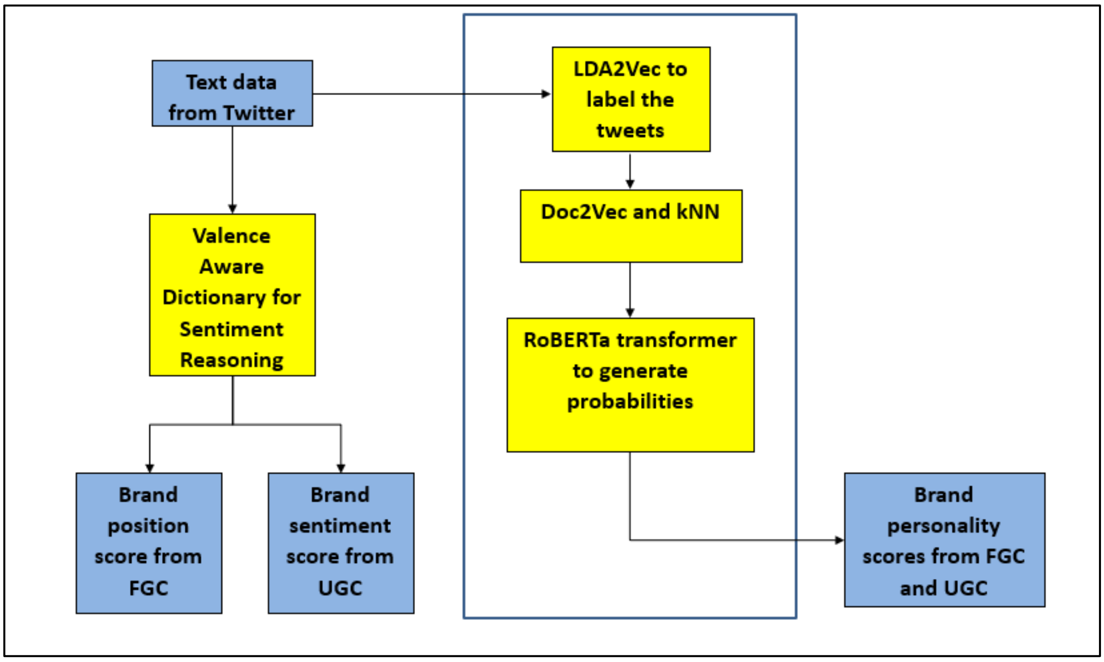

Figure 3.5. Brand personality, position and sentiment computation

2. SMM uses LDA2Vec to label each tweet with a personality dimension label. LDA2Vec combines Latent Dirichlet Allocation (LDA) (Blei et al., 2003) and Word2Vec (Mikolov et al., 2013). LDA forms clusters of words. Word2Vec determines the similarities between words based on word embeddings in the neighborhood of a given word. SMM uses Aaker's BPS mentioned in Table 2.4 and Opoku's brand personality dictionary listed in Appendix A, Table A2 to assign a label to the tweets. If a tweet is not labeled with any personality dimension, it is considered to be a NULL tweet.
3. In order to label the tweets with more than one personality dimension, the model uses Doc2Vec method (Le & Mikolov, 2014) and k-nearest neighbors approach. Doc2Vec is a logical extension of Word2Vec which produces word embeddings in the form of numerical vectors. Each tweet is treated as a document. Doc2Vec depends on document embeddings. In Doc2Vec, the context in which words are used in document 1 are remembered and used in document 2's embeddings and so on. Applying Doc2Vec on a collection of tweets produces vectors and it is possible to identify which vectors are similar or dissimilar to

which other vectors. Using the output of Doc2Vec in conjunction with kNN classifier, the model identifies the k-nearest neighbors of a given tweet and applies the neighbors' labels to the unlabeled or under-labeled tweet. This facilitates the detection of nearest labeled tweets to a given NULL tweet, thereby converting many of the NULL tweets into labeled tweets. It also increases the percentage of tweets labeled with multiple dimensions.

4. Facebook developed Robustly Optimised Bidirectional Encoder Representations from Transformers pre-training approach (RoBERTa) (Liu et al., 2019). RoBERTa's bidirectional processing of sentences results in better understanding of the context of words. RoBERTa is a large language model. The pre-trained model used in our framework was trained with 160 GB text data. It uses dropout regularization to minimize overfitting. SMM uses RoBERTa to generate the probability of each personality dimension in a text message.

The above steps 1 to 4 describe the training and testing of SMM suite. The personality dimension labels attached by SMM for a few tweets are listed in Table 3.4. The average communicated brand personality dimension scores of the 13 brands generated by BAE are listed in Table 4.1 in Chapter 4. The average perceived brand personality dimension scores generated by SMILE are listed in Table 4.5 in Chapter 4.

Table 3.4. Personality dimensions in tweets: A few examples

Tweet	**Personality Dimension**
The announcement of a strategic partnership between @AirtelIndia & @Tech_Mahindra to set up a joint #5G innovation lab for developing #MakeinIndia 5G use cases for the Indian & global market is a laudable initiative towards fostering R& D in telecom sector.	*Excitement, Competence* [in a tweet about AirTel, which received **high** excitement and competence scores]
@[name redacted] Sorry for the inconvenience caused, BSNL assures prompt action to resolve your grievance.	*Sincerity* [in a tweet about BSNL, which received a **high** sincerity score]
.@Brent_Council partnered with Infosys to modernize their legacy applications using #OracleCloud solutions to enable a cloud	*Sophistication* [in a tweet about Infosys, which received a **high** sophistication score]
@[name redacted] Hi there! We did not mean to keep you waiting for your order and our team should have assisted you better. We have checked and can see that delivery partners are currently not available at your location. We know how upsetting this can be.	*Ruggedness* [in a tweet by Swiggy, which received a **low** ruggedness score]

3.4.2. Validation of text labeling

In order to validate the labels assigned to tweets by the Social Media Macroscope, we chose to submit 0.5% of the tweets for validation. Existing precedents in the literature and a few practical considerations related to time and financial resources led us to choose 0.5% tweets. It amounts to a random sample of 416 tweets from the corpus of 83264 tweets. We submitted those tweets to two human validators on Clickworker.com crowdsourcing platform for validation. We paid each validator approximately €15 for this task. Validation involved 21 short surveys of approximately 20 tweets each. A sample invoice related to one of those surveys is shown in Appendix A, Figure A1.

The effectiveness of text labeling could be evaluated using the inter-rater percent agreement metric and the true positive rate metric. In the current context, the inter-rater percent agreement between the two raters is the percentage of labels which was assigned by both the raters to the tweets (McHugh, 2012). The inter-rater percent agreement for the validated data was 0.1495. It means that 14.95% labels were used by both the validators. There is no rigid cut-off value pertaining to the inter-rater percent agreement value and it should be reported as-is.

The agreement between the labeling by the deep learning model and the human raters is measured using the True Positive Rate (TPR).

$$\text{True Positive Rate} = \frac{1}{N} \sum_{i=1}^{N} (SMM_i == H_i) \tag{1}$$

For a tweet *i*, SMM_i is the label assigned by the Social Media Macroscope and H_i is the label assigned by the human rater. The average TPR for the text validation was 0.7103 across all the brands. Upon verifying that the SMM labels for the tweets have an acceptable level of accuracy, we used the labels for the remaining analysis.

RESEARCH OBJECTIVE #2

3.4.3. VADER: Brand sentiment

Hutto and Gilbert (2014) developed a sentiment analysis method known as Valence Aware Dictionary for sEntiment Reasoning (VADER). It is a rule-based learning method which outperformed some of the machine learning methods such as naïve Bayes classifier and support vector classifier on Twitter data. It might not be the best choice for other text data such as reviews.

VADER uses the following hand-crafted rules to quantify the sentiment in a piece of text:

1. Punctuation increases the magnitude of intensity.
2. Capitalization increases the magnitude of intensity.
3. Degree modifiers increase or decrease the magnitude of intensity.
4. “But” signals a shift in the sentiment.

5. Negation flips the sentiment.

VADER uses a gold-standard lexicon which is an improvement over those used by methods such as Linguistic Inquiry and Word Count (LIWC). It was validated by the humans on Amazon Mechanical Turk. It understands the context better than LIWC. Its lexicon is available for download at the website of Georgia Institute of Technology. VADER reportedly outperformed the human classification. Its F1 score was 0.96 and the F1 score of human classification was 0.84.

VADER generates the sentiment scores on a normalized scale of -1 to 1. A score of 0.05 and higher is considered as positive sentiment; a score of less than -0.05 is negative sentiment; a score between -0.05 and 0.05 indicates neutral sentiment.

3.4.4. VADER: Brand Position

We used VADER to also compute brand position from the firm-generated tweets. A score of 0.05 and higher is considered as positive brand position; a score of less than -0.05 is negative position; a score between -0.05 and 0.05 indicates neutral position.

RESEARCH OBJECTIVE #3

An overview of the image analytics performed on the brand images is illustrated in Figure 3.6. Beginning with the Instagram images as the input, the objective is to generate brand clusters that show brands in different light based on firm's and users' images.

3.4.5. Image labels' extraction: Google Cloud Vision

While suggesting new research directions for marketing analytics, Wedel and Kannan (2016) recommended Google's machine learning solutions for interpreting and classifying images on Instagram. Google Cloud Vision (GCV) is a pre-trained machine learning model that was extensively trained with a wide variety of images. It is especially useful for the researchers in business analytics and marketing analytics who might not have access to powerful computational resources for training the computer vision models. It detects people, objects, logos, fonts, handwriting and emotions in a given image. There is no publicly available information to assert that it was trained using brand-related images.

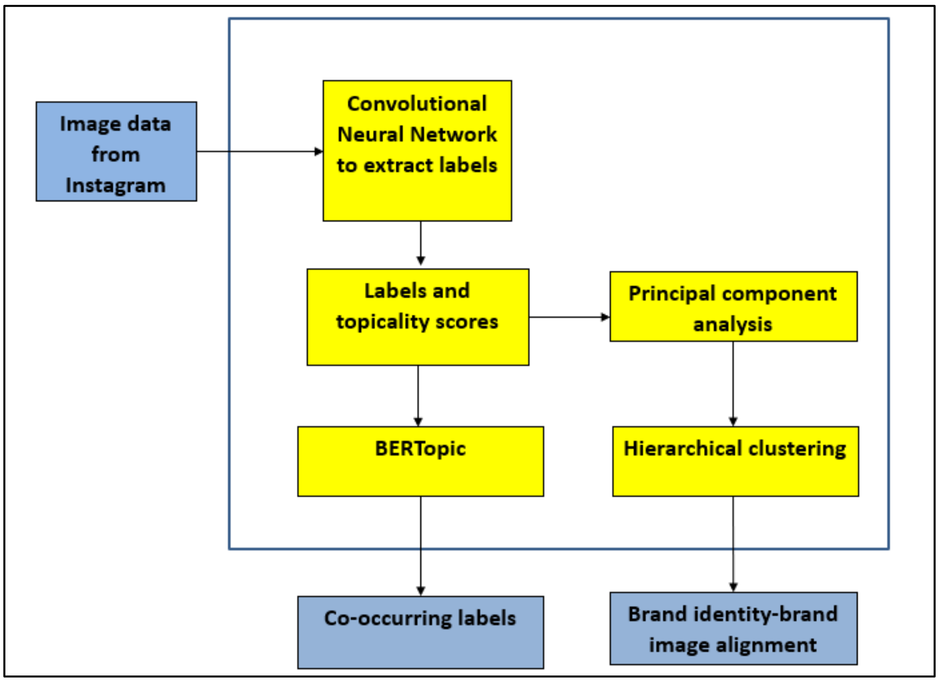

Figure 3.6. Brand-related image analytics

Using the Google Cloud Vision API involves the following broad steps:

1. Linking the Google Cloud account to a billing account. It requires payment by the user or credits from Google Cloud.
2. Setting up a project with the necessary virtual machine and storage buckets.
3. Obtaining the authentication keys from Google Cloud.
4. Enabling Google Cloud Vision API access.
5. Making the API calls, locally on the cloud or remotely from a personal computer, using languages such as C# or Python to extract labels describing a given image. Making API calls on the cloud requires uploading all the images to the cloud,

thereby incurring storage costs. However, the results could be obtained faster when the images are in the cloud storage buckets.

We chose to invoke the API from a laptop. It helped us avoid the cloud storage costs. This resulted in a relatively slower retrieval of results from the cloud. It is not the only way to use GCV API. But it is a trade-off which is appropriate for resource-constrained researchers.

6. Saving the labels and their corresponding topicality scores for further analysis. Other studies such as Klostermann et al. (2018) used topicality scores for investigating some of the brand-related objectives. Converting labels to numerical representations and using them in the quantitative analysis is another valid approach.

As is the case with its commercial competitors, Google does not provide any specific details of the machine learning algorithm, number of images used for training or the hyperparameter values. Revealing these details might possibly erode the competitive advantage of Google. However, the opacity of the GCV did not preclude the researchers from using GCV; instead, it facilitated quick, relatively easy and replicable analysis of images. Without pre-trained public models such as GCV, the researchers setup their own neural networks and sometimes generate non-replicable results.

For each of the 3247 images, GCV provided 7 to 10 labels that describe the contents or emotions in those images. Table 3.5 lists a break-up of the 31,378 labels used in the rest of the data analysis.

The labels, in their original form, could be used in the qualitative analysis involving topic modeling and verbal sentiment analysis. GCV also provided a topicality score, ranging between 0 and 1, for each image label. These topicality scores can be used as the inputs for quantitative analysis.

Table 3.5. Number of image labels per brand

Brand name	**Number of image labels**
AirTel	2501
BSNL	2847
Jio	2472
Swiggy	3104
Zomato	3023
Axis Bank	2486
Canara Bank	2480
HDFC Bank	2381
SBI	2726
Infosys	2714
TCS	1387
TechMahindra	944
Wipro	2313
Total	31378

The correctness of the GCV-generated labels can be understood by taking an image posted by Infosys depicted in Figure 3.7. It is evident that all the labels generated by GCV are not necessarily correct. In the sample image, there is no seesaw, no musical instrument and no recreational activity. This partially erroneous aspect of the GCV output warrants validation of labeling by human beings.

Figure 3.7. A sample image and its GCV-generated labels

3.4.6. Validation of image labeling

All the deep learning-based computer vision models have a certain amount of erroneous output. GCV does not claim all its labels to be 100% relevant to the given image. Before using the GCV output as the input to any other method, the labels should undergo a systematic validation following the extant precedents in the literature. Validation should be done by anonymous validators on a reputable third-party platform. We evaluated appen.com and clickworker.com for the validation task. Based on the price, ease of use and responsiveness of the support teams, we chose to use clickworker.com for validation.

The validation need not be done for all the labeled images. Per the precedent set by the researchers, a random sample of 0.5% to 2.5% is seen in the literature. The number of images submitted to validation is also determined by the time and financial resources at the researcher's disposal. In order to validate the labels assigned to images (Everingham et al., 2010) by Google Cloud Vision, we chose a random sample of 2.5% images from the corpus of 3247 images. We submitted those 82 images to two human validators on clickworker.com crowdsourcing platform for validation. We paid each validator approximately €8 for this task.

Validation involved 4 short surveys of approximately 20 images each. A sample invoice related to one of those surveys is shown in Appendix A, Figure A2.

The effectiveness of image labeling can be evaluated using the inter-rater percent agreement metric and the true positive rate metric. In the current context, the inter-rater percent agreement between the two raters is the percentage of labels which were assigned by both the raters to the images (McHugh, 2012). The inter-rater percent agreement for the validated data was 0.1952. It means that 19.52% labels were used by both the validators.

The agreement between the labeling by the deep learning model and the human raters is measured using the True Positive Rate (TPR).

$$\text{True Positive Rate} = \frac{1}{N} \sum_{i=1}^{N} (GCV_i == H_i) \qquad (2)$$

For an image i, GCV_i is the label assigned by the Google Cloud Vision and H_i is the label assigned by the human rater. The average TPR for the image validation was 0.7958 across all the brands. Upon verifying that the GCV labels for the images have an acceptable level of accuracy, we used the labels for the remaining analysis.

3.4.7. Most frequent, most co-occurring and most-important labels

3.4.7.1. Most co-occurring labels: BERTopic

In order to interpret the overall contents of a brand's images, a topic modeling approach is required. Topic modeling methods such as Latent Dirichlet Allocation (Blei et al., 2003) and Latent Semantic Analysis (Dumais et al., 1988) are often employed for extracting topics from text. BERTopic (Grootendorst, 2022) is a recent method that generates using document embeddings, generates clusters of those embedded vectors and extracts unnamed topics. As the embedding techniques get more refined, BERTopic's architecture might generate better clusters of topics.

While the Latent Dirichlet Allocation (LDA) mainly depends on Bayesian probabilities, BERTopic employs transformers to accomplish the topic modeling. BERTopic is easier to train and easier to use than LDA. When benchmarked with three publicly available datasets, BERTopic's performance was found to be competitive with that of LDA, Non-Negative Matrix Factorization (Fevotte & Idier, 2011), Contextualized Topic Modeling (Bianchi et al., 2020), and Top2Vec (Angelov, 2020).

3.4.7.2. Most important labels: Term Frequency-Inverse Document Frequency

The most commonly occurring words can be obtained using word frequencies and word clouds. But our objective is not limited to finding the most frequent words. We used the Term Frequency-Inverse Document Frequency (TF-IDF) statistic which provides the most important words associated with a brand's images. The Term Frequency (TF) provides information about how frequently a word occurs in a document. In the current context, a tweet is considered as a document. The Inverse Document Frequency (IDF) allocates lower weight to the frequently occurring words and higher weight to the infrequently occurring words.

3.4.8. Brand clusters: Hierarchical clustering

Each of the 13 brands is represented by 97 to 311 images and each image has 7 to 10 topicality scores. In order to reduce the dimensions, we applied principal component analysis on the topicality score numerical vectors. For the sake of brevity, the principal components of FGI images and UGI images are not listed here.

On the first seven principal components generated for each brand, we applied hierarchical clustering using Hellinger's distance. Proposed in the year 1909, the Hellinger's distance is known to be a better distance measure than other techniques such as correspondence analysis in measuring the distance between distributions (Rao, 1995). For two brand topicality vectors p and q, the Hellinger's distance is computed as shown in equation 2.

$$\mathrm{HDist}_{pq} = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2} \qquad (3)$$

In order to compare the distributions of data that emanates from labels, Hellinger's distance is an appropriate choice (Ansari et al., 2018; Culotta & Culter, 2016). Other distance measures such as Canberra distance and absolute correlation distance are also occasionally used for constructing the clusters.

A recently introduced distance measure called Index of Distance in Distance Image (IDDI) appears to have been specifically built in the realm of destination branding (Arabadzhyan et al., 2021). The images posted for destination branding are different from those posted for services branding. Services images are mostly infused with affective dimensions, whereas the destination images also include cognitive and conative aspects.

The uptake of IDDI is entirely in the arena of tourism and destination marketing so far. IDDI's applicability to the service brands in our study is currently unclear. So, we did not empirically evaluate it in our study. A full-fledged comparison of the clusters based on many different distance measures is not within the scope of our study.

RESEARCH OBJECTIVE #4

3.4.9. A machine learning-based framework using text and images

For managing the brands on social media, a machine learning-based framework using text images is illustrated in Figure 3.8.

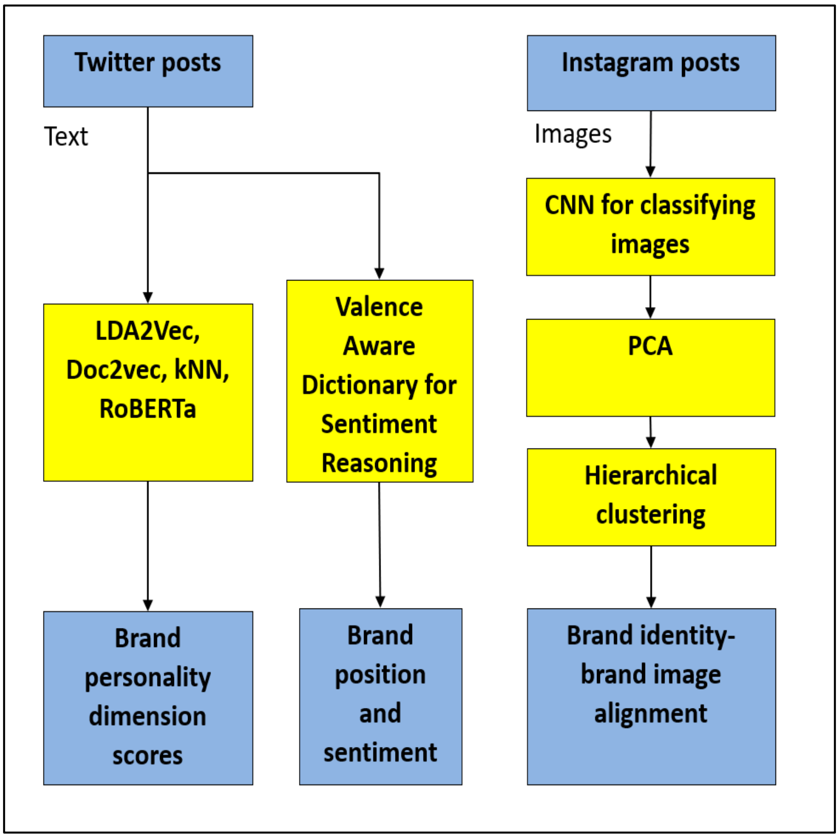

Figure 3.8. A machine learning-based framework using text and images

This framework employs several types of learning:

1. LDA2Vec: Dictionary-based approach combined with unsupervised learning
2. Doc2Vec: Unsupervised learning
3. kNN: Supervised learning
4. RoBERTa: Semi-supervised learning
5. VADER: Rule-based learning
6. CNN: Semi-supervised learning
7. PCA: Unsupervised learning
8. Hierarchical clustering: Unsupervised learning

Samtani et al. (2023) mentioned five types of contributions using deep learning in information systems research: 1. Domain-specific, 2. Representation, 3. Learning, 4. System, Framework or Workflow, 5. Innovation-Accelerating. The framework in Figure 3.8 is a type-4 contribution. It integrates deep learning into the brand management process to meet specific objectives.

Chapter 4

DATA ANALYSIS AND RESULTS

In this chapter, we provide a detailed analysis of the results. These results help us understand the relative performance of brands on Twitter and Instagram.

4.1. Brand personality

4.1.1. Communicated brand personality

The communicated brand personality dimension scores are generated by the Social Media Macroscope. These scores are shown in Table 4.1. Each number in this table represents the probability of a given dimension's presence in the corpus of tweets pertaining to a brand. For example, 0.0018 is the average probability of the presence of sophistication dimension in the tweets posted by AirTel. Most of the brands scored well in sincerity and competence dimensions. It is in line with an extant study by Eisend and Stokburger-Sauer (2013) which found that the mature brands might find it beneficial to focus more on sincerity and competence than on other brands.

Table 4.1. Communicated brand personality scores

Brand Name	Sophistication	Excitement	Sincerity	Competence	Ruggedness
AirTel	0.0018	0.0444	**0.4190**	0.1250	0.0314
BSNL	0.0002	0.0381	**0.5073**	0.1632	0.0171
Jio	0.0508	0.1208	0.0646	**0.1244**	0.0372
Swiggy	0.0081	0.0423	**0.1603**	0.0835	0.0333
Zomato	0.0105	0.0791	**0.1079**	0.0342	0.0159
Axis Bank	0.0218	0.0675	**0.2626**	0.1571	0.0323
Canara Bank	0.0047	0.0129	**0.2870**	0.0942	0.0367
HDFC Bank	0.0118	0.0745	0.1239	**0.2296**	0.0426
SBI	0.0016	0.1201	0.0872	**0.1561**	0.1027
Infosys	0.0224	0.2335	0.1093	**0.2999**	0.0887
TCS	0.0179	0.1849	0.1073	**0.2635**	0.0472
TechMahindra	0.0355	0.2038	0.0913	**0.3192**	0.0897
Wipro	0.0209	0.2454	0.0881	**0.3411**	0.0794

Note: Top-scoring dimension of each brand is in bold numerics.

Among the telecom brands, Jio beat its competitors in sophistication and excitement dimensions. BSNL focuses more on sincerity, competence and ruggedness. AirTel is among the four big players in the oligopoly of Indian telecom market. AirTel has been a major player in the private sector for much longer than Jio. Its communicated brand personality is not better than that of Jio in many dimensions. AirTel's poor scores would be disheartening to its brand managers. They could review its position by starting with a comparison of AirTel's tweets with those of Jio and BSNL. We could not include the fourth major brand, VI (Vodafone-Idea), due to the unavailability of adequate number of tweets at the time of our data collection.

The phenomenon of a young brander outsmarting an older competitor is also evident in other categories. Swiggy is more sincere, competent and rugged than its older competitor Zomato. Zomato scores well in sophistication and excitement. Among the banks, HDFC outperforms SBI in sophistication, sincerity and competence. SBI has a lot more bank branches than HDFC and its wide footprint helped SBI do well in the ruggedness dimension.

Table 4.2. Category-wise mean scores of communicated brand personalities

Category	**Sophistication**	**Excitement**	**Sincerity**	**Competence**	**Ruggedness**
Telecom	0.0176	0.0677	0.3303	0.1375	0.0285
Online food delivery	0.0093	0.0607	0.1341	0.0588	0.0246
Banking	0.0099	0.0687	0.1901	0.1592	0.0535
Technology	0.0241	0.2169	0.0990	0.3059	0.0762

From the category-wise mean scores listed in Table 4.2, the technology brands have the highest mean values in the dimensions of sophistication, excitement and competence. The online food delivery brands have the lowest mean values in all dimensions except sincerity. This result is surprising, because Swiggy and Zomato are the type of brands that are expected to exude qualities like excitement and ruggedness. These two brands need to refine their communications on Twitter to move towards higher scores.

4.1.2. Communicated brand personality clusters

Brand clusters are one of the ways of understanding the brand personality data. We generated clusters based on the communicated brand personality dimension scores listed in Table 4.1. Since the data is not too well-separated, k-means clustering did not seem to be an appropriate clustering approach. We used agglomerative hierarchical clustering to generate the clusters of brands. Hierarchical clustering method can correctly produce non-globular clusters from poorly separated data. Another advantage is that it does not require a pre-determined number of clusters.

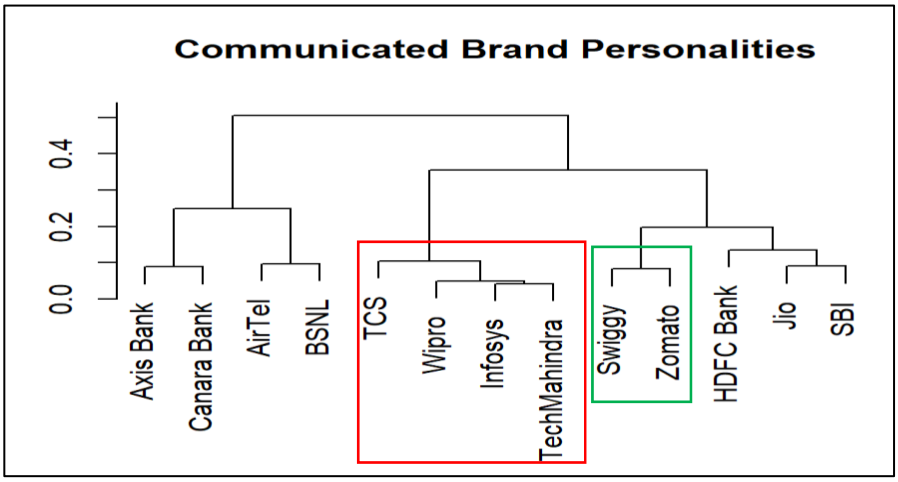

Figure 4.1. Communicated brand clusters

In Figure 4.1, the clustering method shows the communicated brand personalities of all technology brands are neatly grouped into one cluster and online food delivery apps are in another cluster. The banks and telecom brands are however scattered across two clusters. This depiction shows that the personality computation model worked reasonably well in extracting the brand personality dimensions from nearly 41,600 tweets.

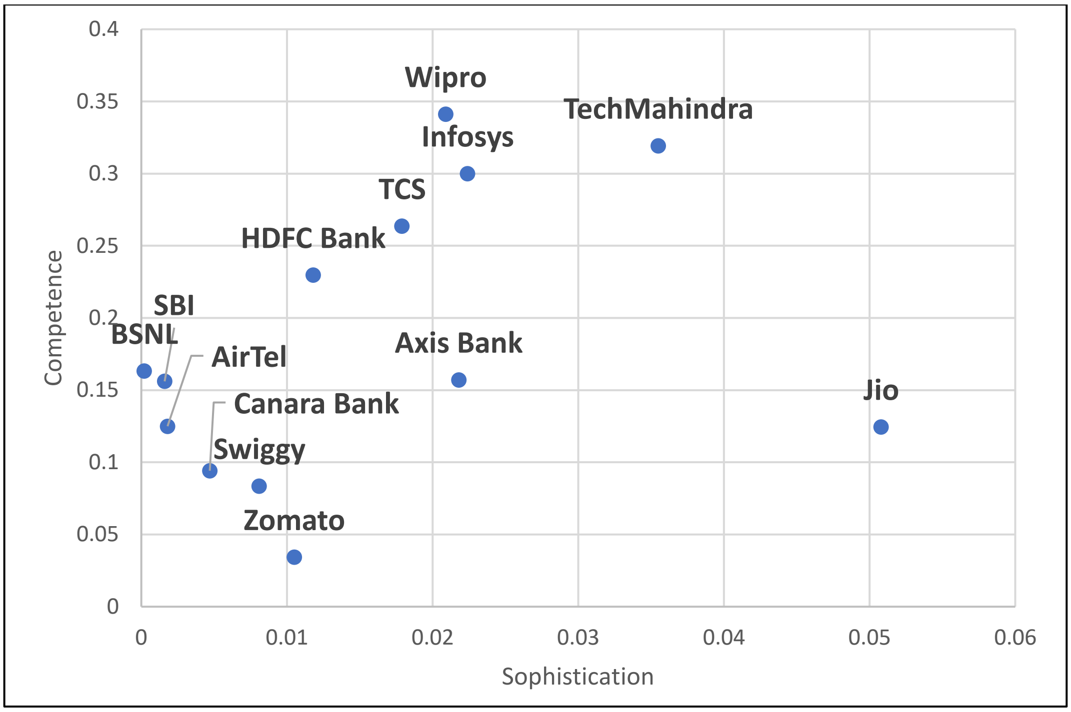

Figure 4.2. Communicated brand personalities: Sophistication vs. Competence

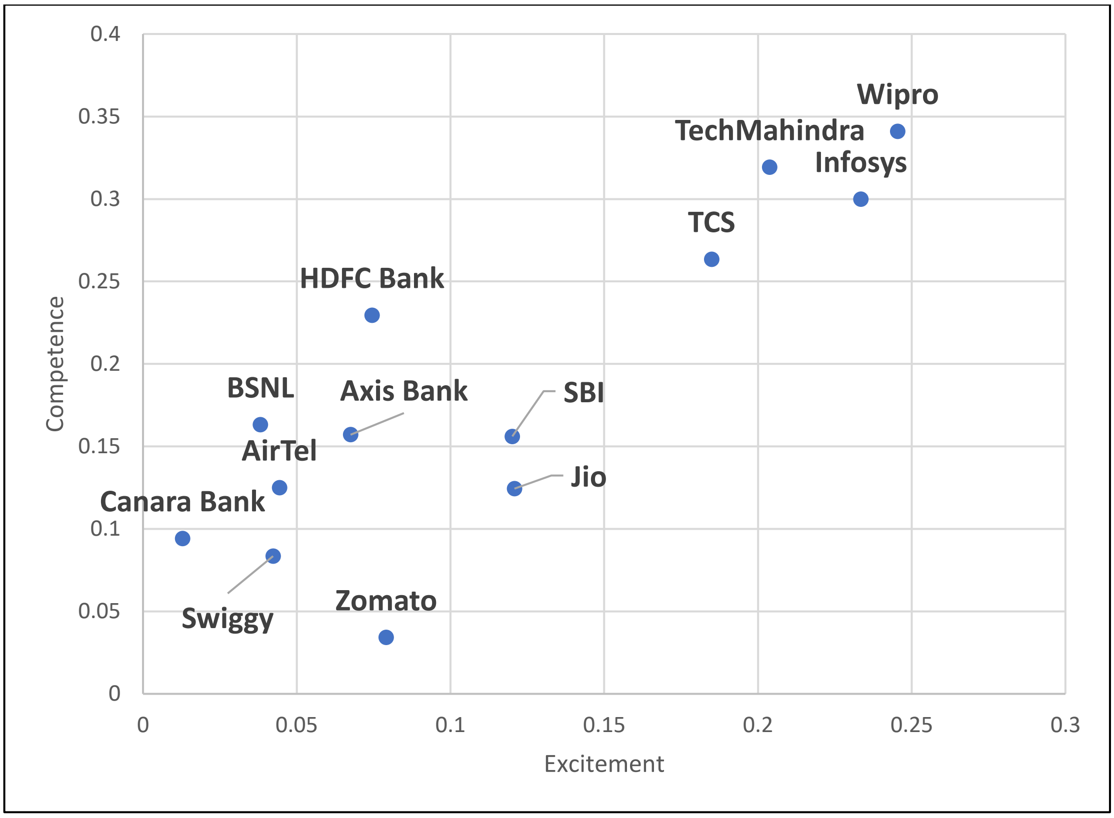

Figure 4.3. Communicated brand personalities: Excitement vs. Competence

The technology brands are posting hundreds of messages each week on Twitter. Despite their tacit acknowledgement that Twitter is a major branding channel, they are unable to differentiate themselves from each other in terms of communicated brand personality. This is also corroborated by Fig. 4.2 and Fig. 4.3 which show how the technology brands are placed along some of the dimensions. In these figures, the technology brands are not too far from each other. Being the technology experts, each of these brands should use phraseology that projects a distinctive personality to stand out among the crowded IT sector.

4.1.3. Perceived brand personality

The perceived brand personality is based on the tweets by the users. We computed the brand personality dimension scores before removing bots and suspicious accounts from the sample. The results are listed in Table 4.3. Table 4.4 contains the category-wise mean scores of the personality dimensions.

Table 4.3. Perceived brand personality scores (before removing suspicious accounts and bots)

Brand Name	Sophistication	Excitement	Sincerity	Competence	Ruggedness
AirTel	0.0052	0.0724	0.1102	**0.1413**	0.0264
BSNL	0.0038	**0.1188**	0.1115	0.1018	0.0286
Jio	0.0089	0.0769	**0.1367**	0.1194	0.0269
Swiggy	0.0103	0.0645	**0.1290**	0.0790	0.0236
Zomato	0.0166	0.0482	0.0942	**0.1085**	0.0201
Axis Bank	0.0162	0.0694	**0.1660**	0.1231	0.0425
Canara Bank	0.0450	0.0690	**0.1372**	0.0595	0.0202
HDFC Bank	0.0052	0.0731	**0.1518**	0.1231	0.0500
SBI	0.0058	0.0671	**0.1943**	0.1110	0.0426
Infosys	0.0123	0.1322	0.0903	**0.2097**	0.0227
TCS	0.0179	0.1000	0.0991	**0.1486**	0.0436
TechMahindra	0.0251	0.1463	0.1041	**0.1617**	0.0475
Wipro	0.0313	0.0654	0.0589	**0.1455**	0.0215

Note: Top-scoring dimension of each brand is in bold numerics.

Table 4.4. Category-wise mean scores of perceived brand personalities (before clean-up)

Category	Sophistication	Excitement	Sincerity	Competence	Ruggedness
Telecom	0.0059	0.0893	0.1194	0.1208	0.0273
Online food delivery	0.0134	0.0563	0.1116	0.0937	0.0218
Banking	0.0180	0.0696	0.1623	0.1041	0.0388
Technology	0.0216	0.1109	0.0881	0.1663	0.0338

Once the bots and suspicious accounts are removed from the sample, the personality dimension scores improved for some of the brands. They are listed in Table 4.5. The sincerity of technology brands decreased, and their sophistication increased after data clean-up. BSNL personality scores decreased in all dimensions after data clean-up. Overall, the predominant focus of the brands on sincerity and competence is still evident in Table 4.5.

Table 4.5. Perceived brand personality scores (after removing suspicious accounts and bots)

Brand Name	**Sophistication**	**Excitement**	**Sincerity**	**Competence**	**Ruggedness**
AirTel	0.0054	0.0699	0.1085	**0.1403**	0.0257
BSNL	0.0037	**0.1152**	0.1101	0.1007	0.0272
Jio	0.0094	0.0683	**0.1399**	0.1217	0.0262
Swiggy	0.0105	0.0644	**0.1290**	0.0786	0.0229
Zomato	0.0160	0.0486	0.0954	**0.1111**	0.0210
Axis Bank	0.0159	0.0692	**0.1691**	0.1226	0.0433
Canara Bank	0.0449	0.0701	**0.1342**	0.0580	0.0204
HDFC Bank	0.0048	0.0713	**0.1516**	0.1245	0.0509
SBI	0.0061	0.0681	**0.1949**	0.1111	0.0436
Infosys	0.0130	0.1343	0.0877	**0.2103**	0.0211
TCS	0.0180	0.1030	0.0976	**0.1460**	0.0451
TechMahindra	0.0259	0.1451	0.1033	**0.1658**	0.0486
Wipro	0.0319	0.0673	0.0570	**0.1390**	0.0222

Note: Top-scoring dimension of each brand is in bold numerics.

Table 4.6. Category-wise mean scores of perceived brand personalities (after clean-up)

Category	**Sophistication**	**Excitement**	**Sincerity**	**Competence**	**Ruggedness**
Telecom	0.0061	0.0844	0.1195	0.1209	0.0263
Online food delivery	0.0132	0.0565	0.1122	0.0948	0.0219
Banking	0.0179	0.0696	0.1624	0.1040	0.0395
Technology	0.0222	0.1124	0.0864	0.1652	0.0342

A comparison of the category-wise mean scores in Table 4.4 and Table 4.6 offers evidence of the bots and suspicious accounts on the brand personalities. The technology brands registered a slight improvement in their mean scores in sophistication, excitement and ruggedness after the data clean-up. The online food delivery brands are once again the poor performers in many dimensions. Banks are perceived to be sincere and rugged before and after data clean-up. The ruggedness in case of banks is related to the geographical coverage and the number of the bank branches.

While the telecom companies, banks and online food apps are directly interacting with consumers daily, the technology services companies are primarily involved in business-to-business marketing. So, the Twitter messages by the technology brands and about the technology brands look different compared to the other messages in this study.

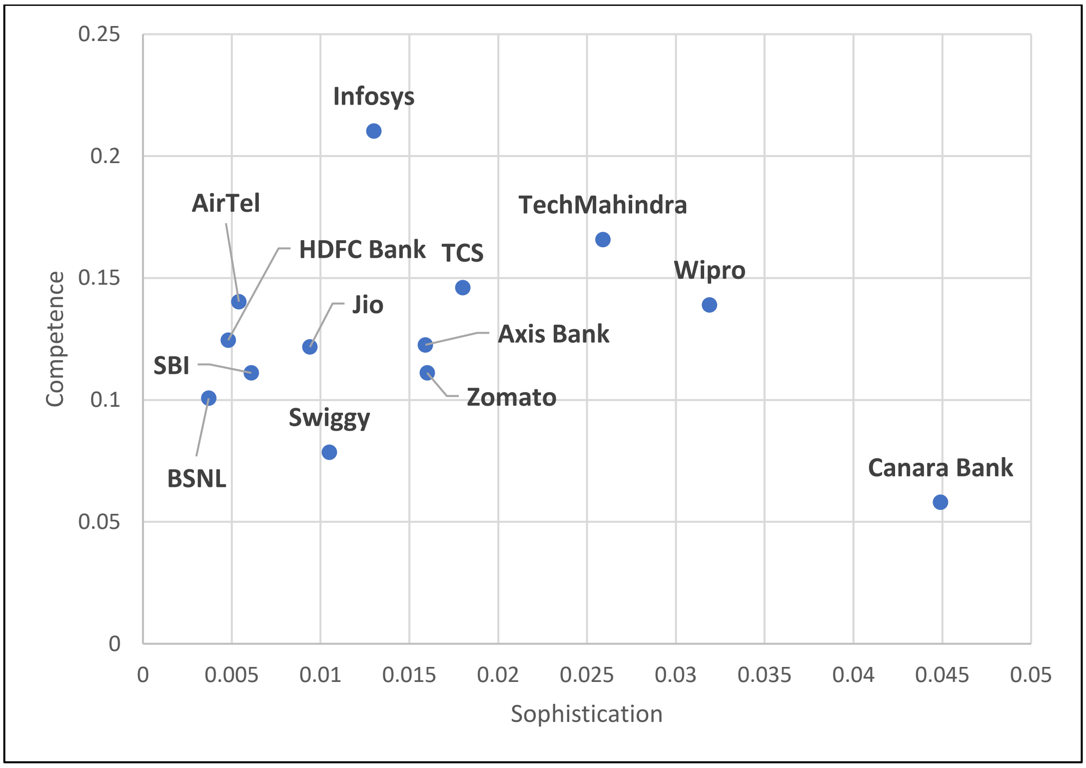

Figure 4.4. Perceived brand personalities: Sophistication vs. Competence

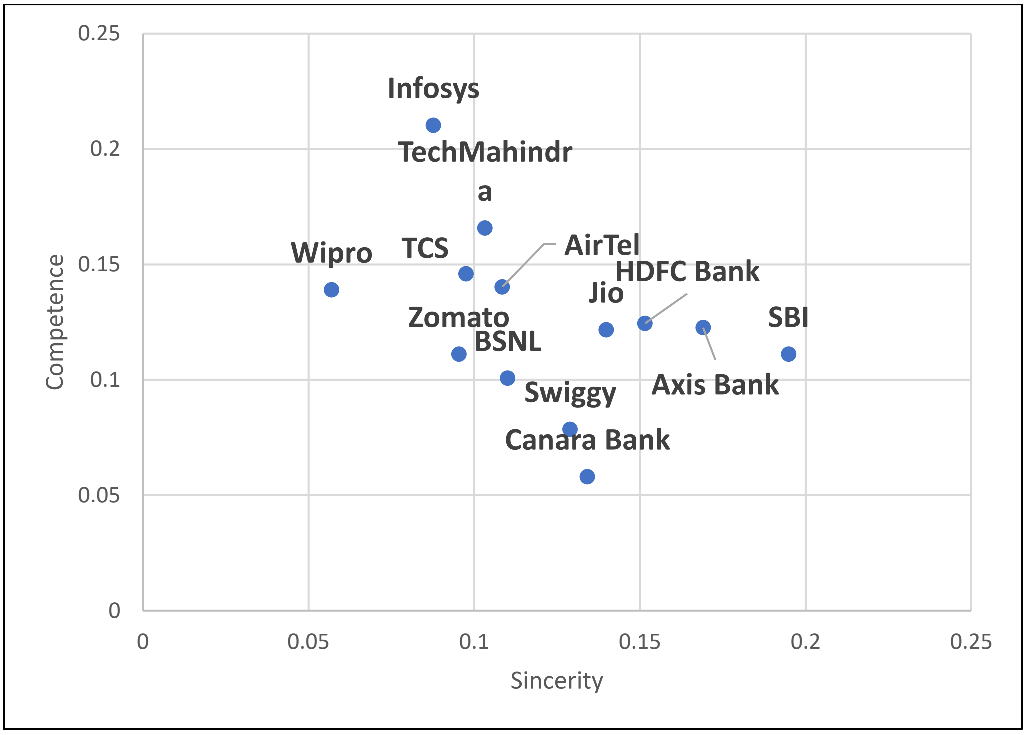

Figure 4.5. Perceived brand personalities: Sincerity vs. Competence

Looking at the perceived brand personalities along some dimensions in Fig. 4.4 and Fig. 4.5, Canara Bank has the highest sophistication among the 4 banks, whereas SBI is perceived to be the sincerest bank. Infosys is the technology brand with the highest perceived competence. BSNL’s perceived sophistication is the lowest among the 3 telecom companies.

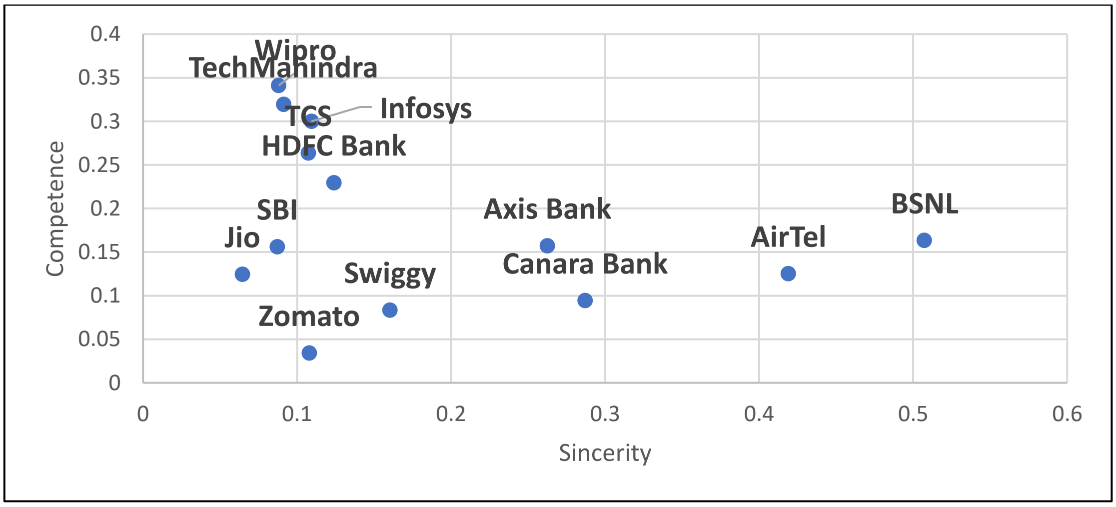

Figure 4.6. Communicated brand personalities: Sincerity vs. Competence

The relative placement of communicated brand personalities and perceived brand personalities can also be understood by using scatterplots such as those shown in Figure 4.6 and Figure 4.7. For 5 personality dimensions, a total of 10 scatterplots are possible. All scatterplots are not shown here.

Communicated brand personality is under the control of a brand; hence, a bigger range of brand personality values resulted in a widespread placement of brands along the sincerity and competence axes in Figure 4.6. However, the perceived brand personality plot in Figure 4.7 shows that the differentiation, in view of the users, is relatively smaller. It explains the crowded placement of brands on smaller ranges of values in Figure 4.7.

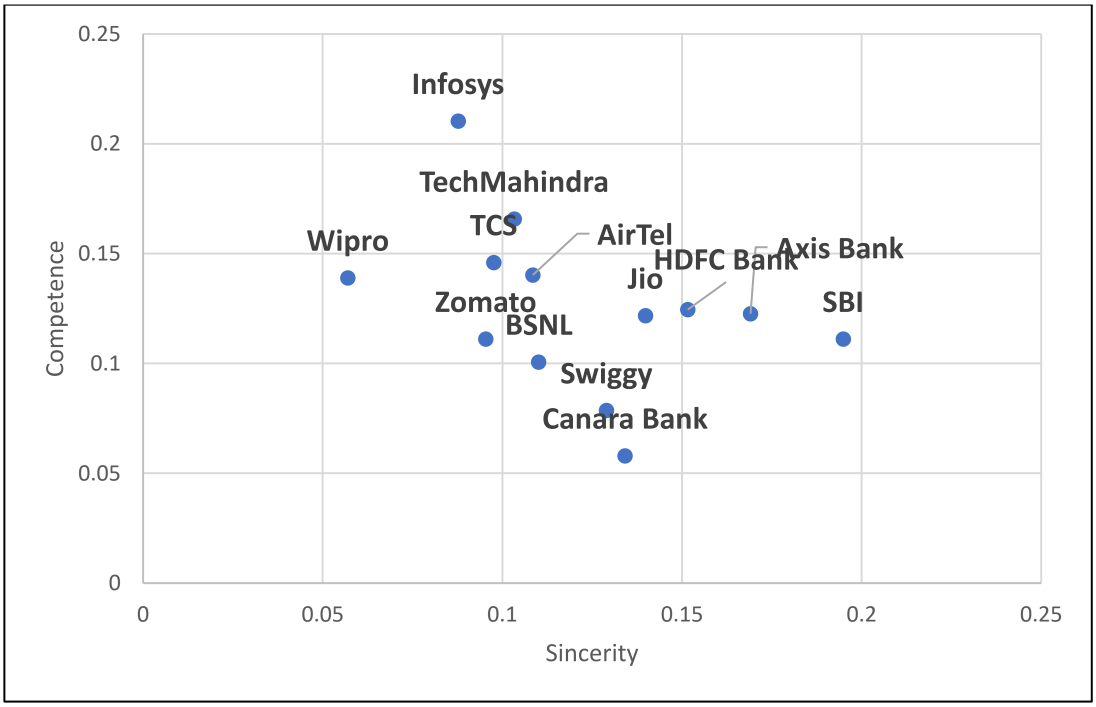

Figure 4.7. Perceived brand personalities: Sincerity vs. Competence

4.1.4. Perceived brand personality clusters

The dendrogram of the perceived brand personalities is illustrated in Figure 4.8.

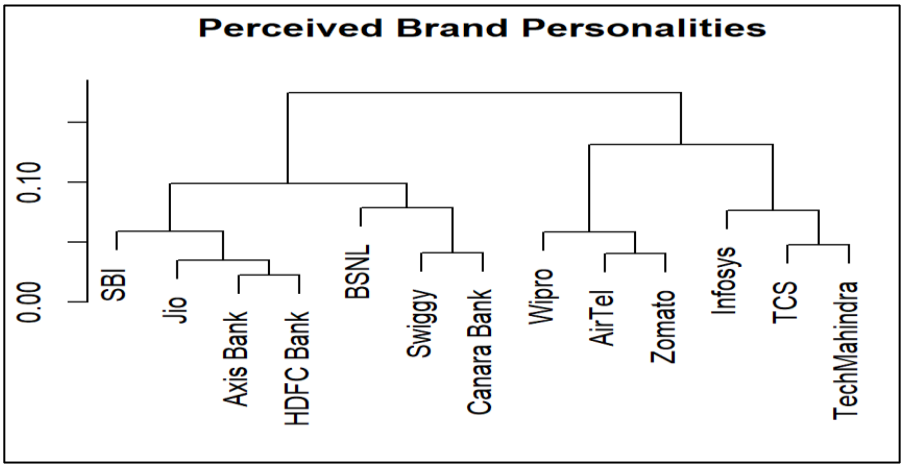

Figure 4.8. Perceived brand clusters after removing suspicious accounts and bots

The clusters based on the perceived brand personalities illustrate that brands across categories could have similar personalities. Hence, a cluster could have brands from unrelated categories. It is non-trivial to decipher insights from such groupings. For example, Jio is in the same cluster as Zomato before data clean-up; it is in the same cluster as Axis Bank, HDFC Bank and SBI after data clean-up. One could prophesy that Jio's parent company, Reliance Industries, might enter the banking sector or food delivery sector at an opportune moment.

4.1.5. Congruence between communicated and perceived brand personalities

A strong similarity between the communicated and perceived brand personalities could mean that the marketing communications efforts by the brand management team are probably successful. Among the many options for measuring similarity, cosine similarity is often used in the social media analytics (Berger et al., 2020; Culotta & Cutler, 2016; Netzer et al., 2012). If *x* and *y* represent the communicated and perceived brand personality vectors, the similarity between them is calculated as

$$cosine\ similarity = \frac{\sum_{i=1}^{n} x_i . y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \qquad (1)$$

A cosine similarity value of 0 means that there is no congruence between the communicated and perceived brand personalities; a value of 1 means that there is perfect congruence. Since SMM does not produce personality dimension vectors with negative values, the cosine similarity in our study will not be a negative value. Cosine similarity value of 1 does not necessarily mean that the brand personality is being managed well. A value of 1 is also possible in the undesirable scenario of a poorly communicated personality being in perfect alignment with a poorly perceived personality. A clearer understanding of cosine similarity value 1 can be obtained by analyzing not only the average scores but also the scores of the individual tweets.

Table 4.7. Congruence between communicated and perceived brand personalities

Brand Name	**Cosine Similarity** (before removing suspicious accounts and bots)	**Cosine Similarity** (after removing suspicious accounts and bots)
AirTel	0.7896 [2]	0.7892 [2]
BSNL	0.7530 [3]	0.7576 [3]
Jio	**0.8844** [1]	**0.8686** [1]
Swiggy	**0.9826** [1]	**0.9826** [1]
Zomato	0.8481 [2]	0.8450 [2]
Axis Bank	**0.9860** [1]	**0.9872** [1]
Canara Bank	0.8941 [3]	0.8881 [3]
HDFC Bank	0.9272 [2]	0.9294 [2]
SBI	0.8241 [4]	0.8257 [4]
Infosys	**0.9843** [1]	**0.9846** [1]
TCS	0.9805 [2]	0.9810 [2]
TechMahindra	0.9672 [3]	0.9708 [3]
Wipro	0.9615 [4]	0.9654 [4]

Note: Within-category congruence rank is indicated in square brackets.

Table 4.7 lists the congruence rank of the brands within each category. A counterintuitive finding is that the congruence rank did not change for any brand after removing bots and suspicious accounts from the data. Technology brands showed better alignment between their communicated and perceived brand personalities; their alignment improved after data clean-up. Technical competence being one of their core strengths, some of them are able to skilfully wordsmith their brand positions on Twitter. Infosys is younger than TCS and Wipro; however, it has a better congruence rank than TCS and Wipro.

Among the banks, Axis Bank is not the top-scorer in most of the brand personality dimensions. But it has the best congruence rank. Axis Bank may want to pay particular attention to its current standing and work on improving its brand personality dimension scores.

Swiggy's congruence rank and some of its brand personality scores are higher than those of Zomato. Though a newer and smaller player than Zomato, Swiggy seems to have a better understanding of how to position its brand on Twitter. Online food delivery companies face frequent comments and criticism from the consumers. Zomato cannot afford to ignore the misalignment of its brand personality on social media.

Telecom brands are among the lowest-ranked brands in the congruence ranking. It shows that the consumers' perception of the brands is quite different from the communicated brand personalities.

The public sector telecom company BSNL has the lowest congruence rank. Similarly, the public sector banks Canara Bank and SBI have worse congruence ranks than the private sector banks. Overall, the public sector brands in India included in our study may need to review their text communications on Twitter in a systematic and comprehensive manner to achieve better alignment of communicated and perceived brand personalities.

4.2. Brand position and brand sentiment

The brand position and sentiment values generated by VADER method are listed in Table 4.8. To allow easier interpretation, the negative position and sentiment values are shown in redcolored font; neutral position and sentiment values are shown in orange-colored font and positive position and sentiment values are shown in green-colored font.

As expected, most of the brands have a positive brand position value. Tweets posted by the brands are under their full control; the brand managers are able to develop messaging that indicates a positive position of their respective brands. Surprisingly, BSNL has a neutral position with a score of -0.0024. There is no rational explanation for BSNL's subdued communication on Twitter. BSNL needs to investigate the tweets and calibrate them to move from a neutral position to a decidedly positive position.

Table 4.8. Brand position and brand sentiment (normalized to a scale of -1 to +1)

Brand Name	**Communicated Brand Position**	**Brand Sentiment** (before removing suspicious accounts and bots)	**Brand Sentiment** (after removing suspicious accounts and bots)
AirTel	**0.5726**	**-0.0623**	**-0.0629**
BSNL	**-0.0024**	**0.1032**	**0.0984**
Jio	**0.3685**	**-0.0424**	**-0.0463**
Swiggy	**0.3097**	**-0.0254**	**-0.0232**
Zomato	**0.3851**	**0.0599**	**0.0559**
Axis Bank	**0.3095**	**0.0122**	**0.0120**
Canara Bank	**0.4379**	**0.1562**	**0.1480**
HDFC Bank	**0.4081**	**0.0412**	**0.0391**
SBI	**0.1991**	**0.1116**	**0.1024**
Infosys	**0.3639**	**0.1970**	**0.1892**
TCS	**0.4244**	**0.2340**	**0.2349**
TechMahindra	**0.4400**	**0.2243**	**0.2256**
Wipro	**0.4583**	**0.2016**	**0.2152**

Interpretation of VADER scores:

- **positive sentiment: score >= 0.05**
- **neutral sentiment: score > -0.05 and score < 0.05**
- **negative sentiment: score <= -0.05**

AirTel is the only brand which has a negative brand sentiment. The ability to freely express negative sentiment is one of the drivers of social media's popularity. A framework to deal with negative sentiment based on five domains, namely, company philosophy, company operations, customer relationships, internal relationships and other stake holder relationships, might help AirTel in reducing the negativity in brand sentiment (Cooper et al., 2019). The brand sentiment value of 3 telecom companies and 4 banks worsened after removing bots and suspicious accounts from the data. The sentiment value of most technology brands improved after the data clean-up.

In all cases where the sentiment value improved or worsened, the sentiment category – negative, neutral or positive – did not change for any of the 13 brands after the data clean-up. This counterintuitive finding is somewhat similar to what was noticed with the congruence ranks of brand personalities before and after the data clean-up.

4.3. Association between communicated brand personality and brand position

Firms post a lot of text messages on social media to communicate the brand personality and brand position. We studied the association between the communicated brand personality and position using Ordinary Least Squares (OLS) regression. In this regression for each brand, we used sophistication, excitement, sincerity, competence and ruggedness as the independent variables, and brand position as the dependent variable. Appendix A, Table A3 summarizes the regression results.

The regression model of every brand is significant in terms of the p-value associated with the F-test. For the sake of brevity, we do not delve into the details of individual regression models. We highlight the overall patterns observed among the 13 regression models. The adjusted R^2 value ranges from Zomato's 0.0095 to Canara Bank's 0.3233. For the 3 telecom brands, the association between sincerity and brand position, and the association between competence and brand position is significant.

For banks and technology brands, the excitement, sincerity and competence have a significant effect on brand position. Interestingly, ruggedness has a significant association with brand position only in case of online food delivery brands.

The multicollinearity was checked using Variance Inflation Factor (VIF). Because none of the VIF values exceed 2 in Table 4.9, the multicollinearity among the communicated personality dimensions does not appear to be a matter of concern. We used multiple linear regression to check for association between the dependent variable and the independent variables. It was not used for inference, prediction or causality. So, we did not perform other checks such as normality of the residuals, homoscedasticity of the residuals, etc. .

Table 4.9. Multicollinearity among communicated brand personality dimensions (VIF values)

Brand	Sophistication	Excitement	Sincerity	Competence	Ruggedness
Airtel	1.002129	1.144252	1.012236	1.01136	1.146665
BSNL	1.006038	1.014935	1.031923	1.037659	1.033635
Jio	1.006242	1.006231	1.002771	1.002839	1.003897
Swiggy	1.000702	1.001832	1.035874	1.031096	1.009143
Zomato	1.001188	1.002146	1.001222	1.000377	1.001505
AxisBank	1.001857	1.01152	1.00661	1.010006	1.011877
CanaraBank	1.249933	1.250797	1.037673	1.010771	1.031374
HDFCBank	1.007057	1.01255	1.000869	1.001007	1.005837
SBI	1.003201	1.013908	1.003336	1.004862	1.01514
Infosys	1.002383	1.00358	1.006415	1.008225	1.000883
TCS	1.00194	1.01319	1.001114	1.020073	1.013316
TechMahindra	1.004557	1.008868	1.004776	1.006885	1.002888
Wipro	1.001223	1.005123	1.002081	1.004532	1.001909

Looking at all 13 brands in Appendix A, Table A3, there is a weak positive association between the brand position and the communicated brand personality dimensions.

4.4. Association between perceived brand personality and brand sentiment

The association between perceived brand personality and brand sentiment has not been empirically studied so far. For each brand, we ran OLS regression with the brand personality dimensions as the independent variables and the brand sentiment as the dependent variable. The regression results are listed in Appendix A, Table A4.

Not all 13 regression models are significant in terms of the p-value associated with the F-test. Among the brands which show a significant association, the adjusted R^2 value ranges from Axis Bank's 0.0022 to TechMahindra's 0.0581. There is no hard cut-off value for adjusted R^2 in the marketing domain. The sincerity of AirTel and Jio has a significant association with brand sentiment.

The public sector banks, Canara Bank and SBI, have a strong association between sincerity and brand sentiment. The private sector banks, Axis Bank and Canara Bank, do not demonstrate a common pattern of association between the personality dimensions and sentiment. Axis Bank demonstrated an association between excitement and sentiment, whereas HDFC Bank has an association between sophistication and sentiment. The online food delivery brands did not reveal any connection between the perceived brand personality dimensions and brand sentiment. All technology brands displayed an association between sophistication, competence and brand sentiment.

None of the VIF values exceed 2 in Table 4.10, so the multicollinearity among the perceived personality dimensions was not a concern.

To summarize, the association between the brand sentiment and the perceived brand personality dimensions is not as strong and as clear as the association between the brand position and communicated brand personality dimensions.

Table 4.10. Multicollinearity among perceived brand personality dimensions (VIF values)

Brand	Sophistication	Excitement	Sincerity	Competence	Ruggedness
AirTel	1.001028	1.002301	1.004427	1.007736	1.004302
BSNL	1.000432	1.003429	1.003468	1.004747	1.002764
Jio	1.001855	1.001883	1.002005	1.002085	1.001886
Swiggy	1.000634	1.003571	1.001732	1.002565	1.00073
Zomato	1.002266	1.013564	1.016372	1.004061	1.00063
AxisBank	1.004573	1.007673	1.018435	1.019951	1.00607
CanaraBank	1.023335	1.046664	1.016243	1.022109	1.00656
HDFCBank	1.001027	1.007293	1.00877	1.002136	1.000455
SBI	1.004348	1.016225	1.009398	1.014666	1.017111
Infosys	1.001222	1.02415	1.009528	1.024524	1.009494
TCS	1.003553	1.017655	1.004324	1.019404	1.00167
TechMahindra	1.015771	1.012404	1.004574	1.020675	1.016558
Wipro	1.002446	1.013294	1.00705	1.010104	1.013167

4.5. Images and the brand identity – brand image alignment

A novel way of understanding the brand identity and brand image using Instagram images is presented here. This approach could be used very frequently and it could be interspersed with periodic primary data-based studies.

4.5.1. Qualitative analysis of image labels

The most frequent, most important and most co-occurring labels for each brand based on FGI are listed in Table 4.11. The labels based on UGI are listed in Table 4.12.

Table 4.11. Brand identity: Labels from firm-generated images

Brand Name	Labels
AirTel	*Most frequent*: device, font, gadget, brand, automotive, rectangle, gesture, logo, communication, design
	Most important: device, automotive, mobile, brand, logo, rectangle, circle, design, graphics, event
	Most co-occurring: design, automotive, event, vehicle, smile, fashion, dress, rolling, mirror, sleeve
BSNL	*Most frequent*: font, event, graphics, poster, advertising, blue, electric, brand, organism, logo
	Most important: blue, electric, parallel, graphics, advertising, event, poster, plant, logo, brand
	Most co-occurring: font, event, poster, graphics, advertising, blue, electric, brand, organism, logo
Jio	*Most frequent*: font, blue, electric, event, circle, gesture, graphics, happy, sleeve, collar
	Most important: blue, electric, event, circle, happy, graphics, rectangle, music, sleeve, logo

	Most co-occurring: flash, photography, gesture, camera, forehead, happy, eyewear, smile, hair, glasses
Swiggy	*Most frequent*: font, food, happy, brand, gesture, ingredient, logo, rectangle, tableware, wood *Most important*: food, brand, logo, font, rectangle, happy, graphics, circle, gesture, paper *Most co-occurring:* food, tableware, ingredient, cuisine, dishware, recipe, dish, wood, utensil, kitchen
Zomato	*Most frequent*: font, brand, magenta, food, graphics, event, logo, cuisine, recipe, tableware *Most important*: graphics, food, magenta, event, brand, logo, blue, electric, device, design *Most co-occurring:* food, recipe, cuisine, tableware, ingredient, dish, staple, dishware, plate, serveware
Axis Bank	*Most frequent*: font, magenta, graphics, logo, brand, happy, design, smile, advertising, circle *Most important*: pattern, art, slope, logo, graphics, happy, brand, circle, smile, design *Most co-occurring:* font, magenta, slope, diagram, rectangle, pattern, advertising, art, illustration, parallel
Canara Bank	*Most frequent*: font, advertising, line, event, screenshot, brand, poster, parallel, logo, product *Most important*: circle, brand, parallel, device, automotive, screenshot, event, product, logo, poster *Most co-occurring:* parallel, line, screenshot, font, blue, electric, brand, rectangle, logo, product
HDFC Bank	*Most frequent*: font, blue, electric, happy, gesture, logo, rectangle, care, event, graphics *Most important*: logo, brand, rectangle, graphics, happy, blue, gesture, automotive, electric, parallel *Most co-occurring:* automotive, happy, font, blue, eyewear, sunglasses, goggles, care, vision, leisure
SBI	*Most frequent*: font, blue, advertising, electric, event, brand, happy, poster, graphics, rectangle *Most important*: advertising, rectangle, brand, event, electric, blue, happy, graphics, poster, circle *Most co-occurring:* happy, smile, font, gesture, event, sleeve, blue, advertising, poster, electric
Infosys	*Most frequent*: font, equipment, blue, electric, gesture, computer, device, sports, display, personal *Most important*: computer, tennis, equipment, device, blue, electric, sports, display, gesture, nature *Most co-occurring:* font, equipment, blue, electric, gesture, computer, device, sports, personal, display
TCS	*Most frequent*: font, event, blue, electric, recreation, design, sleeve, happy, leisure, sky *Most important*: font, document, event, blue, electric, recreation, brand, leisure, tshirt, sky *Most co-occurring:* recreation, event, tree, shorts, crowd, building, sports, fun, city, footwear
TechMahindra	*Most frequent*: font, event, blue, happy, electric, brand, advertising, suit, blazer, collar *Most important*: blue, happy, suit, electric, event, brand, collar, advertising, sleeve, blazer *Most co-occurring:* font, event, happy, blue, electric, brand, advertising, suit, collar, blazer
Wipro	*Most frequent*: blue, font, electric, event, art, circle, photography, sleeve, device, brand *Most important*: brand, circle, blue, font, art, device, electric, plant, event, sleeve *Most co-occurring:* font, blue, electric, event, art, photography, circle, sleeve, device, fashion

The deep learning model expectedly generates labels such as ‘font’ and ‘logo’ for many images. These labels are not helpful in differentiating one brand from the other. Ignoring those common words, we analyzed the other frequent and relevant words to understand the traits that are most emphasized by the brands. The co-occurring words are included since they might reveal the common user-context combinations.

Table 4.12. Brand image: Labels from user-generated images

Brand Name	**Labels**
AirTel	*Most frequent*: font, brand, event, logo, graphics, advertising, rectangle, circle, blue, electric
	Most important: rectangle, circle, advertising, graphics, blue, electric, logo, event, brand, poster
	Most co-occurring: event, device, font, sleeve, smile, happy, table, window, photography, tie
BSNL	*Most frequent*: font, poster, advertising, event, happy, logo, graphics, brand, blue, electric
	Most important: blue, electric, graphics, logo, brand, advertising, event, happy, rectangle, poster
	Most co-occurring: happy, gesture, font, sports, nature, greeting, expression, jersey, organism, Christmas
Jio	*Most frequent*: font, blue, electric, brand, device, logo, event, advertising, graphics, circle
	Most important: device, circle, brand, logo, event, advertising, blue, electric, rectangle, graphics
	Most co-occurring: device, computer, gadget, mobile, communication, output, font, laptop, personal, portable
Swiggy	*Most frequent*: food, font, ingredient, recipe, cuisine, dish, tableware, poster, advertising, natural
	Most important: food, poster, tableware, advertising, dish, brand, cuisine, natural, logo, foods
	Most co-occurring: font, logo, brand, vehicle, graphics, advertising, event, design, poster, tire
Zomato	*Most frequent*: food, ingredient, recipe, cuisine, font, dish, tableware, poster, produce, natural
	Most important: food, tableware, poster, dish, font, produce, advertising, cuisine, cake, natural
	Most co-occurring: cake, food, decorating, table, cuisine, ingredient, supply, dessert, chair, frying
Axis Bank	*Most frequent*: font, event, smile, advertising, happy, brand, design, fashion, magenta, computer
	Most important: advertising, computer, font, event, brand, happy, smile, magenta, fashion, product
	Most co-occurring: font, advertising, brand, magenta, electric, blue, rectangle, poster, parallel, event
Canara Bank	*Most frequent*: font, advertising, brand, blue, electric, event, poster, product, screenshot, parallel
	Most important: blue, advertising, electric, event, screenshot, parallel, brand, rectangle, poster, logo
	Most co-occurring: blue, electric, font, parallel, rectangle, brand, event, logo, advertising, graphics
HDFC Bank	*Most frequent*: font, blue, electric, brand, parallel, rectangle, event, circle, logo, screenshot
	Most important: parallel, circle, rectangle, brand, number, screenshot, blue, electric, event, logo
	Most co-occurring: automotive, vehicle, tire, lighting, car, wheel, motor, bumper, hood, equipment
SBI	*Most frequent*: font, event, blue, advertising, electric, brand, logo, vehicle, rectangle, product
	Most important: vehicle, blue, advertising, brand, event, electric, logo, rectangle, product, plant
	Most co-occurring: smile, shirt, tshirt, event, job, facial, expression, medical, gesture, sleeve
Infosys	*Most frequent*: font, event, plant, blue, smile, electric, sky, brand, building, rectangle
	Most important: blue, event, plant, electric, brand, font, parallel, smile, rectangle, building

	Most co-occurring: botany, leaf, dress, day, flowerpot, houseplant, leisure, waist, smile, plant
TCS	*Most frequent*: font, plant, sleeve, smile, advertising, design, eyewear, tree, gesture, happy *Most important*: plant, font, advertising, tree, sleeve, smile, event, blue, happy, recreation *Most co-occurring:* crowd, snapshot, people, world, black, art, line, happy, gesture, recipe
TechMahindra	*Most frequent*: font, event, advertising, brand, rectangle, building, tree, design, plant, smile *Most important*: brand, rectangle, event, advertising, parallel, building, screenshot, tree, plant, smile *Most co-occurring:* font, event, advertising, brand, rectangle, building, tree, plant, design, smile
Wipro	*Most frequent*: font, circle, brand, event, advertising, rectangle, parallel, blue, electric, plant *Most important*: rectangle, number, parallel, circle, plant, event, symmetry, terrestrial, blue, advertising *Most co-occurring:* font, advertising, product, collar, sleeve, brand, smile, job, event, computer

Telecom brands: Jio's superior brand identity

In case of the telecom brands such as Airtel and BSNL, the most frequent and most relevant words do not facilitate any major distinction between FGI images and UGI images. They both focus on phones, announcements and brand-sponsored events such as marathons. However, the most co-occurring labels of Jio reveal that it is more exuberant than Airtel and BSNL. Although it is true that Jio undercut its competitors with low-priced service plans, Jio also focused beyond phones and vigorously advertised broadband services for laptops.

Online food delivery brands: Swiggy's superior brand image

The FGI images of both Swiggy and Zomato contain cuisines, silverware and happiness resulting from consuming food. The brands do not seem to differentiate themselves in a substantial manner. The UGI images indicate a marked difference between the two brands. Zomato's UGI images show a predominance of celebrations with cakes and desserts that were probably ordered from Zomato.

Swiggy is a relatively newer entrant in this arena. Its users found ways of connecting with the emotions generated by Swiggy's branding efforts. The UGI images show the users expressing how the Swiggy-provided food mattered in their daily lives. More restaurants mentioned Swiggy than Zomato, thereby helping it benefit from the publicity.

Banks: HDFC Bank's misalignment

While Axis Bank uses crimson-coloured FGI images, Canara Bank, HDFC Bank and SBI use blue-coloured images. This is consistent with the colours of their respective brand logos. The contents of UGI images reflect a lot more variety compared to the FGI images, because the users are not bound by the brands' official colours and logos.

HDFC Bank's UGI images reflect many images containing automobiles. At the time of data collection, HDFC Bank was offering various loans for automobile purchase but few customers posted pictures with the automobiles. Other banks' UGI images show branch facilities and celebrations which often involved customers.

Technology brands: Brand image could also depend on non-technological aspects !

Infosys predictably focused on the computational infrastructure and office facilities in its FGI images. Messaging related to its software and IT services is not very prominent in its images. The UGI images, on the other hand, show users posting cheerful and outdoorsy images around Infosys premises. Some images also contain derisive remarks about the company's hiring, compensation, etc. . They offer evidence that a brand's image cannot be driven by employees or non-employees posting exclusively positive images.

TCS built its FGI images around its recreational events, presumably in an effort to promote itself as one of the places with better work-life balance. UGI images are also centered around the similar themes, thereby indicating a fair alignment between brand identity and brand image.

For TechMahindra, the FGI images projected a happy workplace. UGI images instead focused on the camaraderie among its employees; they do not say whether the company's compensation, benefits and facilities resulted in happy employees.

Wipro is posting relatively blander FGI images compared to its competitors. Wipro needs to use the visual channel by posting vibrant and exciting imagery. Wipro's UGI images are similar to TCS's images. In other words, the employees and the customers opinions about these two companies are not too different. There does not appear to be a good alignment between the brand identity and brand image of Wipro.

4.5.2. Quantitative analysis of image labels

In addition to the qualitative analysis of brands using image labels, the relative positions of brands can be understood using brand clusters. It requires using the numerical representations associated with the labels.

4.5.3. Brand clusters

The distance matrix of FGI images is in Table 4.13. The distance between a given pair of brands indicates the dissimilarity between them. The smaller the distance, the greater is the similarity between the brands on Instagram.

Table 4.13. Brand identity: Distance matrices based on firm-generated images

	AirTel	BSNL	Jio
AirTel		1.3647	1.4142
BSNL	1.3647		1.4142
Jio	1.4142	1.4142	

Table 4(a). Telecom brands

	Swiggy	Zomato
Swiggy		1.4142
Zomato	1.4142	

Table 4(b). Online food delivery brands

	AxisBank	Canara Bank	HDFC Bank	SBI
AxisBank		1.4142	1.3713	1.3487
CanaraBank	1.4142		1.4142	1.4142
HDFCBank	1.3713	1.4142		1.3456
SBI	1.3487	1.4142	1.3456	

Table 4(c). Banks

	Infosys	TCS	TechMahindra	Wipro
Infosys		1.4142	1.4142	1.3685
TCS	1.4142		1.4038	1.4142
TechMahindra	1.4142	1.4038		1.4142
Wipro	1.3685	1.4142	1.4142	

Table 4(d). Technology brands

The corresponding dendrogram is in Figure 4.9. A dendrogram shows the manner in which the brands are combined in a bottom-to-top direction. The y-axis shows the Hellinger's distance at which the two brands are linked. The lower the height of the link, the higher is the similarity between the brands.

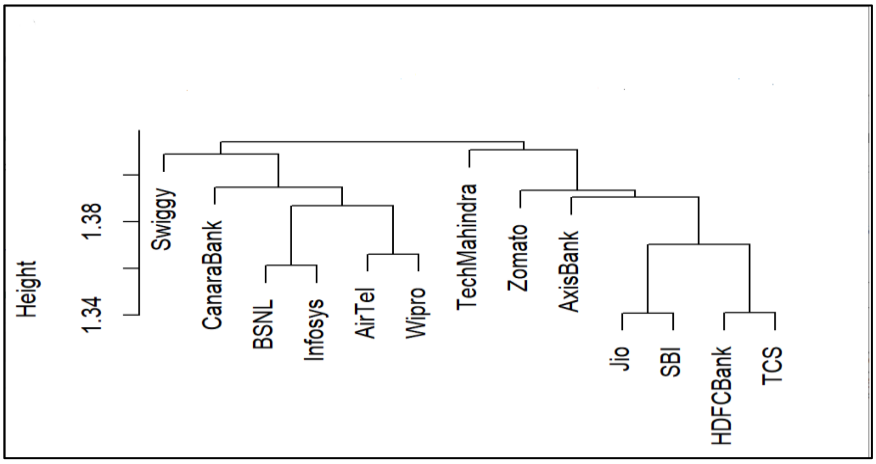

Figure 4.9. Brand identity: Clusters based on firm-generated images

The distance matrix of UGI images is in Table 4.14 and the corresponding dendrogram is in Figure 4.10.

Table 4.14. Brand image: Distance matrices based on user-generated images

	AirTel	BSNL	Jio
AirTel		1.3668	1.4142
BSNL	1.3668		1.4142
Jio	1.4142	1.4142	

Table 5(a). Telecom brands

	Swiggy	Zomato
Swiggy		1.3521
Zomato	1.3521	

Table 5(b). Online food delivery brands

	Axis Bank	Canara Bank	HDFC Bank	SBI
AxisBank		1.4142	1.4142	1.3899
CanaraBank	1.4142		1.2638	1.4142
HDFCBank	1.4142	1.2638		1.4142
SBI	1.3899	1.4142	1.4142	

Table 5(c). Banks

	Infosys	TCS	TechMahindra	Wipro
Infosys		1.4108	1.4142	1.4031
TCS	1.4108		1.4142	1.3799
TechMahindra	1.4142	1.4142		1.4142
Wipro	1.4031	1.3799	1.4142	

Table 5(d). Technology brands

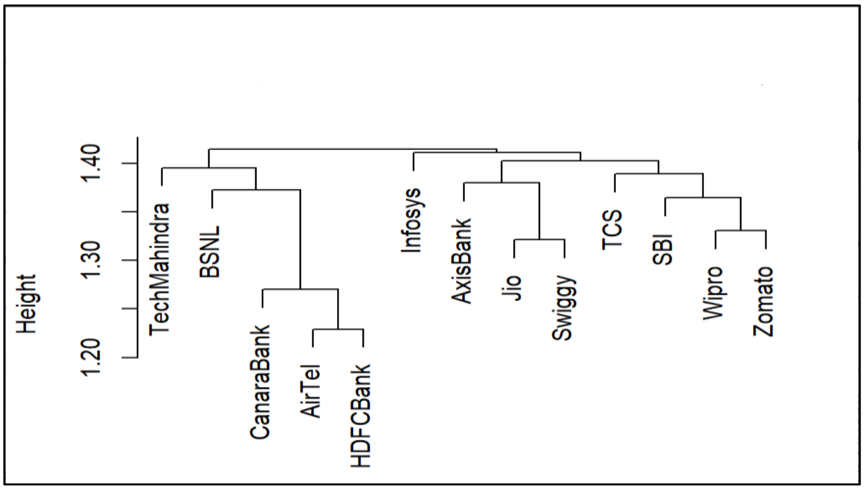

Figure 4.10. Brand image: Clusters based on user-generated images

The range of distances in Table 4.14 (1.2638 – 1.4142) is larger than the range in Table 4.13 (1.3456 – 1.4142). A closer examination of the distances reveals that the brands in the telecom, online food delivery and technology sectors have closely-situated brand image within their respective categories. The banking sector has shown an ability to have a variety of brand images.

AirTel is more similar to BSNL than to Jio both in terms of brand identity and brand image. The curious case of an entrenched private player such as AirTel being similar to a public sector player such as BSNL raises questions about AirTel's brand management on social media.

Swiggy and Zomato are apart by a distance of 1.4142, indicating fairly different brand identities. This gap is wider than the distance of 1.3521 between their respective brand images. It means the customers post relatively similar opinions, experiences or complaints about both the brands in the form of images. The online food delivery brands need to strive towards achieving a distinct impression in the minds of consumers.

It is evident that the banking brands have a broader variety of brand image compared to the other sectors. The services offered by the banks have engendered a wider variety of opinions from the consumers. The brand identities of private sector banks AxisBank and HDFCBank is closer to each other than those of public sector banks Canara Bank and SBI. With a distance of 1.3899, the brand image of AxisBank is closer to SBI than to the other two banks. HDFCBank and Canara Bank have a more similar image at a distance of 1.2638.

At a distance of 1.3685, the brand identities of Infosys and Wipro are closer than those of their competitors in technology sector. However, the brand images of TCS and Wipro are closer to each other than the rest of the brands. These differences resulted in different degrees of brand identity-image alignment.

4.5.4. Alignment between brand identity and brand image

Correlation between FGI and UGI

There does not appear to be a statistically significant correlation between the FGI distance matrix and UGI distance matrix per the Mantel's test (Z = 151.9434; p = 0.345 with 999 permutations). It means that the FGI images and the UGI images of these brands diverge considerably from each other, despite individual similarities. This test result justifies the need to investigate the degree of misalignment between FGI images and UGI images.

Alignment between FGI and UGI

Based on the analysis done using image labels and principal components, an assessment of the degree of alignment between the brand identity and brand image is presented in Table 4.15. The brands are ranked within each category.

It should be emphasized that a good alignment between the brand identity and brand image does not mean that there is no need to improve either of them. At times, the customers opinions are in agreement with a poorly projected brand identity. BSNL appears to be such an example with its top-rank in the alignment. BSNL's communications are quite different than those of the private firms AirTel and Jio. AirTel is closer to BSNL than it is to Jio. Jio is able to differentiate itself better than its competitors in brand identity. This conclusion is in line with its label-based positioning. However, Jio is ranked third because the firm's communications are emphasizing the infrastructure and offers, whereas the users' communications are mostly about the network connectivity and customer service issues.

Table 4.15. Brand identity – brand image alignment

Brand category	Brand name	Euclidean distance	Within-category rank
Telecom	BSNL	0.0330	1
	AirTel	0.1901	2
	Jio	0.2069	3
Online food delivery	Swiggy	0.0509	1
	Zomato	0.4279	2
Banking	CanaraBank	0.0431	1
	SBI	0.1068	2
	AxisBank	0.1637	3
	HDFCBank	0.2181	4
Technology	TCS	0.0781	1
	TechMahindra	0.1422	2
	Infosys	0.1796	3
	Wipro	0.3379	4

Among the banks, the public sector banks Canara Bank and SBI have better alignment compared to the private sector banks Axis Bank and HDFC Bank. In the online food delivery sector, the gap between Swiggy and Zomato's brand images was evident from the image labels and it largely contributed to the dissimilarity between them.

TCS maintains a good alignment between its brand identity and image in the technology brands. This is in line with the insights deduced from the image labels. Wipro appears to find it difficult to achieve a reasonable alignment; its bland image communications would need to be revisited to improve the perceptions of the customers.

Chapter 5
DISCUSSION

This chapter details a few specific suggestions for the brands. It also delves into the methodological implications and managerial implications of our contributions.

5.1. Suggestions for the brands

The suggestions for the 13 leading Indian brands are listed in Table 5.1. Brands which are less than 20 years old are considered as relatively new brands. Although it might seem to be an arbitrary categorization, we chose to use it as one of the ways of delineating the brands in our study.

Table 5.1. Suggestions for brands

	Twitter (X)	**Instagram**
All brands	• Avoid bland tweets • Use Aaker's dictionary and Opoku's dictionary to improve the verbiage • Implement the Cooper et al. (2019) recommendations on reducing negative sentiment	• Reduce wordy announcements in the pictures (Kane & Pear, 2016; Hernández-Méndez & Muñoz-Leiva, 2015) • Improve the brand image by increasing the emotive content
Relatively new brands [< 20 years]	• Increase the emphasis on sophistication and excitement	

5.1.1. Suggestions for improving the tweets

A tweet which is assigned near-zero scores in sincerity, competence, excitement, ruggedness and sophistication dimensions is termed a 'bland tweet'. To begin with, a brand manager should not entertain the notion that some tweet is better than no tweet. Brand managers should examine the words used in the bland tweets. They can improve the scores by crafting tweets based on some of the words from Aaker's dictionary in Table 2.4 and Opoku's dictionary in Appendix A, Table A2.

Negative brand sentiment can emanate from brand hate, anti-brand communities, complaints, customer rage and consumer misbehavior. Reducing negative sentiment is not limited to tweaking the content of tweets on a case-by-case basis. Cooper et al. (2019) suggested the following set of guidelines for reducing the negative sentiment on social media:

1. Understanding the brand's core values.
2. Commitment to uphold those values.
3. Understanding the dynamics of social media communications in order to maintain a positive relationship with the customers.

In case of mature brands, it is not surprising to see strong emphasis on sincerity and competence. In their elaborate meta-analytic review of the brand personality studies, Eisend and Stokburger-Sauer (2013) found that sincerity and competence are the two personality dimensions that had the most impact on a mature brand's success in the marketplace. However, the younger brands should focus on other dimensions such as excitement, ruggedness and sophistication too. Brands such as Swiggy and Zomato would find it beneficial to make their applications look more exciting and sophisticated. As they expand delivery to more geographical locations, their ruggedness score would increase and strengthen the overall brand personality.

5.1.2. Suggestions for improving the Instagram images

Indian brands have shown a preference for using the Instagram to make lengthy announcements in the form of pictures. An occasional wordy message to convey important information does little harm; but, a frequent pattern of wordiness in images appears to ignore the old adage “A picture is worth a thousand words”. Increasing the visual content and decreasing the wordiness were also recommended by earlier studies such as Kane and Pear (2016), and Hernández-Méndez and Muñoz-Leiva (2015). Users take much longer to notice the text than the images in an Instagram post. As the images with lesser number of words start appearing on Instagram, they might increase the likelihood of emotive reactions from the users.

Compared to the global brands in telecom, banking and food delivery categories, the Indian brands do not post too much content on Instagram. They could try to increase the quantity of images while considering the above suggestions.

5.2. Implications

Our empirics-first approach resulted in findings that would have implications for both the researchers and the managers.

5.2.1. Implications for researchers

5.2.1.1. Text data-based implications

We used VADER to calculate the brand position. The researchers may find other ways of calculating brand position. Given that VADER is more suited for handling Twitter data, other methods are certainly appropriate if the researchers decide to use text data from sources other than Twitter.

The traditional approach of examining the most frequent words has its merits. However, analyzing the most frequent words with or without the word clouds is not the only option available to the researchers. They could also examine the long-tail of the word distribution that uncovers the infrequently used words in the text messages. These words could provide insights about the newly unfolding attitudes of consumers. An example of the long-tail distribution, illustrated in Figure 5.1, is based on reviews the from epinions.com, Amazon.com and Yahoo.com.

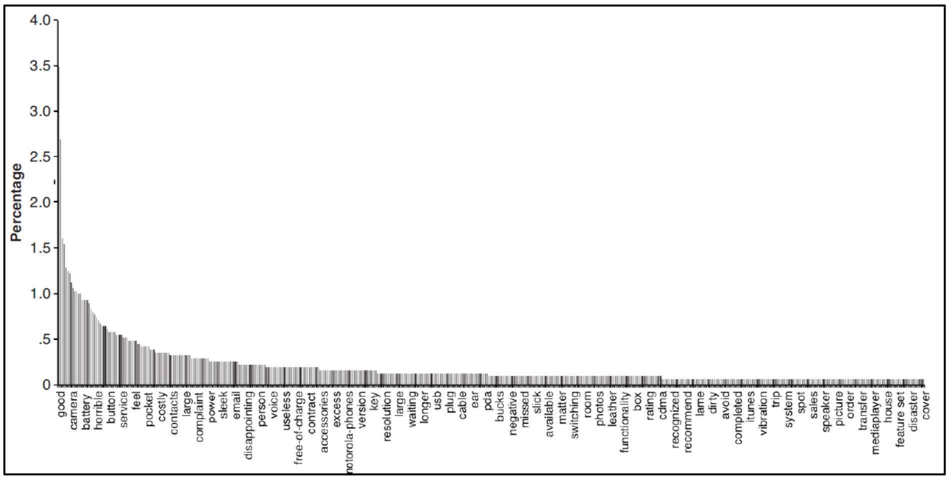

Figure 5.1. Long-tail distribution of words from Twitter data (Source: Tirunillai and Tellis, 2014)

Aaker’s brand personality dimensions faced some criticism in recent times (Kumar, 2018). Aaker’s scale might need to undergo modifications in the Indian context. Instead of sincerity and ruggedness, Ahmad and Thyagaraj (2017) proposed using popularity, trendiness and integrity dimensions in the brand personality of Indian brands. Since there is no noticeable uptake of these modified versions of Aaker’s scale in the Indian context, we did not use any of them in our study. However, other researchers could download the SMM’s open-source code to include these new dimensions and measure the brand personalities afresh.

The position and sentiment can be studied at tweet level instead of corpus level. Some of the tweets with poor position score or sentiment score could be closely examined to see which words are most likely to generate positive position and sentiment.

We make three significant contributions to the literature on brand personality and brand sentiment. To the best of our knowledge, this is the first study to empirically compare the Indian brand personalities using Pamuksuz, Yun and Humphreys model. Its transformers-based technology processes a text segment in both forward and backward directions, thereby producing better results than a long short-term memory network (LSTM) model (Hochreiter &

Schmidhuber, 1997). Since the transformers use the phenomenon of self-attention, they are considered to understand the context better than a typical LSTM model.

Brand sentiment present in the user-generated content is being studied for more than a decade. The brand-related posturing present in the firm-generated content has not been explored in detail. By measuring the brand position of 13 brands, we provided the baseline data. The future studies can be compared with this baseline data. A baseline can also be developed for other brands and other sample sizes.

Ours is the first study to establish a positive association between communicated brand personality and brand position. Our study did not yield clear evidence of any association between perceived brand personality and brand sentiment. Researchers may expand upon our work to build predictive models (Kumar & Vannan, 2021) for these constructs.

5.2.1.2. Image data-based implications

Two primary challenges slowed down the image analytics research in marketing. First, the heterogeneity of images results in different amounts of useful information from each image. Therefore, the images require careful evaluation prior to being included in the sample. Second, there does not exist a method that directly translates image content into marketing variables such as usability, performance, aesthetics and price.

Despite these challenges, the current study makes a strong argument in favour of using computer vision for brand management purposes. Apart from contributing to the literature, the results from the image data analysis have both methodological and managerial implications.

5.2.2. Methodological implications

Having contributed a new approach of brand identity – brand image alignment measurement using images, our work has a few methodological implications.

5.2.2.1. Image data-based implications

A clear advantage of our approach is that it avoids manual analysis of image contents. Considering the hundreds of images per brand that appear on Instagram every week, a fast and reliable approach is needed to rapidly intervene in cases of serious misalignment of brand identity and image. Our study expanded the knowledgebase of brand management approaches that use unstructured data.

By combining a pre-trained deep learning model with other open-source methods, we demonstrated a scalable approach. It is also a model of parsimony due to the usage of image data alone. But the dendrograms indicate that the inherent heterogeneity in the images would warrant a bigger sample to refine the brand clusters. As the monitoring increases, the corpus of images increases. The deep learning model's scalability makes it an appropriate choice for analyzing large amounts of image data.

Since the Instagram API provides information about how many users liked an image, our approach could be directly extended to investigate which images make a better advertising copy. It could lead to an insightful predictive model which estimates the number of 'likes' that would be garnered by an image. Valence is the number of likes and dislikes garnered by a post. Though valence by itself is insufficient to reach any conclusions about the effectiveness of social media posts, it could become part of a set of metrics that present the holistic picture of a brand's identity or image.

An inflexible analytics approach would be of little interest in the current era of open-source software and pre-trained models available on the cloud platforms. Our approach has a flexible architecture. A company could replace Google Cloud Vision with another deep learning model of its choice and still infer valuable marketing insights. The effectiveness of other deep learning models should be evaluated while considering the costs and efforts involved in integrating it into the approach. Alternatives to GCV such as Clarifai offer a limited-use free trial; YOLO offers the complete source code.

5.3. Implications for brand managers

5.3.1. Text data-based implications

This study has practical implications for the brand managers. They could monitor the brand performance as often as daily and make necessary changes in their Twitter messaging. Such frequency and rapidity are impossible with survey-based methods. Machine learning methods in our study clearly demonstrated the feasibility of mining impactful insights from text data.

As propounded by the associative network theory (Collins & Loftus, 1975), consumers have networks wherein information about products resides in the consumer's knowledge network as interconnected nodes. Brand managers can build these networks from Twitter data using open-source python packages such as NetworkX. These networks help the brand managers identify the appropriate influencers for their brands. The endorsement of well-chosen celebrities can possibly generate positive returns in the stock market in India (Agnihotri & Bhattacharya, 2018).

The employee-generated content can be as important as user-generated content in shaping the perceived brand personality (Hu et al., 2019). If the employees of certain firms are allowed to write unbiased reviews on social media, brand managers could use them as additional inputs to decipher an arguably different set of brand personality scores.

At times, the incongruence between the communicated and perceived brand personalities could be an indication that the consumers are nudging the brand towards a new position in some personality dimensions. It is advisable for the firms to heed to the voluntary cues by the consumers in their social media posts.

The effectiveness of a message may also depend on the language used by the firm. For example, Swiggy and Zomato's brand managers should study the impact of Twitter messages about local delicacies in local languages. Nederstigt and Hilberink-Schulpen (2018) undertook a questionnaire-based study to observe that the messaging in local languages may increase the consumers' positive perception about the brand as well as purchase intention. The text-processing toolkits in Indian languages are gradually springing up in the open-source software arena; they would expedite the managers' efforts to understand the positive impact of local language messaging.

The brand managers could undertake studies like ours to benchmark their brands' perceptions against competing brands. Since the competitors' tweets are publicly available via Twitter API, the analysis of the competing brands is economical and feasible.

The branding difficulties of technology companies are detailed in Huang and Dev (2020), Keller and Swaminathan (2019), and Blankson and Kalafatis (1999). A major company such as Wipro has shown the least alignment between communicated and perceived brand personality on Twitter. Wipro and other technology brands in a similar predicament might find it beneficial to implement some of our suggestions.

Our study included brands as young as the 9-year-old Swiggy to 117-year-old Canara Bank. There is no indication that older brands can certainly outperform younger brands, or vice versa, on Twitter. A younger brand such as Jio demonstrated superior positioning on Twitter. It could behove a mature brand to avoid complacency and utilize the social media communications judiciously.

5.3.2. Image data-based implications

This study has several implications for the brand managers.

High-technology firms often struggle with branding (Keller and Swaminathan, 2019). Brand management requires more than creative names, hoary past and bombarding the social media with humongous number of images. These firms need to further calibrate their messaging on Instagram to increase engagement and revenues. While the employees are focusing on posting what makes them happy at work, the firms are posting text announcements in the form of images. Technology brands may need to focus on the emotional aspects of the workplace environment and the employee value proposition in the images posted on Instagram.

Another dynamic that is unique to the technology brands is that there is no new entrant among the top-75 list or top-100 list considered by us. So, the firms that have been operating in India for decades need to compare themselves with some of the newer technology firms or startups. Such an exercise would keep them nimble and help them avoid any complacence.

It is well-known that the customers' impressions of events, encounters or episodes coalesce into the brand image. Customers at times compare two or more brands in a single image. The brand manager of a firm may analyze those images in close detail using TF-IDF and topic modeling to understand their brand’s relative standing.

The user-generated images may also include sarcastic or negative content. There are a few technological challenges involved in correctly interpreting sarcasm in the text embedded in the images. The brand managers could look at those image labels to understand the specific causes of dissatisfaction.

A brand that has been in the market for many decades could be outperformed by a relatively newer brand on social media. This was noticed in the case of older brands such as Zomato and Wipro. These brands need to pro-actively monitor their brand performance. A longitudinal study over a period of time helps these older brands how to improve their image contents.

The brand image is predominantly driven by the users, in spite of the concerns about spurious posts and employees posting as regular users. Temporary blips in alignment might not need an intervention. In cases where a brand image is extremely different from the brand identity over a longer period of time, the managers should heed to the signals by the users. If the users are telling the firm how a brand should position itself, the firm cannot afford to ignore the message.

To summarize succinctly, it is recommended that a brand manager should (a) strike a judicious balance between text and pictures in an image, (b) use more pictorial content to evoke the emotions associated with the brand, (c) understand and reduce the reasons for dissatisfaction expressed in the users' images, (d) realize that there might be no linear improvement in the brand identity or brand image with the years in existence, (e) utilize the signaling by the customers in the identity-image alignment information in order to maneuver the brand positioning in the brand clusters.

Chapter 6

CONCLUSION

This chapter summarizes our contributions. It also lists the limitations and future directions for research in the area of brand management on social media.

6.1. Summary of the contributions

Our contributions in the areas of marketing analytics using text data and image data are succinctly summarized in Table 6.1. They are explained under 'What's new ?' and 'So what ?' headings as suggested in Faff (2013).

Table 6.1. Recap of the contributions

Contribution	What's new ?	So what ?
1.	Studied the communicated and perceived brand personality of major Indian brands using Twitter (X) data.	Tracking the users' response to messaging could help reduce dissonance and disengagement.
2.	Developed an approach for comparing brand identity with brand image using Instagram images.	A reliable way of extracting insights using images is now available. Information other than valence is essential.
3.	Proposed a framework for managing the brands on social media.	Significant progress towards an integrated framework for analyzing data from multiple platforms.

Our first contribution provided valuable quantitative information about the communicated and perceived brand personalities of major Indian brands on Twitter (X). This information allows the brands to understand what is the gap and explore ways to bridge the gap. A good alignment between the communicated and perceived personalities could possibly result in a competitive advantage because of increased customer engagement with the brand.

Our second contribution is a novel way of comparing brand identity with brand image using Instagram images. This approach does not seem to exist in the marketing analytics literature. Our scalable approach facilitates a consistent and objective way of tracking a given brand's standing on Instagram.

Our third contribution is the creation of a framework that currently performs text analytics and image analytics on brand-related unstructured data. This framework is a result of incremental changes to Pamuksuz, Yun and Humphreys framework (Pamuksuz et al., 2021). It is quite different from the framework proposed by Klostermann et al. (2018) in terms of parsimony, marketing objectives and the technical stack.

6.2. Limitations and directions for future research

Our study has a few limitations which provide opportunities for further research.

6.2.1. Future research with text data

- Algorithms to detect misinformation in text data have been appearing in computer science literature for more than a decade. Using one of these algorithms with appropriate training data, it is possible to produce a machine learning model which will detect misinformation in Twitter data. This model can then be used to remove misinformation from a given sample of tweets.

  However, the misinformation and disinformation datasets for the 13 Indian brands used in our study do not exist in the public domain. So, we could not use any machine learning algorithm for detecting misinformation from the tweets. Many misleading tweets are deleted by their authors after other users or Twitter flag their dubious nature.

- We investigated tweets from more than 27,000 unique Twitter users. Still, there is the possibility of online herding behavior (Huang and Chen, 2006) resulting in biased tweets. To offset its effects, a bigger sample size is needed.
- SMM was available at socialmediamacroscope.org till August 31, 2022. It then switched from online hosting mode to free download mode during the period of our study. SMM's complicated software architecture could create challenges to the social science researchers in running SMM on their own machines. University of Illinois at Urbana-Champaign is attempting to revive SMM at smm.ncsa.illinois.edu at the time of writing this thesis.
- Combining the tweets with data from other social media platforms could lead to different sentiment values as suggested by Schweidel and Moe (2014). Fortunately, the SMM architecture can be modified to feed data from Facebook, Reddit and other sources.
- Regional differences in brand perceptions may exist in a vast country like India. For example, it is possible that the consumers in Karnataka state of India might have a more favorable impression of Canara Bank compared to others. However, any effort to study such differences would be hampered by the fact that some users choose to disclose their location and others do not.

6.2.2. Future research with image data

- Google Cloud Vision is probably not trained using brand-related images. So, it does not always provide the image labels that resonate with the marketing people. Other options such as the artificial intelligence-based Clarifai model should be investigated to see if they generate better results. Although the YOLO version 2 model did not receive a favorable evaluation in the past (Nanne et al., 2020), its newer version 7 could be a worthy alternative to Google Cloud Vision. YOLO is a free product, while GCV and Clarifai are commercial products. Clarifai is relatively less opaque compared to GCV; it disclosed a certain amount of information about the possible number of labels that could be generated by its computer vision model. Also, Clarifai offers a general model and a few specialized models.

- It is a documented fact that the young people in the age group of 18-19 years constitute the biggest Instagram user group in India (Statista, 2023). Since the demographical information was not included, our study does not offer actionable intelligence about specific user groups. Improvements in this direction are predicated upon the changes in Instagram API.
- The current study focused on Indian service brands. However, the methodology is not specifically designed for Indian brands; it is applicable to brands from other countries as well. Since a cross-cultural or cross-country comparison of brands is feasible, it deserves serious consideration.
- Our study built the brand clusters without using any specific attributes. If a method is developed for extracting marketing attributes from image contents, it would lead to useful predictive and prescriptive analytics. Although some researchers might attempt to extract attributes by displaying images in the form of survey to participants, it is costly, slow and not scalable. Deep learning methods are required for consistent and replicable extraction of marketing attributes. Such methods would help in building brand clusters based on specific attributes. Brand clusters based on income groups, age groups or geographic locations are obviously more interesting to the chief marketing officers.
- Instagram CEO Adam Mosseri stated that "more and more of Instagram is going to become video over time". It is imperative to include the short videos known as reels from Instagram in the next-level analysis of visual branding. We did not employ advanced methods for integrating image data and video data to expand the brand management knowledgebase.

6.3. Conclusion

Twitter is a free communications channel with millions of users and hundreds of daily posts per brand. Communicated brand personality might not change too often. But the consumers' sentiments and perceptions about a brand on social media can change very rapidly. Some of the commercial social media analytics tools operate as black-boxes and their results are unreliable (Hayes et al., 2021); our study used several open-source methods. By choosing brands from several categories, we increased the likelihood of generalizability of our results. The incongruence between firm-generated content and user-generated content in our results warrants further research on the interplay between brand personality, position and sentiment in the context of Indian brands.

Posting a large number of catchy images on social media does not result in effective differentiation of a brand. The advances in image analytics by prominent studies such as Nanne et al. (2020) did not focus on firm-generated images. The comparison of firm's images with users' images needed to be addressed, especially in the context of Indian brands. To address this crucial gap, the current study employed an affordable mix of open-source and commercial software. Instead of using technology for operations tools alone, it is prudent to utilize the technology for better positioning on the electronic marketing channels (Yu, 2022). Our research could trigger relatively more frequent monitoring of the alignment between the brand identity and brand image on the digital media platforms.

Comparing brand identity with brand image using Instagram is, to the best of our knowledge, a novel contribution. By using text and images as input to the cutting-edge deep learning models, we are on the path to establishing an integrated framework for tracking the brand consistency across the social media platforms.

6.4. Disclosure statement

We received a $1000 grant [Grant ID 209475195] under the Google Cloud Research Credits program in March 2022. Google LLC played no role in the research design, data collection, data analysis, articles' preparation and thesis preparation.

We paid approximately €80 to clickworker.com for evaluating the validation platform, validating the labeled tweets and validating the labeled images. Clickworker.com was unaware of our overall research objectives. Also, the validation was done in a double-blind manner.

REFERENCES

1. Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research*, 34(3), 347-356.
2. Aaker, J. L., Benet-Martinez, V., & Garolera, J. (2001). Consumption symbols as carriers of culture: A study of Japanese and Spanish brand personality constructs. *Journal of Personality and Social Psychology*, 81(3), 492.
3. Aaker, D. A. (2012). Building strong brands. Simon and Schuster.
4. Agnihotri, A., & Bhattacharya, S. (2018). The market value of celebrity endorsement: Evidence from India reveals factors that can influence stock-market returns. *Journal of Advertising Research*, *58*(1), 65-74.
5. Ahmad, A., & Thyagaraj, K. S. (2017). An empirical comparison of two brand personality scales: Evidence from India. *Journal of Retailing and Consumer Services*, *36*, 86-92.
6. Ahmad, S., Srivastava, A. and Sharma, S. (2021). Customer sentiment towards freebies in telecom sector: a social media mining approach. *International Journal of Business Information Systems*, *38*(2), pp.240-253.
7. American Marketing Association (1960), Definition of Service, American Marketing Association, Chicago, IL.
8. Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.
9. Ansari, A., Li, Y., & Zhang, J. Z. (2018). Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach. *Marketing Science*, *37*(6), 987-1008.
10. Arabadzhyan, A., Figini, P., & Vici, L. (2021). Measuring destination image: a novel approach based on visual data mining. A methodological proposal and an application to European islands. *Journal of Destination Marketing & Management,* 20, 100611.
11. Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index-insights from Facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*, 49, 86-101.
12. Arora, N., Prashar, S., Vijay, T. S., & Parsad, C. (2023). Exploring the Effect of Personality Congruencies on Brand Identification and Purchase Intentions. *Journal of Global Scholars of Marketing Science*, *33*(2), 186-209.
13. Austin, J. L. (1962). How to Do Things with Words: The William James Lectures Delivered in Harvard University in 1955.

14. Bakri, M., Krisjanous, J., & Richard, J. E. (2020). Decoding service brand image through user-generated images. *Journal of Services Marketing*.
15. Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, *46*, 557-590.
16. Batra, R., Ahuvia, A., & Bagozzi, R. P. (2012). Brand love. *Journal of Marketing*, 76, 1–16.
17. Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1-25.
18. Bhatia, R., Gupta, A., Vimalkumar, M., & Sharma, D. (2023). Factors affecting Consumer Brand Sabotage virality: a study of an Indian brand #boycott. *Information Systems and e-Business Management*, 1-28.
19. Bianchi, F., Terragni, S., & Hovy, D. (2020). Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *arXiv preprint arXiv:2004.03974*.
20. Blankson, C., & Kalafatis, S. P. (1999). Issues and challenges in the positioning of service brands: a review. *Journal of Product & Brand Management*, *8*(2), 106-118.
21. Blei, D.M., Ng, A.Y. & Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, Vol. 3, No. 2, pp.993–1022.
22. Brand Finance (2021). The Brand Finance India 100. https://brandfinance.com/wp-content/uploads/1/the_brand_finance_india_100.pdf.
23. Cayla, J., & Eckhardt, G. M. (2007). Asian brands without borders: regional opportunities and challenges. *International Marketing Review*, *24*(4), 444-456.
24. Chen, Q., & Rodgers, S. (2006). Development of an instrument to measure web site personality. *Journal of Interactive Advertising*, *7*(1), 4-46.
25. Clarifai (2023), https://clarifai.com, (Last accessed October 28, 2023).
26. Colicev, A., Kumar, A., & O'Connor, P. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *International Journal of Research in Marketing*, *36*(1), 100-116.
27. Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82(3), 407–428.
28. Cooper, T., Stavros, C. & Dobele, A.R. (2019), Domains of influence: exploring negative sentiment in social media. *Journal of Product & Brand Management*, Vol. 28 No. 5, pp. 684-699.

29. Culotta, A., & Cutler, J. (2016). Mining brand perceptions from Twitter social networks. *Marketing Science*, 35(3), 343–362.
30. Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web* (pp. 273-274).
31. de Chernatony, L. (1999). Brand management through narrowing the gap between brand identity and brand reputation. *Journal of Marketing Management*, Vol. 15, No. 1–3, pp. 157–179.
32. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
33. Dibb, S., Simkin, L., Pride, W. M., & Ferrell, O. C. (2019). *Marketing: Concepts and Strategies*. Cengage Learning EMEA.
34. Donelson, C., Sutter, C., Pham, G. V., Narang, K., Wang, C., & Yun, J. T. (2021). Using a Machine Learning Methodology to Analyze Reddit Posts regarding Child Feeding Information. *Journal of Child and Family Studies*, 30(5), 1290-1298.
35. Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 281-285).
36. Dvornechuck, A. (2020). Brand personality: Traits of top brands. Ebaqdesign Blog.
37. Eisend, M., & Stokburger-Sauer, N. E. (2013). Brand personality: A meta-analytic review of antecedents and consequences. *Marketing Letters*, *24*, 205-216.
38. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, *88*, 303-338.
39. Faff, R. (2013). Mickey Mouse and the IDioT principle for assessing research contribution: discussion of 'Is the relationship between investment and conditional cash flow volatility ambiguous, asymmetric or both?'. *Accounting & Finance*, *53*(4), 949-960.
40. Ferwerda, B., & Tkalcic, M. (2018). You are what you post: What the content of Instagram pictures tells about users' personality. In *The 23rd International on Intelligent User Interfaces, March 7-11, Tokyo, Japan*. CEUR-WS.
41. Févotte, C., & Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. *Neural Computation*, *23*(9), 2421-2456.

42. Freling, T. H., Crosno, J. L., & Henard, D. H. (2011). Brand personality appeal: conceptualization and empirical validation. *Journal of the Academy of Marketing Science*, 39(3), 392-406.
43. Fromkin, Howard L. (1970). Effects of Experimentally Aroused Feelings of Indistinctiveness upon Valuation of Scarce and Novel Experiences. *Journal of Personality and Social Psychology*, 16 (3), 521-29.
44. Gesing, S., Brandt, S., Bradley, S., Potkewitz, M., Kee, K., Whysel, N., Perri, M., Cleveland, S., Rugg, A., & Smith, J. (2021). A Vision for Science Gateways: Bridging the Gap and Broadening the Outreach. *Practice and Experience in Advanced Research Computing*, (pp. 1-8).
45. Geuens, M., Weijters, B., & De Wulf, K. (2009). A new measure of brand personality. *International Journal of Research in Marketing*, 26(2), 97-107.
46. Ghorbani, M., Karampela, M., & Tonner, A. (2022). Consumers' brand personality perceptions in a digital world: A systematic literature review and research agenda. *International Journal of Consumer Studies*, 00, 1-32.
47. Golder, P. N., Dekimpe, M., An, J. T., van Heerde, H. J., Kim, D., & Alba, J. W. (2023). Learning from data: An empirics-first approach to relevant knowledge generation. *Journal of Marketing*.
48. Google (2022). Vision API. Retrieved from https://cloud.google.com/vision/?hl=de, (Last accessed December 28, 2022).
49. Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
50. Ham, C. D., & Lee, H. S. (2015). Internet media personality: scale development and advertising implications. *International Journal of Advertising*, 34(2), 327-349.
51. Harris, F. & de Chernatony, L. (2001). Corporate branding and corporate brand performance. *European Journal of Marketing*, Vol. 35, No. 3–4, pp. 441–456.
52. Hayes, J. L., Britt, B. C., Evans, W., Rush, S. W., Towery, N. A., & Adamson, A. C. (2021). Can social media listening platforms' artificial intelligence be trusted? Examining the accuracy of Crimson Hexagon's (now Brandwatch Consumer Research's) AI-Driven analyses. *Journal of Advertising*, *50*(1), 81-91.
53. Heinberg, M., Ozkaya, E. & Taube, M. (2018), "Do corporate image and reputation drive brand equity in India and China? Similarities and differences", *Journal of Business Research*, Vol. 86, pp. 259–268.

54. Hernández-Méndez, J., & Muñoz-Leiva, F. (2015). What type of online advertising is most effective for eTourism 2.0? An eye tracking study based on the characteristics of tourists. *Computers in Human Behavior*, *50*, 618-625.
55. Hewett, K., Rand, W., Rust, R. T., & Van Heerde, H. J. (2016). Brand buzz in the echoverse. *Journal of Marketing*, *80*(3), 1-24.
56. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), pp. 1735-1780.
57. Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*, Sage; Newbury Park, CA.
58. Hsieh, M.H. (2004), "Measuring global brand equity using cross-national survey data", *Journal of International Marketing*, Vol. 12 No. 2, pp. 28–57.
59. Hu, Y., Xu, A., Hong, Y., Gal, D., Sinha, V., & Akkiraju, R. (2019). Generating business intelligence through social media analytics: Measuring brand personality with consumer-, employee-, and firm-generated content. *Journal of Management Information Systems*, 36(3), 893-930.
60. Huang, J. H., & Chen, Y. F. (2006). Herding in online product choice. *Psychology & Marketing*, *23*(5), 413-428.
61. Huang, M. H., & Dev, C. S. (2020). Growing the service brand. *International Journal of Research in Marketing*, *37*(2), 281-300.
62. Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).
63. Ibrahim, N. F., & Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decision Support Systems*, *121*, 37-50.
64. Internetlivestats.com (2023), Twitter usage statistics. https://www.internetlivestats.com/twitter-statistics.
65. Jaakonmäki, R., Müller, O., & Vom Brocke, J. (2017). The impact of content, context, and creator on user engagement in social media marketing.
66. Kane, G. C., & Pear, A. (2016). The rise of visual content online. *MIT Sloan Management Review*.
67. Kantar (2020). BrandZ Top 75 most valuable Indian brands 2020. https://www.kantar.com/campaigns/brandz-downloads/brandz-top-75-most-valuable-indian-brands-2020

68. Kantar (2021). India’s most purposeful brands 2021. https://www.kantar.com/inspiration/brands/2021-indias-most-purposeful-brands-amazon-asian-paints-and-tata-tea.
69. Kaplan, Andreas & Michael Haenlein (2010), “Users of the World, Unite! The Challenges and Opportunities of Social Media,” *Business Horizons*, 53, 1, 59–68.
70. Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*, 57(1), 1-22.
71. Keller, K. L., & Swaminathan, V. (2019). *Strategic brand management: Building, measuring, and managing brand equity*. London: Pearson.
72. Kim, Y. E., Lee, J. W., & Lee, Y. K. (2008). Relationship between brand personality and the personality of consumers, and its application to corporate branding strategy. *Journal of Global Academy of Marketing Science, 18*(3), 27-57.
73. Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2018). Extracting brand information from social networks: Integrating image, text, and social tagging data. *International Journal of Research in Marketing*, 35(4), 538-556.
74. Kotler, P. (2000). Marketing Management. New Jersey: Prentice Hall
75. Kuksov, D., Shachar, R., & Wang, K. (2013). Advertising and consumers' communications. *Marketing Science*, 32(2), 294-309.
76. Kumar, A. (2018). Story of Aaker’s brand personality scale criticism. *Spanish Journal of Marketing-ESIC*, Vol. 22 No. 2, pp. 203-230.
77. Kumar, U. D. (2017). *Business analytics: The science of data-driven decision making*. Wiley.
78. Kumar, V., & Vannan, M. (2021). It takes two to tango: Statistical modeling and machine learning. *Journal of Global Scholars of Marketing Science*, *31*(3), 296-317.
79. Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. 31st International Conference on Machine Learning, Beijing, China, 1188–1196.
80. Lieven, T. & Hildebrand, C. (2016), “The impact of brand gender on brand equity”, *International Marketing Review*, Vol. 33 No. 2, pp. 178–195.
81. Liu, X., Burns, A. C., & Hou, Y. (2017). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising*, *46*(2), 236-247.
82. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

83. Liu, L., Dzyabura, D., & Mizik, N. (2020). Visual listening in: Extracting brand image portrayed on social media. *Marketing Science*, 39(4), 669-686.
84. Martineau, P. (1958). The personality of the retail store. *Harvard Business Review*, 36, 47–55.
85. Masiello, B., Bonetti, E., & Izzo, F. (2020). Multiple identities of a festival: Intended, communicated and perceived brand personality in the social media environment. *International Journal of Contemporary Hospitality Management,* Vol. 32 No. 2, pp. 749-768.
86. Mazloom, M., Rietveld, R., Rudinac, S., Worring, M., & Van Dolen, W. (2016, October). Multimodal popularity prediction of brand-related social media posts. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 197-201).
87. McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276-282.
88. Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, https://doi.org/10.48550/arXiv.1301.3781 (accessed 19 July 2022).
89. Mint (2022, August 21). *Zomato withdraws 'Mahakal' thali ad featuring Hrithik Roshan following protests*. https://www.livemint.com/news/india/zomato-withdraws-mahakal-thali-ad-featuring-hrithik-roshan-following-protests-11661077919341.html.
90. Mitra, S., & Jenamani, M. (2020). OBIM: A computational model to estimate brand image from online consumer review. *Journal of Business Research*, *114*, 213-226.
91. Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, *40*(10), 4241-4251.
92. Moussa, S. (2019). An emoji-based metric for monitoring consumers' emotions toward brands on social media. *Marketing Intelligence & Planning,* Vol. 37 No. 2, pp. 211-225.
93. Mowen, J. C. & M. Minor (2001). Consumer Behavior: A Framework, USA: Prentice Hall
94. Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). What yelp fake review filter might be doing?. In *Proceedings of the international AAAI conference on web and social media* (Vol. 7, No. 1, pp. 409-418).
95. Nagarkar, P., Amarnani, L., & Doshi, D. (2021). Social Media Intelligence for Brand Analysis. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT),* (pp. 1-7). IEEE.

96. Nanne, A. J., Antheunis, M. L., van der Lee, C. G., Postma, E. O., Wubben, S., & van Noort, G. (2020). The use of computer vision to analyze brand-related user generated image content. *Journal of Interactive Marketing*, 50, 156-167.
97. Nebenzahl, I. D., Jaffe, E. D., & Lampert, S. I. (1997). Towards a theory of country image effect on product evaluation. *MIR: Management International Review, 37*(1), 27–49.
98. Nederstigt, U., & Hilberink-Schulpen, B. (2018). Advertising in a Foreign Language or the Consumers' Native Language?. *Journal of International Consumer Marketing*, *30*(1), 2-13.
99. Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543.
100. Opoku, R., Abratt, R., & Pitt, L. (2006). Communicating brand personality: are the websites doing the talking for the top South African business schools?. *Journal of Brand Management*, 14(1), 20-39.
101. Ordenes, F.V., Ludwig, S., De Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research*, *43*(6), 875-894.
102. Ordenes, F. V., Grewal, D., Ludwig, S., de Ruyter, K., Mahr, D., & Wetzels, M. (2018). Cutting through content clutter: How speech and image acts drive consumer sharing of social media brand messages. *Journal of Consumer Research*.
103. Pamuksuz, U., Yun, J. T., & Humphreys, A. (2021). A Brand-New Look at You: Predicting Brand Personality in Social Media Networks with Machine Learning. *Journal of Interactive Marketing*, 56, 55-69.
104. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in information retrieval*, 2(1–2), 1-135.
105. Park, C. W., Jaworski, B. J. & MacInnis, D. J. (1986). Strategic brand concept-image management. *Journal of Marketing*. Vol. 50, October, pp. 135–145.
106. Park, J. W., Cho, E. Y., & Kim, H. W. (2016). Examining context-specific social media marketing strategies. *Asia Pacific Journal of Information Systems*, 26(1), 143-162.
107. Pareek, V., & Harrison, T. (2020). SERVBID: the development of a B2C service brand identity scale. *Journal of Services Marketing*, *34*(5), 601-620.
108. Ranjan, S., Singh, I., Dua, S., & Sood, S. (2018). Sentiment analysis of stock blog network communities for prediction of stock price trends. *Indian Journal of Finance*, *12*(12), 7-21.

109. Rao, C. Radhakrishna (1995), “Use of Hellinger Distance in Graphical Displays,” in *Multivariate Statistics and Matrices in Statistics*. Leiden, The Netherlands: Brill Academic Publisher, 143–61.

110. Ray, K., & Sharma, M. (2022). Benchmarking global IT majors' brand strength towards global branding of Asian IT organisations. *International Journal of Business and Globalisation*, *31*(2), 170-197.

111. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

112. Rust, R. T., Rand, W., Huang, M. H., Stephen, A. T., Brooks, G., & Chabuk, T. (2021). Real-time brand reputation tracking using social media. *Journal of Marketing*, *85*(4), 21-43.

113. Samiee, S. (1994). Customer evaluation of products in a global market. *Journal of International Business Studies, 25*(3), 579–604.

114. Samtani, S., Zhu, H., Padmanabhan, B., Chai, Y., Chen, H., & Nunamaker Jr, J. F. (2023). Deep learning for information systems research. *Journal of Management Information Systems*, *40*(1), 271-301.

115. Schweidel, D. A., & Moe, W. W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research*, *51*(4), 387-402.

116. Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

117. Sengupta, A., Wesley, S., Cavender, R., & Lee, M. Y. (2022). Global vs local: Analysis of the consumer-brand relationships in India. *International Journal of Retail & Distribution Management*, *50*(3), 361-376.

118. Sharma, A., Patro, S., & Chaudhry, H. (2022). Brand identity and culture interaction in the Indian context: a grounded approach. *Journal of Advances in Management Research*, *19*(1), 31-54.

119. Sindhani, M., Parameswar, N., Dhir, S., & Ongsakul, V. (2019). Twitter analysis of founders of top 25 Indian startups. *Journal for Global Business Advancement*, *12*(1), 117-144.

120. Sirgy, M. J. (1982). Self-concept in consumer behavior: A critical review. *Journal of Consumer Research*, 9(3), 287–300.

121. Sirgy, M. J. (2018). Self-congruity theory in consumer behavior: A little history. *Journal of Global Scholars of Marketing Science*, *28*(2), 197-207.

122. Smith, A.N., Fischer, E. & Yongjian, C. (2012). How does brand-related user generated content differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing,* Vol. 26 No. 2, pp. 102-113.

123. Sridhar, S., & Srinivasan, R. (2012). Social influence effects in online product ratings. *Journal of Marketing*, 76(5), 70–88.

124. Srinivasan, N., Jain, S. C., & Sikand, K. (2004). An experimental study of two dimensions of country-of-origin (manufacturing country and branding country) using intrinsic and extrinsic cues. *International Business Review, 13*(1), 65–82.

125. Srivastava, V., & Kalro, A. D. (2019). Enhancing the helpfulness of online consumer reviews: the role of latent (content) factors. *Journal of Interactive Marketing*, *48*(1), 33-50.

126. Srivastava, K., & Sharma, N. K. (2016). Consumer perception of brand personality: An empirical evidence from India. *Global Business Review*, *17*(2), 375-388.

127. Statista (2022). Leading countries based on Instagram audience size as of January 2022, https://www.statista.com/statistics/578364/countries-with-most-instagram-users. (Accessed 01, January 2023).

128. Statista (2023). Instagram – Statistics and Facts. https://www.statista.com/topics/1882/instagram/

129. Tang, T., Fang, E., & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, *78*(4), 41-58.

130. Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research,* 51(4), 463-479.

131. Triandis, H. C., Individualism and Collectivism 1995 Westview Press Boulder.

132. Trivedi, S. K., & Singh, A. (2021). Twitter sentiment analysis of app based online food delivery companies. *Global Knowledge, Memory and Communication*. Vol. 70 No. 8/9, pp. 891-910.

133. *UN DESA Policy Brief No. 153: India overtakes China as the world's most populous country | Department of Economic and Social Affairs*. (2023). https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-no-153-india-overtakes-china-as-the-worlds-most-populous-country

134. Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97-121.

135. Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, March). Who says what to whom on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 705-714).

136. Wu, L., Dodoo, N. A., & Choi, C. W. (2023). Brand anthropomorphism on Twitter: communication strategies and consumer engagement. *Journal of Product & Brand Management*.

137. Yadav, M. L., Dugar, A., & Baishya, K. (2022). Decoding Customer Opinion for Products or Brands Using Social Media Analytics: A Case Study on Indian Brand Patanjali. *International Journal of Intelligent Information Technologies (IJIIT)*, *18*(2), 1-20.

138. Yu, B. (2022). How consumer opinions are affected by marketers: an empirical examination by deep learning approach. *Journal of Research in Interactive Marketing*, 16(4), 601-614.

139. Yun, J. T., Vance, N., Wang, C., Troy, J., Marini, L., Booth, R., Nelson, T., Hetrick, A., & Hodgekins, H. (2018). *The Social Media Macroscope*. In Gateways 2018. https://doi.org/10.6084/m9.figshare.6855269.v2

140. Yun, J. T., Pamuksuz, U., & Duff, B. R. (2019). Are we who we follow? Computationally analyzing human personality and brand following on Twitter. *International Journal of Advertising*, 38(5), 776-795.

141. Yun, J.T., Vance, N., Wang, C., Marini, L., Troy, J., Donelson, C., Chin, C.L. & Henderson, M.D. (2020). The social media macroscope: a science gateway for research using social media data. *Future Generation Computer Systems*, Vol. 111, pp.819–828.

142. Zarantonello, L., Grappi, S., Formisano, M., & Brakus, J. (2020). How consumer-based brand equity relates to market share of global and local brands in developed and emerging countries. *International Marketing Review*, *37*(2), 345-375.

143. Zhang, S., de Jong, M. D., & Gosselt, J. F. (2022). Microblogging for Engagement: Effects of Prior Company Involvement, Communication Strategy, and Emojis on Western and Chinese Users. *Journal of International Consumer Marketing*, 1-15.

APPENDIX A

Table A1. Theories cited in marketing studies using unstructured data

(Source: Balducci, B. & Marinova, D., 2018)

S. No.	Theory name
01.	Agency Theory
02.	Associative Network Theory
03.	Behavioral Consistency Theory
04.	Cognitive Response Theory
05.	Information Processing Theory
06.	Optimal Stimulation Level Theory
07.	Persuasive Argumentation Theory
08.	Random Utility Theory
09.	Social Impact Theory
10.	Search Theory
11.	Emotion Theory
12.	Utility Theory
13.	Characteristics Theory
14.	Customer Utility Theory
15.	Grounded Theory
16.	Self-Congruity Theory
17.	Social Identity Theory
18.	Social Influence Theory
19.	Organizational Identity Theory
20.	Social Cognition Theory
21.	Social Proof Theory
22.	Gatekeeper Theory
23.	Contrast Theory
24.	Outsourced Regulation Theory
25.	Role Theory
26.	Attribution Theory
27.	Consumer Information Search Theory
28.	Communication Accommodation Theory
29.	Communication Theory
30.	Diagnosticity of Information Theory
31.	Diffusion Theory
32.	Human Communication Theory
33.	Information Theory
34.	Prospect Theory
35.	Social Influence Theory
36.	Speech Act Theory

37.	Theory of Cognitive Dissonance
38.	Theory of Information Accessibility and Influences
39.	Theory of Weak Ties
40.	Social Learning Theory
41.	Emotion Regulation Theory
42.	Accessibility–Diagnosticity Theory
43.	Context Theory of Classified Learning
44.	Categorization Theory
45.	Two-Factor Theory
46.	Behavioral Theory

Table A2. Opoku's brand personality dictionary

Sincerity	Competence	Excitement	Ruggedness	Sophistication
ABOVE_BOARD	ABLE	ACTIVE	AL_FRESCO	A_LA_MODE
ACCOMMODATING	ABLE_BODIED	AGGRESSIVE	ALFRESCO	ALLURING
ACCURATE	ADEPT	ARTISTIC	ANIMAL	AMIABLE
ACTUAL	ADROIT	ARTY	ANIMALS	ANGELIC
AFFABLE	ASSIDUOUS	AUDACIOUS	ARDUOUS	APPEALING
APPROACHABLE	ASSURED	AUDACITY	BEEFY	ARISTOCRACY
APPROACHING	ASTUTE	AUTONOMOUS	BOISTEROUS	ARISTOCRAT
AUTHENTIC	AWARD_WINNING	AVANT_GARDE	BRUTAL	ARISTOCRATIC
BENEFICIAL	BLOOMING	AWE_INSPIRING	BUMPY	ARISTOCRATICAL
BENEVOLENT	BOOMING	AWESOME	CALLOUS	ATTRACTIVE
BENIGN	BRAINY	BOLD	CHALLENGE	BARONIAL
BLUNT	CELEBRATORY	BOLDNESS	CHALLENGING	BEAUTIFUL
BONAFIDE	CERTIFIED	BOOST	COARSE	BLUE_BLOOD
BRIGHT	COMPETENCE	BRACING	CONFRONTATI	BLUE_BLOODED
BUOYANT	COMPETENT	BRANDNEW	ON	BRUSH_UP
CANDID	COMPLETE	BRAND_NEW	COWBOY	CAPTIVATE
CHARITABLE	COMPREHENSIVE	BRAVE	CRAGGED	CAPTIVATING
CHEERFUL	CONCERN	BRAVERY	CRAGGY	CELEBRATED
CIVIL	CONCLUSIVE	BREATHTAKING	CRIMSON	CHARISMATIC
CIVILISED	CONFINED	BRISK	CRUDENESS	CHARM
CIVILITY	CONGLOMERATE	COLORFUL	CRUDITY	CHARMING
CIVILIZED	CONQUERING	COLOURFUL	CRUEL	CHERUBIC
CLEAN_CUT	CONSCIENTIOUS	COOL	DANGEROUS	CLASSY
CLEAR_CUT	CONSISTENT	COURAGE	DAUNTING	COSMOPOLITAN
COMMON	CONSTANT	COURAGEOUS	DAYBREAK	COTOURE
COMMONPLACE	CRAFTINESS	COURAGEOUSNESS	DAYSPRING	COURTIER
COMPANIONABLE	CRAFTY	COURANT	DEMANDING	CULTIVATED
COMPASSIONATE	CUNNING	CRAZY	DESERT	CULTURED
CONGENIAL	DEPENDABLE	CREATIVE	DIFFICULT	CUTE
CONTENT	DEXTEROUS	CREATIVITY	DURABLE	DANDYISH
CONVENTIONAL	DILIGENCE	CRISP	EFFORTFUL	DE_LUXE
CONVIVIAL	DILIGENT	CURRENT	ENDEAVOR	DELICATE
COOPERATIVE	DOINGWELL	DARING	ENDEAVOUR	DIGNIFIED
CORDIAL	DOMINANT	DAZZLING	ENDURE	DISTINCTION
CORRECT	ENTERPRISE	DESIGNER	EXTERNAL	DISTINGUISHED
COURTEOUS	ENTERPRISINGNESS	DETERMINED	EXTINCT	DOWNY
CUSTOMARY	EQUIPOTENT	EARLY	EXTREME	DULCET
DECENT	ESTABLISHMENT	ELECTRIFYING	EXTREMUM	EDIFICATION
DEFENSIBLE	EVERLASTING	ELEVATE	FEROCIOUS	ELEGANT
DIRECT	EXHAUSTIVE	EMANCIPATE	FORCIBLE	ELOQUENT
DISTINCTIVE	EXPERIENCED	EMANCIPATED	FRESCO	ENCHANT
DOWN_TO_EARTH	EXULTANT	ENERGISE	FRESHAIR	ENCHANTING
EARNEST	FAIL_SAFE	ENERGISING	FRONTIER	ENDEARING
EBULLIENT	FIRM	ENERGIZE	FURROW	ENGAGING
EMOTIONAL	FIRST_PLACE	ENERGIZING	GODFORSAKEN	ENNOBLING
EVERYDAY	FLOURISHING	ENLIVEN	GRANITELIKE	ENRAPTURE
EXISTENT	FOOLPROOF	ENLIVENING	GRANITIC	ENTHRAL
EXISTING	FOR_CERTAIN	ENTERPRISING	GRATING	ENTHRALL
FACT_BASED	FOREFRONT	EXALT	GRAVEL	ENTHRALLING
FACTUAL	GAINFUL	EXALTING	GRUELING	ENTICING
FAITHFUL	GENIUS	EXCITATION	GRUELLING	ENTRANCING
FORTHCOMING	GET_AHEAD	EXCITE	HARD	EPICUREAN
FORTHRIGHT	GIFTED	EXCITED	HARD_BOILED	ESTEEMED
FRANK	GLORIOUS	EXCITEMENT	HARDENED	ESTHETIC
FRIENDLY	GOVERNANCE	EXCITING	HARD_HITTING	EXCELLENT
GENEROUS	GUARANTEE	EXHILARATE	HARSH	EXCLUSIVE
GENIAL	GUARANTEED	EXHILARATING	HAZARDOUS	EXCLUSIVITY
GENUINE	HARDWORKING	EXUBERANT	HEAVY_DUTY	EXPENSIVE
GLAD	HARD_WORKING	FEISTY	HUNT	EXQUISITE
GOOD	HI_TECH	FORCEFUL	HUNTING	EXQUISITELY
GOOD_HEARTED	ILLUSTRIOUS	FRESH	HUSKINESS	EXTRAVAGANT

Sincerity	Competence	Excitement	Ruggedness	Sophistication
GOOD_HUMOURED	IMPERISHABLE	FRESHNESS	INHUMANE	EYE_CATCHING
GRACIOUS	IN_FRONT	GUTSY	INSENSITIVE	FABULOUS
GREGARIOUS	IN_CHARGE	HAPPENING	IRREGULAR	FANTABULOUS
GUILELESS	INDUSTRIAL	HEROIC	JAGGED	FASCINATING
HALE AND HEARTY	INDUSTRIALISE	HEROISM	JEANS	FASHIONABLE
HEALTHFUL	INDUSTRIALISED	HIGH_SPIRITED	JERKING	FEMALE
HEARTFELT	INDUSTRIALIZED	HIP	JERKY	FEMININE
HEARTY	INDUSTRIOUS	IMAGINATIVE	JOLTING	FIRST_CLASS
HELPFUL	INDUSTRIOUSNESS	INDEPENDENT	JOLTY	FIRST_RATE
HONEST	INDUSTRY	INDIVIDUAL	JUNGLE	FLOSSY
HONESTNESS	IN_NO_DOUBT	INNOVATIVE	LABOURIOUS	FLUENT
HONESTY	INTELLECTUAL	INSPIRING	LEATHERY	FRAGILE
HONORABLE	INTELLIGENT	INTREPID	MACHO	FRAGRANT
HONORABLENESS	JUBILANT	INVENTIVE	MANFULLY	FULGID
HONOURABLE	KNOWING	INVIGORATING	MANLY	GENTEEL
HONOURABLENESS	KNOWLEDGEABLE	IN_VOGUE	MANNISH	GENTLE
HUMANE	KNOWLEDGEABLE_A	JUVENILE	MASCULINE	GENTLEMANLIKE
HUMBLE	BOUT	LATEST	MAVERICK	GENTLEMANLY
INDISPUTABLE	LASTING	LIBERATED	MOUNTAINOUS	GENTLEWOMAN
INIMITABLE	LEADER	LIFTING	MOUNTAINS	GILDED
INSPIRED	LICENSE	LIVEN_UP	NERVE_RACKI	GLAMOROUS
IRREPLACEABLE	LOGICAL	LONE	NG	GLAMOUR
JOVIAL	LONG_LASTING	MODERN	NERVE_WRAC	GLAMOUROUS
KIN	LONG_LIVED	MODERN_DAY	KING	GLIB
KIND	LONG_SUFFERING	MODERNISTIC	OPEN_AIR	GLIB_TONGUED
KINDLY	LOYAL	MODERNNESS	OUTDOOR	GLITTERING
KINSHIP	LUCRATIVE	MODISH	OUTDOORS	GLOSSY
LEGITIMATE	MANUFACTURE	MOVING	OUTDOORSY	GOOD_LOOKING
LEGITIMATISE	MARKETABLE	NERVE	OUTER	GOOD_NATURED
LEGITIMATIZE	MECHANICAL	NEW	OUT_OF_DOOR	GORGEOUS
LEGITIMISE	MERCANTILE	NEWLY_ARISEN	OUT_OF_DOOR	GRACEFUL
LEGITIMIZE	METHODOLOGICAL	NIFTY	S	HANDSOME
LUCKY	METICULOUS	PLUCKY	OUTSIDE	HAUTE_COTOURE
MATCHLESS	MONEYMAKING	PRESENT	PACHYDERMA	HIGH_BORN
MAUDLIN	MONEY_MAKING	PRESENT_DAY	TOUS	HIGH_BROW
MERCIFUL	OUTSTANDING	PREVAILING	PERDURABLE	HIGHBROWED
MERRY	PAINSTAKING	PRISTINE	PERILOUS	HIGH_CLASS
MODEST	PARTNERSHIP	RECENT	PHYSICAL	HIGH_PROFILE
NATURAL	PERPETUAL	REFRESHED	POINTY	HIGH_STATUS
NOURISHING	PERSEVERING	REFRESHEN	POTHOLED	IN_STYLE
NOVEL	PERSISTENT	REFRESHFUL	PRAIRIE	IN_VOGUE
OBLIGING	POISED	REFRESHING	PRECARIOUS	INDULGENT
OLD_FASHIONED	POTENT	REPRESENT	PROHIBITED	LADY
OPEN	PRIZE_WINNING	RESOLUTE	PUNISHING	LORD
OPENHEARTED	PROCEDURAL	RESOURCEFUL	RAMPAGEOUS	LUSTROUS
ORDINARY	PRODUCTION	RISKY	RESILIENT	LUXURIOUS
ORIGINAL	PROFIT	ROUSING	RIGOROUS	MAGNANIMOUSNES
ORIGINALITY	PROFITABLE	SHAKE_UP	ROBUST	S
PERKY	PROMISING	SHARP	ROCKLIKE	MAGNIFICENT
PLAINSPOKEN	PROSPER	SMASHING	ROCKY	MELLIFLUOUS
PLEASANT	PROSPERING	SOLE	ROUGH	MELLISONANT
PLENTIFUL	PROSPEROUS	SOLITARY	ROUGHENED	NICE_LOOKING
POLITE	PROTECTED	SOLO	ROUGHISH	NOBILITY
POLITENESS	PROUD	SOVEREIGN	ROUGH_TEXTU	NOBLE
POSITIVE	PUNCTILOUS	SPECIFIC	RED	NOBLEMAN
PRACTICAL	RESPONSIBLE	SPINE_TINGLING	RUGGED	NOBLE_MINDED
PRAGMATIC	PROFIT_MAKING	SPIRITED	RUGGEDNESS	NOBLENESS
PROPER	SAFE	STATE_OF_THE_AR	RUTHLESS	NOBLESSE
PROPERNESS	SALABLE	T	RUTTED	NOBLEWOMAN
REAL	SALEABLE	STIMULATING	SAFARI	PATRICIAN
REALISTIC	SCIENTIFIC	STIRRING	SALOON	PHOTOGENIC
RELATION	SECURE	STOUT	SAVANNA	PICTURESQUE
RELATIONS	SELF_ASSURED	STRONG_WILLED	SAVANNAH	PLEASING
RELATIONSHIP	SELF_CONFIDENT	THRILLING	SCRAGGY	POLISHED
RELIABLE	SELF_POSSESSED	TONIC	SCRATCHY	POSH
REMARKABLE	SELLABLE	TRENDY	SERRATED	PRECIOUS

Sincerity	Competence	Excitement	Ruggedness	Sophistication
RESPECTABLE	SHELTERED	TURN_ON	SEVERE	PREEMINENCE
RESPONSIVE	SMART	UNCONSTRAINED	SPARTAN	PRESTIGIOUS
SCRUPULOUS	SOLID	UNDEVELOPED	STONY	PRETTY
SELF_EFFACING	STAUNCH	UNFETTERED	STRAINING	PRINCELY
SENTIMENTAL	STEADFAST	UNIQUE	STRAPPING	PROFLIGATE
SIMPLE	STEADY	UNMARKED	STRENUOUS	PROMINENT
SIMPLE_MINDED	STRONG	UNMATCHED	STRONG_ARM	QUEENLIKE
SINCERE	SUCCESSFUL	UP_TO_DATE	STRUGGLE	QUEENLY
SINGLE	SUPERIOR	UP_TO_THE_MINUTE	STURDY	REFINED
SMALL_TOWN	SYSTEMATIC	VALIANT	SUNRISE	RENOWNED
SMILING	TALENTED	VENTURESOME	SUNSET	RIGHTEOUS
SOCIABLE	TECHNICAL	VIBRANT	SURVIVOR	ROYAL
SPRIGHTLINESS	THOROUGH	VIBRATE	TESTING	SATIN
STANDARD	THOROUGHGOING	VITAL	THICK_SKINNED	SATINY
STRAIGHT	THRIVING	VOUGE	TIMBERLAND	SCINTILLANT
STRAIGHTFORWARD	TIRELESS	YOUNG	TOOTHED	SCINTILLATING
STRAIGHTNESS	TOPPLACE	YOUTHFUL	TOUCH_AND_GO	SEDUCTIVE
SYMPATHETIC	TRADE		TOUGH	SENSATIONAL
THE_RIGHT_WAY	TRANSNATIONAL		TOUGHENED	SERAPHIC
TRUE	TRIUMPHAL		TOUGHNESS	SHINING
TRUE_LIFE	TRIUMPHANT		TREACHEROUS	SHINY
TRUSTWORTHY	TRUSTY		TREK	SILK
TRUTHFUL	UNATTACKABLE		TREKKING	SILKY
TYPICAL	UNBEATABLE		TRICKY	SILVER_TONGUED
UNADULTERATED	UNBEATEN		UNCHARITABLE	SNOBBISH
UNASSUMING	UNBENDABLE		UNCIVILISED	SOPHISTICATE
UNCHANGING	UNDEVIATING		UNCIVILIZED	SOPHISTICATED
UNDERSTANDING	UNFAILING		UNCOMFORTABLE	SOPHISTICATION
UNGLAMOROUS	UNFALTERING		UNCOVERED	SPECTACULAR
UNGLAMOUROUS	UNFLAGGING		UNDOMESTICATED	SPLENDID
UNINTERESTED	UNFLUCTUATING		UNEVEN	STRIKING
UNOSTENTATIOUS	UNSHAKABLE		UNFORGIVING	STUNNING
UNPRETENTIOUS	UNSWERVING		UNKIND	STYLISH
UNQUESTIONABLE	UNWAVERING		UNMERCIFUL	SUAVE
UNSPOILT	UP_AND_COMING		UNPADDED	SUGARINESS
UP_FRONT	VENDABLE		UNPLEASANT	SUPERFINE
VALID	VENDIBLE		UNPOLISHED	TOP_NOTCH
WARM	VENTURE		UNREFINED	UPMARKET
WELCOMING	VICTORIOUS		UNRELENTING	UPPER_CLASS
WELL_FOUNDED	WILY		UNRESTRAINED	UPPERCLASSES
WELL_MANNERED	WINNING		UNSMOOTH	WELL_BRED
VERITABLE	WITTY		UNSTEADY	VELVET
VERITY			UNTAMED	VELVETY
WHOLESOME			WEATHER_BEATEN	VOGUISH
VIGOROUS			WEATHERED	VOLUPTUARY
VIRTUOUS			WEATHERWORN	VOLUPTUOUS
VIVACIOUS			VENTUROUS	WOMANLIKE
			WESTERLY	WOMANLY
			WESTERN	
			WILD	
			WILDLIFE	
			VIOLENT	
			WOODLAND	

Table A3. Association between communicated brand personality dimensions and brand position

Brand	**Sophistication**	**Excitement**	**Sincerity**	**Competence**	**Ruggedness**	**R^2**	**Adjusted R^2**	**F**
AirTel	-0.0359, p = 0.7675	0.0328, p = 0.1910	**0.09925, p < 2e-16**	**0.0412, p = 0.0051**	**0.0775, p = 0.0098**	0.0342	0.0327	**F(5, 3242) = 23.01, p < 2.2e-16**
BSNL	**4.3732, p = 0.0156**	0.0476, p = 0.1794	**-0.1816, p < 2e-16**	**-0.3298, p <2e-16**	**-0.1318, p = 0.0146**	0.1497	0.1484	**F(5, 3244) = 114.3, p < 2.2e-16**
Jio	**0.2630, p < 2e-16**	**0.1354, p = 1.23e-12**	**0.1240, p = 1.07e-06**	**0.1683, p < 2e-16**	0.0448, p = 0.174	0.0739	0.0725	**F(5, 3224) = 51.48, p < 2.2e-16**
Swiggy	**0.3520, p = 3.18e-07**	**0.1749, p = 4.76e-09**	**0.1689, p < 2e-16**	**0.1890, p < 2e-16**	**-0.0926, p = 0.00625**	0.0768	0.0753	**F(5, 3242) = 53.94, p < 2.2e-16**
Zomato	**0.1870, p = 0.0048**	**0.1028, p = 3.1e-05**	0.0188, p = 0.3818	0.0678, p = 0.0734	**-0.1184, p = 0.0291**	0.0111	0.0095	**F(5, 3088) = 6.954, p = 1.816e-06**
Axis Bank	**0.3930, p < 2e-16**	**0.1635, p = 1.95e-10**	**-0.2218, p < 2e-16**	**0.1678, p < 2e-16**	0.0632, p = 0.0854	0.1351	0.1337	**F(5, 3180) = 99.33, p < 2.2e-16**
Canara Bank	**0.2848, p = 0.0012**	**0.1138, p = 0.0315**	**0.4093, p < 2e-16**	**0.2778, p < 2e-16**	-0.0273, p = 0.3413	0.3244	0.3233	**F(5, 3244) = 311.5, p < 2.2e-16**

Brand	Sophistication	Excitement	Sincerity	Competence	Ruggedness	R^2	Adjusted R^2	F
HDFC Bank	**0.1868, p = 0.0011**	**0.0759, p = 0.0012**	**0.2047, p < 2e-16**	**0.1251, p < 2e-16**	-0.0127, p = 0.6777	0.0633	0.0618	**F(5, 3220) = 43.55, p < 2.2e-16**
SBI	0.3511, p = 0.0860	**-0.3844, p < 2e-16**	**0.3279, p < 2e-16**	**0.0760, p = 0.0001**	**0.2009, p < 2e-16**	0.1438	0.1425	**F(5, 3244) = 109, p < 2.2e-16**
Infosys	**0.1520, p = 0.0002**	**0.0596, p = 3.18e-05**	**0.0449, p = 0.0225**	**0.0966, p = 5.41e-13**	0.0128, p = 0.5507	0.0269	0.0254	**F(5, 3244) = 17.94, p < 2.2e-16**
TCS	**0.1205, p = 0.0055**	**0.0832, p = 1.97e-08**	**0.1975, p < 2e-16**	**0.1362, p < 2e-16**	0.0523, p = 0.0563	0.0798	0.0784	**F(5, 3242) = 56.24, p < 2.2e-16**
TechMahindra	0.0336, p = 0.3062	**0.0468, p = 0.0018**	**0.1015, p = 1.95e-06**	**0.0992, p = 3.49e-14**	-0.0200, p = 0.3548	0.0290	0.0275	**F(5, 3221) = 19.26, p < 2.2e-16**
Wipro	**0.2209, p = 6.39e-07**	**0.0693, p = 2.42e-06**	**0.0762, p = 0.0007**	**0.1180, p < 2e-16**	0.0347, p = 0.1395	0.0435	0.0420	**F(5, 3162) = 28.81, p < 2.2e-16**

Table A4. Association between perceived brand personality dimensions and brand sentiment

Brand	Sophistication	Excitement	Sincerity	Competence	Ruggedness	R^2	Adjusted R^2	F
AirTel	-0.0628, p = 0.5954	0.0237, p = 0.4755	**0.2109, p = 2.55e-14**	**0.0795, p = 0.0013**	-0.0495, p = 0.3606	0.0219	0.0204	**F(5, 3283) = 14.76, p = 2.424e-14**
BSNL	0. 0582, p = 0.673	**0.1745, p = 4.19e-12**	0.0116, p = 0.651	-0.0224, p = 0.421	-0.0279, p = 0.589	0.0153	0.0138	**F(5, 3245) = 10.13, p = 1.211e-09**
Jio	**0.2932, p = 0.0010**	0.0349, p = 0.3053	**0.0509, p = 0.0409**	0.0012, p = 0.9619	-0.0481, p = 0.3782	0.0050	0.0035	**F(5, 3220) = 3.286, p = 0.0057**
Swiggy	0.0522, p = 0.5089	0.0440, p = 0.1775	0.0158, p = 0.5065	0.0323, p = 0.2777	0.0846, p = 0.1161	0.0020	0.0004	F(5, 3232) = 1.298, p = 0.2619
Zomato	0.0295, p = 0.629	-0.0297, p = 0.403	0.0098, p = 0.707	-0.0343, p = 0.168	-0.0756, p = 0.157	0.0015	-2.614e-05	F(5, 3197) = 0.9833, p = 0.4264
Axis Bank	0.0713, p = 0.3251	**0.0858, p = 0.0165**	0.0085, p = 0.7264	**0.0605, p = 0.0300**	-0.0068, p = 0.8775	0.0037	0.0022	**F(5, 3174) = 2.41, p = 0.0343**
Canara Bank	-0.0105, p = 0.7814	**0.1669, p = 7.49e-08**	**0.0992, p = 2.22e-05**	-0.0654, p = 0.0527	**-0.1625, p = 0.0036**	0.0180	0.0165	**F(5, 3311) = 12.16, p = 1.063e-11**
HDFC Bank	**0.3596, p = 0.0050**	0.0372, p = 0.2730	0.0429¸ p = 0.0797	0.0415, p = 0.1161	**-0.0992, p = 0.0126**	0.0066	0.0051	**F(5, 3261) = 4.362, p = 0.0005**

Brand	Sophistication	Excitement	Sincerity	Competence	Ruggedness	R^2	Adjusted R^2	F
SBI	0.2004, p = 0.0583	0.0138, p = 0.6671	**0.0723, p = 0.0004**	-0.0208, p = 0.4187	-0.0028, p = 0.9427	0.0054	0.0038	**F(5, 3228) = 3.53, p = 0.0034**
Infosys	**0.1837, p = 0.0054**	-0.0055, p = 0.8018	**0.1173, p = 8.82e-06**	**0.1438, p = 6.51e-15**	-0.0219, p = 0.6739	0.0279	0.0264	**F(5, 3137) = 18.04, p < 2.2e-16**
TCS	**0.4207, p = 5.06e-14**	**0.0705, p = 0.0035**	**0.1549, p = 4.94e-10**	**0.1168, p = 3.17e-08**	0.0095, p = 0.7921	0.0420	0.0405	**F(5, 3147) = 27.65, p < 2.2e-16**
TechMahindra	**0.3002, p = 3.65e-09**	**0.1574, p = 5.46e-12**	**0.1215, p = 4.32e-06**	**0.1698, p = 7.56e-15**	**0.0970, p = 0.0145**	0.0596	0.0581	**F(5, 3044) = 38.63, p < 2.2e-16**
Wipro	**0.2236, p = 1.62e-06**	**0.1025, p = 0.0015**	0.0405, p = 0.2497	**0.0528, p = 0.0244**	**0.1149, p = 0.0389**	0.0151	0.0135	**F(5, 3087) = 9.475, p = 5.575e-09**

clickworker

clickworker GmbH · Hatzper Straße 30
DE-45149 Essen

Univ of Hyderabad
Anand V

Gachibowli
IN 500046 Hyderabad

Invoice#: COI23-75153
Customer#: 117959

Essen, April 14, 2023

Invoice

We invoice you the following orders:

Pos	Order#	Description	Order date	Net Amount (EUR)
1	439519	Survey - Tagging text messages with labels - 10	Apr 14 23	1.540

Sum Total Net	1.54
+ 19% VAT	0.00
Gross Amount	1.54

Prepaid invoice.

Kind regards
clickworker

Figure A1. Sample invoice pertaining to the validation of text classification task on clickworker.com clickworker

clickworker GmbH · Hatzper Straße 30
DE-45149 Essen

Univ of Hyderabad
Anand V

Gachibowli
IN 500046 Hyderabad

Invoice#: COI23-74809
Customer#: 117959

Essen, April 03, 2023

Invoice

We invoice you the following orders:

Pos	Order#	Description	Order date	Net Amount (EUR)
1	438891	Survey - Tagging image with labels - VI	Apr 03 23	1.820

Sum Total Net	1.82
+ 19% VAT	0.00
Gross Amount	1.82

Prepaid invoice.

Kind regards
clickworker

Figure A2. Sample invoice pertaining to the validation of image classification task on clickworker.com

APPENDIX B: ORIGINALITY REPORT

BRAND MANAGEMENT ON SOCIAL MEDIA: A MACHINE LEARNING-BASED FRAMEWORK USING TEXT AND IMAGES

by Anand V

Submission date: 06-Dec-2023 11:23AM (UTC+0530)
Submission ID: 2249739048
File name: V_Anand.pdf (1.53M)
Word count: 24708
Character count: 132701

BRAND MANAGEMENT ON SOCIAL MEDIA: A MACHINE LEARNING-BASED FRAMEWORK USING TEXT AND IMAGES

ORIGINALITY REPORT

6%	5%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	v-des-dev-lnx1.nwu.ac.za Internet Source	1%
2	www.slideshare.net Internet Source	<1%
3	libweb.kpfu.ru Internet Source	<1%
4	bura.brunel.ac.uk Internet Source	<1%
5	Ton Duc Thang University Publication	<1%
6	Submitted to Sheffield Hallam University Student Paper	<1%
7	Tanya (Ya) Tang, Eric (Er) Fang, Feng Wang. "Is Neutral Really Neutral? The Effects of Neutral User-Generated Content on Product Sales", Journal of Marketing, 2014 Publication	<1%
8	Submitted to Anglo American University Student Paper	

		<1%
9	elib.uni-stuttgart.de Internet Source	<1%
10	gigapaper.ir Internet Source	<1%
11	repository.usmf.md Internet Source	<1%
12	Submitted to The Independent Institute of Education (IIE) Student Paper	<1%
13	Submitted to University of Brighton Student Paper	<1%
14	scholarworks.waldenu.edu Internet Source	<1%
15	Submitted to University of Malaya Student Paper	<1%
16	Xuan Tran, Camille Dauchez, Anna-Milena Szemik. "Hotel brand personality and brand quality", Journal of Vacation Marketing, 2013 Publication	<1%
17	uu.diva-portal.org Internet Source	<1%
18	m.moam.info Internet Source	<1%

19	www.diam.unige.it Internet Source		<1%
20	www.researchgate.net Internet Source		<1%
21	aaltodoc.aalto.fi Internet Source		<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		