



UNIVERSITÀ
DI PAVIA

**Dipartimento di Scienze Economiche e
Aziendali**

**Corso di Laurea Magistrale in
International Business and
Entrepreneurship**

**The impact of aspect-based
sentiment analysis of customer
reviews on product strategy**

Relatore:

Chiar.mo Prof. Veglio Valerio

Tesi di Laurea di

Bellati Simone

Matr. n. 549269

Anno Accademico 2025-2026

Abstract - Italian	5
Abstract - English	7
Introduction	9
1. Theoretical foundations: main Machine Learning and Natural Language Processing techniques.....	12
1.1 Main Machine Learning techniques and their functioning	18
1.1.1 Supervised learning techniques	21
1.1.2 Unsupervised learning techniques	37
1.2 Natural Language Processing	45
1.2.1 Latent Dirichlet Allocation.....	47
1.2.2 The Bag-of-Words model.....	51
1.2.3 Rule-based sentiment analysis.....	54
1.3 From Neural Networks to the Transformer architecture.	58
1.3.1 Neural Networks.....	58
1.3.2 The transformer architecture	60
1.3.3 The BERT architecture.....	63
1.3.4 The DeBERTa model	64
1.3.5 Transformer use-cases in Marketing	66
2. Research Methodology	70
2.1 Model and product choice	70

2.2 Data collection and processing	72
2.3 Aspect selection.....	77
2.4 Advantages and limits of ABSA	78
3. Empirical Findings	84
3.1 Structure and interpretation of the model output.....	84
3.2 Confidence scores and the reliability of the classifications	85
3.3 Aspect salience: what customers spontaneously discuss	87
3.4 Sentiment polarity and the salience-sentiment matrix	89
3.5 Implications for marketing strategy.....	93
Conclusions	97
References	103
Appendix	118

Abstract - Italian

Nelle ultime decadi, la crescita dell'economia digitale ha portato con sé un aumento esponenziale del numero di dati a disposizione delle aziende. Tra questi, le recensioni online dei prodotti spiccano come una fonte di informazione strategica di immediata rilevanza, in quanto offrono una valutazione precisa e spontanea da parte dei consumatori. Tuttavia, la natura testuale e non strutturata di tali recensioni ne rende l'analisi sistematica difficile e costosa su larga scala. Questa tesi studia un tipo di analisi chiamato Aspect-Based Sentiment Analysis (ABSA), che sfrutta l'architettura di reti neurali chiamata transformer per individuare in modo automatico, all'interno delle recensioni dei clienti, le caratteristiche di un prodotto di cui si parla. Dopodiché, il modello utilizzato per condurre l'analisi assegna a ciascuna caratteristica del prodotto un sentiment positivo, negativo o neutrale, in base al contenuto della recensione stessa.

La domanda di ricerca che guida il lavoro è dunque la seguente: l'aspect-based sentiment analysis delle recensioni dei clienti può migliorare la formulazione strategia di marketing?

Per rispondere a tale quesito, lo studio propone un'analisi empirica basata su un corpo di 426 recensioni del visore di realtà virtuale Meta Quest 3, raccolte da Amazon in sei mercati e definendo sette aspetti: batteria, comfort, risoluzione, prezzo, passthrough, controller e fascia (strap).

I risultati rivelano informazioni strategiche significative come la frequenza di menzione di ogni aspetto e il sentimento generale ad esso associato nelle recensioni. Nel campione di dati analizzato, gli attributi più discussi dai clienti, ovvero la batteria e la fascia, sono anche quelli valutati più negativamente, mentre

la risoluzione è tra i meno menzionati ma, quanto viene citato, è quasi sempre per parlarne in maniera positiva. L'analisi evidenzia inoltre una marcata dipendenza del sentiment relativo al prezzo dal mercato di riferimento, con una discrepanza insolita tra Italia e Germania, dove il prezzo viene visto sotto una luce rispettivamente positiva e negativa.

Lo studio dimostra infine che, pur entro i limiti propri di un campione ristretto di dati, il tipo di analisi esplorativa e predittiva condotto è in grado di generare informazioni di diverso grado di granularità e con il potenziale di influenzare decisioni strategiche di marketing e di prodotto, fornendo così una risposta affermativa alla domanda di ricerca e contribuendo a colmare una lacuna nella letteratura sull'applicazione dei transformer alle recensioni di prodotto.

Abstract - English

In recent decades, the growth of the digital economy has led to an exponential increase in the amount of data available to companies. Among this data, online product reviews stand out as a strategic source of information of immediate relevance, as they offer an accurate and spontaneous assessment by consumers. However, the textual and unstructured nature of these reviews makes their systematic analysis difficult and costly on a large scale.

This thesis examines a type of analysis called Aspect-Based Sentiment Analysis (ABSA), which uses a neural network architecture known as a transformer to automatically identify, within customer reviews, the features of a product being discussed. The model used to conduct the analysis then assigns each product feature a positive, negative, or neutral sentiment based on the content of the review itself.

The research question guiding this study is therefore as follows: Can aspect-based sentiment analysis of customer reviews improve the formulation of marketing strategy?

To answer this question, the study presents an empirical analysis based on a corpus of 426 reviews of the Meta Quest 3 virtual reality headset, collected from Amazon across six markets, and defines seven aspects: battery, comfort, resolution, price, passthrough, controllers and head strap.

The results reveal significant strategic insights, such as the frequency with which each aspect is mentioned and the overall sentiment associated with it in the reviews. In the data sample analyzed, the attributes most frequently discussed by

customers (namely, the battery and the head strap) are also the ones rated most negatively, while resolution is among the least mentioned. However, when it is mentioned, it is almost always in a positive light. The analysis also highlights a marked dependence of price sentiment, with an unusual discrepancy between Italy and Germany, where the price is viewed in a positive and negative light, respectively.

Finally, the study demonstrates that, although within the limitations of a small data sample, the type of exploratory and predictive analysis conducted is capable of generating information at various levels of granularity and has the potential to influence strategic marketing and product decisions, thereby providing an affirmative answer to the research question and helping to fill a gap in the literature on the application of transformers to product reviews.

Introduction

Over the last decades, digital technology evolved increasingly fast, putting roots into most people's lives. This had many socioeconomic consequences, among which we note the rapid expansion of digital commerce. This has transformed the way people shop on one side and the way companies access, interpret and leverage customer feedback on the other. It is important to note that in 2021, the European e-commerce market was estimated to be EUR 718 billion in size (Roszko-Wójtowicz et al., 2024), with an annual growth rate of 13%. Amazon.com is one such platform where millions of reviews for products have been posted, making it a very detailed source of information. Unlike other methods of collecting information from markets, such as surveys and focus groups, reviews offer an unfiltered, granular assessment based on actual use of products.

Despite being a resourceful asset to assess customer feedback, product reviews present an unstructured textual nature that makes scaling systematic reviews analysis extremely costly in terms of time and money.

However, along with e-commerce, digital products and services in general grew at the same time, increasing exponentially the amount of data generated each year. In 2019, for example, the International Data Corporation forecasted that the "Global Datasphere" would grow from 45 Zettabytes to 175 Zettabytes in 2025 (Rydnig et al., 2018).

As suggested by a study (Halevy et al., 2009), it is only thanks to these huge amounts of data that companies and researchers managed to train natural language processing (NLP) and machine learning algorithms effectively. In fact, recent advances in these two fields have enabled the automated interpretation of textual

data at scale. Among these techniques, sentiment analysis has emerged as a prominent tool for identifying customers' attitudes toward products and services. Still, traditional sentiment analysis operates primarily at the document or sentence level, therefore producing aggregated positive, neutral or negative classifications that often obscure the specific product features driving satisfaction or dissatisfaction (Mao et al., 2024). From a managerial perspective, such insights provide limited guidance for product strategy, where attributes-specific insights can greatly enhance the intrinsic value of data.

Aspect-Based Sentiment Analysis (ABSA) addresses this limitation by decomposing customer feedback into predefined or detected product aspects and assigning sentiment polarity to each of them (Mao et al., 2024). This approach allows firms to identify which specific features (e.g. battery life, price, comfort etc...) generate positive evaluations and vice versa.

This thesis investigates how transformer-driven, aspect-based sentiment analysis can enhance marketing strategy through the systematic examination of online customer reviews. The empirical setting focuses on reviews of the Meta Quest 3, a virtual reality headset positioned within the consumer electronics market. The selected aspects for analysis (battery, comfort, resolution, price, passthrough, controllers and strap) reflect both technical specifications and experiential attributes that are central to the product's value proposition.

The study is based on a dataset of 426 Amazon reviews, collected and cleaned to remove incomplete entries. To conduct the aspect-based sentiment analysis, the pre-trained transformer model yangheng/deberta-v3-base-absa-v1.1, built upon the DeBERTa-v3 architecture, is employed. Rather than fine-tuning a model or evaluating predictive performance against labeled data, this research adopts an

exploratory analytical approach. The objective is not to benchmark algorithmic accuracy but to assess the managerial relevance of the insights generated by ABSA outputs. In this context, the machine learning model basically acts as an analytical tool designed to extract structured signals from unstructured textual data.

The central research question guiding this work is:

“Can aspect-based sentiment analysis of customer reviews enhance marketing strategy?”

To address this question, this thesis unfolds in three main parts. The first part establishes the theoretical foundations by providing a basic understanding and examining the role of machine learning and natural language processing in marketing. The second goes over the methodology applied to carry out the empirical aspect-based sentiment analysis, with a focus on the logic underlying the choice of the DeBERTa model and on the data collection procedure. Finally, we’ll present the findings and the managerial insights that such a type of study implies from a strategic point of view. Particular attention will be given to translating quantitative sentiment outputs into actionable product strategy recommendations.

1. Theoretical foundations: main Machine Learning and Natural Language Processing techniques.

Before diving into the core of the analysis, it is utterly important to provide an overview of the different types of machine learning algorithms, their functioning, how they are applied to solve specific marketing problems and the specific limitations that these *modus operandi* carries along.

In order to better frame this specific subset of artificial intelligence, we'll start by giving a definition of machine learning itself.

During the course of time, there have been different definitions of machine learning. A comprehensive study (Kühl et al., 2020) clarifies the terminology around machine learning, defining it as the set of techniques used to solve a wide range of problems by making computational systems learn instead of entirely programming the solution step by step. Another study showcasing machine learning applications to communication systems defines it as an alternative approach to algorithmic design (Simeone, 2018), where the development of an algorithmic solution doesn't stem from the immediate acquisition of domain knowledge, but starts instead from a training dataset from which the model learns and outputs a selected hypothesis.

More formally, machine learning can be framed as the study of algorithms that improve their performance at a given task through experience (Mitchell, 1997). With this in mind, a computer program is said to be able to learn from experience E regarding a class of tasks T and performance measure P if its performance on tasks in T , measured by P , improves with experience E (Mitchell, 1997). From this definition, three different elements are evident: experience, a task to be performed

and a performance metric. In a similar context, Goodfellow et al. (2016) define machine learning as a field that enables computers to learn without explicit programming, and this definition emphasizes the role of statistical and optimization techniques in finding patterns in data.

If we flip the perspective on the methodological side, machine learning can also be observed within a much broader framework: that of statistical learning theory. In this view, learning consists in estimating a function that maps inputs to outputs while minimizing a predefined loss function and controlling model complexity to avoid overfitting (Bzdok, 2017).

Within this context, it is important to consider that a major factor contributing to the spread of machine learning techniques is the rapid growth in the amount of data generated in digital environments. The rise of what is commonly called “big data”, usually described in terms of high volume, velocity and variety, has significantly changed how organizations approach data analysis (Mazzieri and Soares, 2016).

At the beginning, traditional statistical methods used to work with smaller datasets and limited computing power, whereas, modern machine learning algorithms are designed to handle large datasets and tend to improve their performance as the quantity of training data increases (Halevy et al., 2009).

Research has shown that, in many cases, having more data can lead to greater performance improvements than making small adjustments to the model itself (Halevy et al., 2009). At the same time, the development in the capabilities of computers, as well as the development in the capabilities of cloud storage and data processing technologies, have enabled companies to use these techniques on a large scale. Therefore, in the last ten years, machine learning has moved from being

a merely academy-oriented subject to being a significant aspect in the world of data-driven decision-making in various industries (McAfee & Brynjolfsson, 2012).

Moving on, when talking about machine learning it is useful to differentiate between supervised, unsupervised and reinforcement learning.

According to Simeone (2018), this taxonomy reflects the type of feedback signal available to the learning system and the nature of the task it aims to solve.

Supervised learning refers to the class of algorithms trained on labeled datasets, where each input observation is associated with a corresponding output variable (Hastie et al., 2009). The objective of these algorithms is to learn a mapping function that can generalize unseen data. Typical supervised learning tasks include classification and regression, where the output variable is respectively categorical or continuous (James et al., 2013). Algorithms such as logistic regression, support vector machines and neural networks fall within this category. The key strength of supervised learning lies in its predictive accuracy when sufficient labeled data are available; however, its performance strongly depends on the quality and representativeness of the training data (Goodfellow et al., 2016).

On the other hand, unsupervised learning works with unlabeled data and aims to find hidden patterns without any output variables (Hastie et al., 2009). Hence, learning is more exploratory than predictive. The most common examples of unsupervised learning include clustering, dimensionality reduction and topic modeling (James et al., 2013). We will later analyze k-means clustering, hierarchical clustering and principal component analysis. As there is no output available for these techniques, evaluating their performance is naturally more difficult (Hastie et al., 2009).

The third type of machine learning is reinforcement learning. This type of learning involves an agent learning how to make sequential decisions by interacting with an environment and obtaining feedback in the form of rewards and penalties (Sutton and Barto, 1998). Unlike supervised learning, there is no output available for an agent. The agent has to learn to find out what actions lead to maximum reward over time. Reinforcement learning typically involves a decision process where an agent perceives a state and takes an action and changes its state probabilistically (Sutton and Barto, 1998). This type of learning is very suitable for fields of application where decisions need to be made based on continuous adaptation to changing environments.

In conclusion, supervised learning is used for prediction with labeled data whereas unsupervised learning addresses structure discovery with unlabeled data and reinforcement learning points at sequential decision-making tasks with an agent and an environment. This classification of machine learning techniques helps understanding the conceptual distinctions between different types of machine learning and, at the same time, offers a framework for understanding how these techniques can be used to solve different types of marketing and product related problems and what each of these techniques entails.

Now that we have discussed a general classification of these different types of machine learning techniques, let us finally position our project within this framework.

It is important to remember that the empirical analysis conducted in this research is based on a pre-trained transformer model, whose architecture will be briefly explained later in the analysis. Plus, the objective of the model was to assign a sentiment polarity (positive, neutral or negative) to predefined product aspects

such as battery, comfort, resolution or price.

From a methodological standpoint, the learning mechanism that characterizes the adopted model relates to supervised learning. As previously discussed, supervised learning relies on labeled datasets, in which each input is associated with a corresponding output variable (Hastie et al., 2009; James et al., 2013). In our case, the model has been trained on textual data where each aspect and its corresponding sentiment polarity were explicitly labeled.

Although the present thesis does not involve fine-tuning the model on a custom labeled dataset, it still relies on a supervised learning system whose parameters have been estimated through exposure to annotated data. Therefore, we can say that the predictive mechanism at the core of the whole analysis aligns with the supervised learning approach. The model receives textual input (customer reviews) and outputs discrete sentiment classifications for each predefined aspect, thus performing a multi-class classification task.

More on the rationale behind the choice of this specific model will be explained in another section of this study.

On the other hand, we can argue that the project does not fall within the unsupervised learning taxonomy, since it does not aim to discover unobserved structures without predefined labels or to cluster the different reviews. Similarly, it does not belong to reinforcement learning, as there is no sequential interaction between an agent and an environment, nor any reward-based optimization process (Sutton and Barto, 1998).

It is crucial to understand the differences among these three paradigms of machine learning since they provide unique opportunities for gaining competitive benefits

within practical marketing problems, which is also the reason why this chapter provides substantial and extensive theoretical background to understand the evolution of such techniques.

For example, supervised learning may help large e-commerce platforms like Amazon in creating purchase propensity models that predict which customers will be more willing to buy products in a certain product category using historical transaction data. As a result, a company will know where to allocate resources on advertising, optimizing marketing expenses. This technique is advantageous since instead of targeting all consumers indiscriminately, the firm can invest money in people who have a higher chance of buying advertised items, thus achieving better returns from its campaigns: with the number of customers and data modern companies have nowadays, this cannot possibly be achieved without machine learning.

Similarly, a consumer electronics firm like Samsung can use supervised classification algorithms to train models that predict whether some product lines will receive complaints from dissatisfied clients after purchasing it or reading its product reviews. Thus, a firm will be able to make proactive interventions aimed at improving its offerings and avoiding bad publicity.

At the same time, with the help of clustering algorithms, companies like Unilever can perform customer segmentation by behavior using unsupervised learning techniques. It means that the organization would no longer depend on assumptions made by marketers about the segments' preferences but actually see the patterns behind consumer behavior. Instead of guessing how to position their product or modify its packaging or messages used during the marketing campaign, they will be able to do this accurately.

Furthermore, topic modeling that involves unsupervised learning enables firms like Booking.com to perform a similar operation, discovering what themes guests mention in their reviews and then creating products according to their expectations.

Finally, reinforcement learning allows organizations to design self-improving marketing strategies. For instance, Google Ads can apply reinforcement learning models to bid on keywords and change bidding rates based on the outcomes of auction rounds.

These and other, more detailed examples that will be explored in the subsequent sections of this literature review clearly show that nowadays, machine learning enables the development of winning marketing strategies as they provide management with competitive, additional knowledge, allowing them to better allocate resources, take more informed decisions and, ultimately, improve marketing strategy.

1.1 Main Machine Learning techniques and their functioning

Now that we have introduced machine learning itself, along with the different types of approaches related to it, we can provide an overview of the different types of machine learning algorithms and how they are applied to solve specific marketing problems.

The progressive permeation of machine learning into marketing activities is likely due to the fact that the ability to process large volumes of behavioral and textual data translates directly into competitive advantage in this field. As previously mentioned and as argued by Huang and Rust (2021), machine learning is repositioning marketing from a discipline grounded in human intuition and small-scale research to one increasingly driven by automated pattern recognition, a shift

with profound implications for how firms understand and interact with their customers.

At a high level, machine learning applications in marketing can be grouped around a few recurring problems. Prediction tasks such as customer churn forecasting, lead scoring and demand estimation have benefited greatly from supervised learning approaches, where models are trained on historical labeled data to anticipate future outcomes. Geiler et al. (2022), for instance, provide a comprehensive survey of machine learning methods applied to churn prediction, showing how classification algorithms trained on transactional data can effectively identify customers at risk of churn. Dynamic pricing represents another important use case: Yoshi et al. (2025) show how regression-based machine learning models can help e-commerce platforms adjust prices in real time based on demand signals and competitor behavior, performing better than traditional rule-based pricing approaches.

Other typical marketing problems can be better addressed through unsupervised approaches. Customer segmentation is perhaps the most natural application of such methodology: rather than relying on predefined demographic categories, clustering machine learning algorithms can identify groups within a customer base based on actual behavioral patterns. Hicham and Karim (2022) demonstrate how clustering methods applied to customer data can produce detailed and useful segmentation results that go beyond what conventional approaches would offer. Recommendation systems, which are a key part of modern e-commerce personalization strategies, represent another area where unsupervised and hybrid methods have proven very effective. The item-to-item collaborative filtering approach introduced by Linden, Smith and York (2003) at Amazon is a foundational reference in this space, showing how large-scale interaction data can be used to generate relevant product suggestions without requiring explicit user

input.

In the recent years, the rise of deep learning and transformer-based architectures has expanded what machine learning can do with unstructured data, particularly text. These models have greatly improved natural language processing capabilities, making it possible to carry out tasks such as sentiment analysis, topic modeling and text classification at a scale and level of accuracy that was not reachable before (García-Díaz et al., 2024). In marketing, this has made it possible to systematically analyze consumer-generated content such as reviews, social media posts and customer service transcripts, and turn it into structured strategic insight. As noted by Hartmann and Netzer (2023), transformer-based models have not only improved performance on established NLP tasks, but have also enabled entirely new applications with direct relevance to marketing practice.

It is important to consider, however, that across all these approaches, a persistent tension exists between predictive performance and interpretability. As Luo et al. (2021) observe, even if machine learning methods can process large-scale, unstructured data with impressive accuracy, their complexity often limits their use in managerial contexts where decision-makers need to understand and justify the outputs they act upon. This limitation is not only technical but also carries real strategic consequences, and it will come up again as a relevant consideration at several points throughout this study.

With this general overview in place, we can now turn to a more structured examination of the main machine learning techniques and discuss how each of them address specific marketing applications and methodological constraints, starting from supervised learning techniques.

1.1.1 Supervised learning techniques

Linear and Logistic Regression

Linear and logistic regression used to be statistical tools, but with the evolution of machine learning, they became the baseline of this set of techniques thanks to their interpretability compared to other, more modern machine learning approaches. One of the first moments in which they are treated as part of machine learning rather than statistical learning models is witnessed in a study by Tibshirani (1996), where the introduction of the LASSO technique (L1 type of regularization) effectively managed to make the models automatically ignore the least relevant features.

One of the fundamental differences between these two models lies in their objective function. While linear models aim to predict continuous dependent variables, logistic regression aim is to predict the correct class of a categorical target variable. Another piece of literature (Nelder et al., 1972) better emphasizes how logistic regression relates to a binomial distribution, whereas linear regression is suited for normal distribution. However, the study unifies them under the generalized linear models category due to their relevant common aspect: they can both fit a maximum likelihood estimation framework.

As mentioned at the beginning of this section, these models are widely considered as the baseline when comparing them to other machine learning techniques. In fact, they are often referred to as “white box models”, as opposed to other “black box models” that we are going to see in the subsequent chapters. In a marketing application, this would mean that every prediction could be explained by a specific marketing lever. Salih and Wang (2024) affirm that despite these linear regression

models being widely considered fully interpretable, their fundamental assumption of linearity and their limitations such as multicollinearity and covariate effects interfere with the reliability of these models' results, proposing that such models should be treated as if they were less interpretable ones, such as neural networks, especially in high-stakes domains like healthcare. Dervovic et al. (2024) states a similar claim but for logistic regression models, highlighting the difficulties in interpreting the coefficients of the results and proposing to fix this through the use of Linearised Addictive Models, allowing direct interpretation of coefficients as probability changes.

Given the already explained characteristics of these models, we can find a wide range of marketing applications related to them. A typical marketing problem is tied to the use of advertising. Qader et al. (2022), for example, assessed advertisement's effectiveness across 5 different channels using multiple linear regression, resulting in online advertisement being the most effective. Another piece of literature (Jemal and Kassegn, 2025), explores in the same way the impact of advertisement on buying behavior.

Sometimes, these models are used in an even broader framework: Ali et al. (2024) used linear regression to evaluate the relationship between Customer Lifetime Value (CLV) and financial performances.

Effendi and Ali (2017)'s research investigates another marketing problem, typical of the digital age: predicting Click-Through-Rate for contextual advertisement. This study emphasizes the relevance of feature engineering (i.e. choosing the most appropriate predictors) to enhance model performance and concludes that linear regression represents a validated model to solve this specific marketing problem.

Linear regression also shines when combined with other techniques. Dong et al. (2020)'s literary contribution showcases how the analysis of customer reviews, star ratings and helpfulness votes of Amazon reviews help unveiling behavioral trends. In this context, linear regression helped assess the type and intensity of the relationship between star ratings and sentiment scores, along with their statistical significance.

Moving on to logistic regression, Constantin (2015) analyzes binary survey data to discover how this model can enhance and support marketing strategies. The data analyzed in this case study regards student housing and, through data analysis, the author manages to predict the likelihood of students leaving a residence hall in case they had the opportunity to do so, discovering, however, that the only significant variable in predicting this type of outcome is the quality of the accommodation facilities.

A greatly explored marketing problem is churn prediction. Chajia et al. (2024), proposes an innovative solution to that, combining logistic regression with an LLM, manages to identify customers that are more likely to leave in order to implement retention strategies such as discounts and special offers in the telecommunications industry.

Nie et al. (2011), use a more traditional approach to churn prediction and forecasting, comparing it at the same time with decision trees, another classification model that we will explore later in this section.

Li (2022)'s implementation of logistic regression compares it to other machine learning techniques, focusing on consumer engagement through the identification of customers that are most likely to positively respond to digital marketing

campaigns, by measuring user interactions such as purchases or clicks. The performance, even when compared with other methods, is moderately good.

Similarly, Chen (2022), uses logistic regression to predict customers that would purchase on an e-commerce website, aiming at an improvement in the so-called “database marketing”. In this case, the model was outperformed by tree-based models that better capture non-linear behavioral patterns in the data.

We can conclude from this overview about linear models that their key strength lies in their simplicity and interpretability, making them a useful and empowering tool for practitioners. However, they suffer from some strong limitations. Despite various advancements have been made to enhance these models' performances, such as feature selection algorithms, they will always assume linearity in the data, which makes the models fail whenever there is some slightly more intricate pattern within the dataset in analysis.

This is why researchers have started applying new, more complex techniques to solve both the same and new, unexplored marketing problems.

Tree-based methods.

Tree-based models are the second category of supervised learning techniques that we are going to analyse in this machine learning's state of the art overview.

The tree-based category contains several techniques that all share a common base model, which is the decision tree, used for both classification and regression tasks.

So, decision tree models can predict a target class label based on a series of hierarchical questions (Kingsford and Salzberg, 2008). In this type of model, the data starts from a root node and flows through each node, representing a question

that will influence the outcome predicted in each final node which is called leaf. At the node-level, each split threshold is decided based on impurity measures, such as the Gini index or entropy.

According to Blockeel et al. (2023), decision trees shine when it comes to interpretability: every predicted outcome can easily be traced back to the root node by following backwards the rule-based classification.

One of their biggest limits, on the other hand, is definitely their natural tendency to overfitting: as datasets increase in complexity (for example, with a high number of features compared to the total observations), decision trees' flexibility allow them to capture hidden patterns in the data, even irrelevant ones (Ying 2015). This eventually translates into extremely high performances in the training phase and poor evaluation metrics as the model tries to generalize its predictions to unseen data. However, this can be prevented in different ways. Amro et al. (2021) suggests, as an overfitting mitigation strategy, a combination of instance reduction and pruning, which basically consists in reducing the dataset size and the tree complexity by cutting off the branches with the least predictive power.

As mentioned in the first paragraph of this subsection, decision trees are just the simplest methods entangling tree-based models. All these other techniques fall into the ensemble category.

Opitz and Maclin (1999) describe them as a mix of multiple decisions trained together, combining their predictions to provide a final outcome. The different outputs of each tree is aggregated through a combining function that can leverage different methodologies, such as majority voting or prediction averaging. The key idea underlying ensemble methods, they explain, is that several models together

perform better than a single one alone.

One of the ensemble methods mentioned in the paper is bagging, also deepened by Breiman (1996) in another work. It is a way to reduce the algorithm instability, generating multiple versions of a predictor through randomized sampling with replacement. There are two points here: the first is that this randomized sampling with replacement defines another statistical expedient called bootstrapping: bagging, in fact, is the short version for bootstrapping aggregating. The second point relates to the fact that by “reducing the algorithm instability”, it is intended that the core idea of bagging is decreasing variance itself: by averaging several models, bagging stabilizes the model results mitigating overfitting at the same time.

Bagging is only one of the methods that belongs to the wider category of ensemble learning techniques that seek to enhance predictive accuracy through the combination of several models. Random Forest models are related to bagging, since they also build several trees through bootstrapped resampling. However, the key difference is that the Random forest technique applies random feature selection at each split of each tree (Yanni, 2025). It means that while bagging-based trees evaluate their predictions based on all the available features, trees built using the Random Forest technique will do it using a random subset of features, ultimately reducing correlation among trees and improving accuracy (Breiman, 2001).

The third main technique used to improve decision trees' accuracy and limit overfitting is boosting. Boosting refers to a class of algorithms that enhance the performance of a base learner (in this case, decision trees) by combining multiple models in a sequential manner. In this approach, each new model is trained with a greater focus on observations that were poorly predicted by previous models,

thereby progressively improving overall performance (Freund et al., 1999).

Friedman (2001) proposes a framework where a specific type of boosting, gradient boosting, is implemented in the context of decision trees. What happens when you do such thing is that many trees are built sequentially, each focusing on the previous tree's mistakes, making the final results more reliable.

Now that we provided a clear outlook on the main tree-based machine learning methodologies, let's investigate how these tools are implemented to solve specific marketing problems.

A first example of decision tree-driven problem solving in marketing can be found in Tirelli et al. (2007)'s work. The paper analyzes more than 450000 customers' data from a major European airline, with two main goals: performing customer segmentation over three different value tiers based on annual revenue generated and predicting customer's future value. In the end, previous spending turned out to be the most powerful predictor for future value, while regarding customer tiers it was found that the best target customers, or the ones with the highest value, were corporate customers with seats preferences.

Another study (Liu and Yang, 2022) compares different algorithms in the context of content marketing. The problem in this analysis concerns the prediction of customer responses to content marketing initiatives and leverage high volume datasets with many dimensions available. The target variable in question is a binary variable stating whether or not the customer accepted the marketing offer. Surprisingly, a simple decision tree here outperformed other models, including ensemble methods such as random forest, achieving an accuracy of 91%. It is important to note that this was possible only through a proper fine tuning process,

with special attention posed to the depth of the tree, but it still shows how even a simple tool like the decision tree can enhance managerial decision making. For example, by targeting mainly highly profitable customers in content marketing campaigns.

The article just observed shows that any of these tools can find an appropriate use-case under the right conditions and there is no such thing as a perfect technique. The next paper we examine (Hayadi and Ibrahiem, 2024), in fact, shows an opposite example as the previous one.

In this research, in fact, Hayadi and Ibrahiem offer a comparative evaluation of decision trees and random forest models for forecasting the return on investment (ROI) of digital marketing campaigns. The results show that random forest clearly surpasses individual decision trees on various assessment metrics, such as mean squared error, root mean squared error and mean absolute error. This performance gap with respect to decision trees is due to the ensemble characteristic of random forest, which, as previously explained, enables it to effectively capture nonlinear relationships and interactions between variables, while also reducing overfitting. On the other hand, decision trees displayed lower predictive power and accuracy, as indicated by their less favorable performance metrics. The research also highlights key ROI drivers, with particular regards to metrics such as the conversion rate and the customer acquisition cost, providing high level guidance on managerial decision making.

Going deeper in this overview, we bring up Gültür and Muhammed (2025)'s work. Similarly to what we saw earlier, they aim at predicting a marketing campaign's success with different machine learning approaches. In order to do so, they leverage a publicly available dataset containing approximately 10000 marketing

campaigns, with various features such as ad spend, duration, audience reach, budget allocation and engagement metrics (e.g. click-to-views ratio). In this case, gradient boosting was the top performer model, with an accuracy of 93.3%, followed closely by random forest's 92.9% of accuracy. The single decision tree, here, achieved a modest level of predictive power, with an accuracy around the 82% threshold. The managerial finding here is that in most of the marketing initiatives analyzed, engagement and conversion metrics were the strongest success indicators, while budget allocation and ad spend proved themselves to be secondary in terms of variable importance. So, one can assume that structuring compelling and appealing marketing campaigns should be prioritized over injecting more money into ad spend or reorganizing the budget allocation strategy.

The latter is surely not the only study that compares different ensemble methods. Singh (2025) uses the tree-based techniques seen so far to forecast airline ticket prices by utilizing a vast dataset of over 300,000 observations collected over a period of 50 days. Methodologically speaking, the study involves a comparison of the results of a single regression tree, built using a splitting procedure that minimizes mean squared error (MSE), with an ensemble method that uses random forests of decision trees built using bootstrapping and averaging. The results of the empirical study show that the single decision tree yields a root mean squared error (RMSE) of 70.27, whereas the results of the random forest show a lower test RMSE of 64.24. From a marketing perspective, the paper's implications are significant, as they prove the efficacy of ensemble methods over single decision tree methods in forecasting airline ticket prices, which will be beneficial to firms seeking to use such methods to create more effective dynamic pricing strategies.

Support Vector Machines

Other than Tree-based methods, support vector machines (SVMs) are a widely used machine learning technique when it comes to solving marketing-specific problems.

Just like decision trees and their extensions, they are used in both classification and regression tasks and often considered reliable in terms of performance (Saravanan, 2023). Plus, they are generally effective in handling noisy datasets, even in high-dimensional data contexts.

Saravanan (2023) also explains their underlying mathematical foundation clearly: support vector machines aim at finding the optimal hyperplane between different groups of data. It means that these types of models aim to find the dividing line's function that maximizes a margin. This margin represents the distance between the hyperplane itself and the closest data point for each group. This characteristic helps the model capture non-linear patterns in the data. Additionally, different mathematical methods such as kernel functions and Lagrange multipliers are part of the technique, respectively to transform the data into a higher dimensional space and to efficiently solve the optimal boundary problem.

If decision trees could be improved with ensemble methods, we have to recognize that support vector machines also have their own variants, each with their own specifics.

Following this, it can be noted from the literature that there are now many different versions of support vector machines, each attempting to solve different issues inherent in the traditional model, all of which maintain their core idea of maximising margin.

A very important contribution to the literature is represented by a review of SVMs applications in healthcare (Guido et al., 2024)

According to the study, the first set of extensions aims at making it easier for practitioners to control and understand the complexity of their model. ν -SVM does this by reformulating the traditional optimization problem. Instead of using a regularisation parameter, ν -SVM uses a new parameter, more easily controlled because it can be set based on the proportion of support vectors and the proportion of errors in the decision boundary. This is particularly significant because, unlike traditional models, users can now easily control the number of data points that are taken into consideration for decision making.

The second SVMs variant deals with making it easier for users of SVMs to extend their model to more complex classification problems. In fact, when it comes to more complex classification problems, it is often required for users of SVMs to extend their model. This is done using fuzzy SVMs, which attempts to solve this problem. The idea behind this is that not all data points are equally important for decision making (Tanveer et al., 2025), which is especially true when working with class-imbalanced datasets. Fuzzy SVMs assign fuzzy membership values to samples, giving less weight to outliers and noisy instances in the data.

Moving on with the third extension of SVMs mentioned by Guido et. al (2024) relates to one of the main limitations of classical SVM: their computational cost, particularly for large datasets. Least squares SVM attempts to overcome this limit by reformulating the optimization problem to make the computation simpler. The resulting problem is to solve a system of linear equations rather than the traditional quadratic optimization problem, which reduces the computational burden of the algorithm significantly. The downside is that this also drastically reduces the

sparsity of the solution, as the resulting model typically depends on all training data points rather than a subset of support vectors.

One of the largest extensions of the basic SVM idea is the so-called Twin Support Vector Machine. In this case, instead of one optimal decision hyperplane, two non-parallel decision hyperplanes are constructed, one for each class of the binary target variable (even if in some versions it can be used for multi-class prediction and regression). New instances are classified according to the closest decision hyperplane. This approach decreases the computational cost by effectively solving two smaller optimisation problems instead of one large one (Tanveer et al., 2024). This approach also enhances the ability to handle flexible decision functions for the two classes, which can improve performance for unbalanced class distributions. Various extensions exist for this approach, including the least squares formulation for further reducing the computational cost and the addition of further regularisation parameters.

Additionally, cost-sensitive SVMs have been developed to have different penalty parameters for different classes, ultimately ensuring that instances of the minority class are given a higher penalty than instances of the majority class and contributing to the overall problem of imbalanced classes predictions (Guido et al., 2024).

Uncertainty and noise have also led to the development of probabilistic and fuzzy versions of these models. Models such as intuitionistic fuzzy twin SVMs have improved the weighting of data points by taking into consideration membership and non-membership functions (Rezvani et al., 2019). Other models have also been developed to incorporate probabilistic reasoning with the traditional SVMs, allowing for more interpretable outputs in relation to class likelihoods.

More recent developments involve the incorporation of multiple sources of information within the learning process. For example, Universum-based SVMs make use of auxiliary data that is not a member of any particular class but that nonetheless gives some insight into the overall distribution of the data. This allows for even greater generalisation, particularly in environments where there is a lack of data (Dhar and Cherkassky, 2019).

Finally, there are some SVM variants that make use of feature selection and even hybrid modelling techniques. One such technique is recursive feature elimination, which again uses the SVM framework but this time in order to eliminate certain less useful features. This is particularly useful in datasets that are very high-dimensional (Sanz et al., 2018).

Overall, all of the above variants serve to highlight that SVMs are not a static model, but rather a whole range of models that all make use of a common framework. Each of the different models has addressed a particular limitation of the original formulation, whether this is with regards to computation, robustness, imbalanced data, or adaptability.

Let's now see how SVMs in general are implemented by marketing practitioners to solve different types of product and marketing problems.

A recent application of support vector machines in the field of marketing can be found in the study by Zhao & Hu (2025), in which the authors develop a Particle Swarm Optimization-based (PSO) SVM model for consumer segmentation and optimization of marketing strategies. Specifically, the study attempts to address the limitation associated with the high dependence on hyperparameters through the development of a model based on the integration of the particle swarm

optimization technique, which is a computational method inspired by nature. From a methodological point of view, the study combines the nonlinear modeling potential of SVM models, in which the radial basis function kernel was adopted, with the evolutionary search potential of the PSO algorithm, aiming to enhance the search for an optimal decision boundary in a high-dimensional dataset. The results of the experimental analysis, based on behavioral and demographic characteristics of 2,000 consumers, showed that the proposed model significantly outperformed standard SVM and decision trees, along with another SVM variant. Specifically, the results showed that the model achieved a more than decent accuracy rate of 89.4% and an F1 score of 0.89, indicating better performance compared to the benchmarks, preserving at the same time a relatively low computational cost. The study also explains that, from a managerial point of view, the model can be used to segment consumers based on different characteristics, identifying high-value consumers, price-conscious users or potential growth users, thus enabling the development of effective marketing strategies.

Another significant contribution in the use of support vector machines in marketing research can be found in Xiahou et al. (2022)'s application in customer churn prediction in a B2C electronic commerce environment. The paper's main goal is to improve the accuracy of customer churn prediction by incorporating customer segmentation into machine learning algorithms. In this context, the paper's methodology is based on the following steps: customer segmentation via k-means clustering, the use of random forest for selecting the most significant features and the use of both SVM and linear regression for customer churn prediction. The SVM algorithm, which, as already mentioned, is able to process high-dimensional nonlinear customer data, is used as a binary classifier to differentiate between churners and non-churners. The results show that the proposed approach improves

the predictive performance in all the evaluation criteria. Moreover, the proposed SVM approach performs better than the linear regression algorithm in all the evaluation criteria, including the accuracy, recall, precision. For instance, the proposed SVM approach has a higher average accuracy, which equals 0.916 compared to 0.907 for the linear regression model. Nonetheless, it is a good example showing how all of the tools and techniques we are examining can work all together to provide an overall insight into the marketing problem that is as accurate as possible. Regarding the paper's implications for marketing practices, the results clearly show that the proposed approach, which combines customer segmentation with SVM-based prediction, helps marketers better identify customer groups at risk of churning. The latter improves customer relationship management as well as operational costs.

Yuan (2025) builds on the concept of retail precision marketing by developing a radial basis function kernel SVM algorithm optimized by fuzzy clustering to improve the quality of transactional data. By performing feature engineering in the form of a weighted Recency, Frequency, Monetary (RFM) analysis, the research was able to segment the customer base into four groups, thereby helping to identify high-value customers. The algorithm was able to achieve an accuracy rate of 87.4%, which was significantly higher than the conventional decision trees and logistic regression algorithms. In the context of business strategy, the application of the research was able to achieve a 12.7% increase in conversion rates and a 18.3% increase in marketing ROI. However, the research states that the effectiveness of the algorithm decays after four months, thereby requiring the algorithm to be re-trained to sustain the competitive edge.

Rogić et al. (2022) also tried to address the problem of class imbalance that is usually experienced in digital marketing, by analyzing an online sports store

dataset that had an extremely low response rate of only 0.41%. This is where the number of actual customers is extremely low when compared to the number of people visiting a given website. In this case, the authors came up with a Balanced SVM (B-SVM) that is used as a pre-processor. This is where random undersampling is combined with SVM to remove noise in the dataset, to ensure that there is a balance between customers and non-customers. This approach, in the end, improved the system's ability to identify actual customers. One of the key takeaways from this paper is that the pre-processing technique enabled the model to find 7.3 times more potential customers than usual, improving the sensitivity score from 10% to over 83%. The framework in question can help businesses target their ads more precisely so that they do not waste their budget on customers that are not likely to respond.

To conclude our SVMs overview, we analyze the study by Zhou et al. (2019), which presents an intriguing example of SVM usage in the telecommunication industry, filling in a gap in the subject matter that most other relevant studies on customer churn prediction tend to ignore. While most of the research in the field has mostly focused on predicting the likelihood of customer churn, Zhou's paper attempts to turn its attention towards the most loyal customer segment, asking whether they might be interested in a new product (leveraging their appetency) or in upselling the service the customers are currently using. The study uses a publicly available UCI telecom dataset containing 3,333 customer records to develop a prediction framework that combines these three different prediction types through a clever intermediate calculation based on support vector machines result, where the output is first converted to a posterior probability estimate for the non-churn class and then normalized to a customer satisfaction score. Customers with satisfaction scores above certain thresholds are then marked as potential upselling

or new product targets.

Technically speaking, two different kernel functions are compared in the study, with the polynomial kernel consistently performing better than the linear kernel, attaining a maximum accuracy in predicting customer churn at 91.67% while also recording significantly lower minimum prediction errors in all possible train-test splits. From a marketing perspective, the framework would provide telecommunication experts with more relevant signals than just a binary churn prediction result, potentially allowing for more effective customer relationship management strategies. When viewed in conjunction with the other SVM usages presented in the other studies in this section, the study by Zhou et al. (2019) shows how a more creative approach to model building can increase the potential usage of support vector machines beyond just binary classification.

1.1.2 Unsupervised learning techniques

K-Means and Hierarchical Clustering

In this sub-section, we will review the functioning and marketing relevance of our first unsupervised learning methodologies: k-means clustering and hierarchical clustering. While classification and regression algorithms that have been previously discussed require some labeled information to learn from, clustering algorithms do not have this restriction. In fact, they are designed to look for clusters in a given dataset where observations within a cluster are more similar one to each other, compared to how similar they are to observations in other clusters (Jain, 2010). This makes clustering algorithms particularly relevant to various marketing problems because identifying clusters within a given dataset is a problem that is frequently pursued for the sake of insights discovery rather than being used to make predictions about a specific target variable. Still, this kind of analysis can be

of great help, as Wedel and Kamakura (2000) note in their work on market segmentation. According to them, that of segmenting a market into homogeneous groups within a heterogeneous market can be a crucial capability to enhance marketing strategy because it helps marketing practitioners target specific clusters of customers based on shared characteristics.

The above mentioned clustering algorithms are two of the most popular techniques as they are frequently used in various machine learning problems. While both algorithms are designed to perform clustering on a given dataset, they do so in rather different ways.

K-means is an iterative algorithm that was originally formulated in a formal sense by MacQueen (1967). This algorithm is designed to partition a given dataset into a pre-specified number of clusters. In doing so, it repeatedly updates the position of the cluster centers such that observations within a given cluster are as similar to one another as they can possibly be while being as dissimilar as they can possibly be to observations within other clusters at the same time (Hartigan and Wong, 1979). Its efficiency in computation and the ability to handle big data have made it one of the most widely used clustering algorithms in many fields, including marketing. Some applications in marketing include customer segmentation, product positioning, forecasting, etc.

The algorithm's biggest limitation lies in the fact that it requires the number of clusters to be specified in advance, which is not always easy in many cases. Several techniques have been proposed to solve this problem. The first one is known as the elbow method, which selects the number of clusters by finding the point at which adding more clusters results in diminishing returns in the form of decreased cluster variance. The second is the silhouette coefficient,

which measures the fit of each data point in the cluster compared to the points in the adjacent clusters (Rousseeuw, 1987).

Hierarchical clustering, on the other hand, uses a completely different philosophy by constructing a tree-like structure of nested clusters, called dendrogram, that reflects the progressive splitting of observations according to their similarity or distance one with the others (Murtagh and Contreras, 2012). This may be performed in a bottom-up manner, where every observation starts in its own class and clusters are formed by merging the two most similar clusters (agglomerative clustering), or in a top-down manner, where the set of observations starts in a single cluster and splits into two or more clusters in a series of divisive clustering steps.

Regarding the computation of the distance between two clusters in hierarchical clustering different linkage criteria have been established: single linkage considers the minimum distance between two observations in the two clusters, complete linkage considers the maximum distance, average linkage considers the mean of all distances and Ward's linkage considers the minimum increase in the total variance of the clusters (Ward, 1963). The choice of the linkage criteria can significantly impact the shape of the clusters and their interpretability with Ward's linkage in particular showing to yield compact and well-separated clusters (Murtagh & Legendre, 2014) that can be more beneficial in applied marketing contexts.

One of the major advantages of using hierarchical clustering over k-means is that the number of clusters does not have to be predefined and the dendrogram plot allows the researcher to visually inspect the data at different levels of granularity using the same analysis, cutting the dendrogram tree at different heights. However, the disadvantage is the substantially increased computational burden of

agglomerative clustering methods, with running times extremely higher relatively to the number of cases in the dataset, making it potentially unfeasible for large datasets without the use of approximations (Müller et al., 2011). Researchers have also proposed hybrid approaches that can potentially leverage the advantages of using both clustering methods, such as using k-means clustering to reduce the dimensionality of the data and then using hierarchical clustering on the centroids of the clusters rather than the original data (Bouguettaya et al., 2015).

It is also worth noting that both techniques are also sensitive to the scale of the input variables, as it is expected that the distance-based computation would naturally favor the variables that have the widest numerical scale if the data is not standardized in advance. Preprocessing operations such as z-normalization or min-max normalization are typically performed in advance of the clustering operation to ensure that all variables are equally weighted in the distance computation (Raymaekers and Zamar, 2020). In addition, the distance metric computation, which could be Euclidean distance, Manhattan distance, cosine distance, etc., is also expected to impact the cluster structure in meaningful ways, depending on the choice made in this regard, although the choice would naturally depend on the specific data set as well as the specific marketing question being asked in the process.

As has already been mentioned in the course of this section, both techniques have seen significant use in the context of marketing research. Plus, it is noteworthy that both techniques have also seen significant use in conjunction with supervised learning algorithms which have seen significant use in the context of enhancing the predictive accuracy of the supervised algorithm by training it on the clusters rather than the entire data set (Xiahou et al., 2022).

It is now time to delve into the specific application of these two unsupervised clustering techniques to solve different marketing problems.

Customer segmentation is arguably the most commonly cited marketing application of clustering algorithms. A considerable body of empirical research has been conducted to verify the effectiveness of both k-means and hierarchical clustering in various marketing contexts. Punj and Stewart (1983) provided a review of cluster analysis as a classification technique in market segmentation. The authors concluded that iterative partition methods, such as the K-means method, are superior to hierarchical methods because they are less affected by outliers and irrelevant variables. To obtain better results, the authors of the review recommend a two-stage process: first, using a hierarchical method to determine the number of clusters and initial cluster centroids, and then using an iterative method to refine the results. The resulting clusters must have external validity, meaning they must relate to variables not included in the initial cluster analysis and must have management utility.

Kansal et al. (2018) conducted a research on customer segmentation using k-means, agglomerative and mean shift clustering techniques in their research paper. Their research data was based on a local retail shop. With a sample size of 200 customers, two important features of the data were analyzed: average annual amount spent on shopping and frequency of visiting the shop. The data was preprocessed by applying a standard scaler. The elbow curve approach was also implemented to identify the optimal number of clusters. Based on their analysis, five distinct segments of customers were identified: Careless, Careful, Standard, Target and Sensible customers. The results were validated by using silhouette scores and it was found that k-means and Agglomerative clustering techniques performed better than mean shift clustering. This helped the retail shop to move

away from mass marketing to a more personalized approach to increase revenue by improving customer relationship management.

Kumar et al. (2025) offer a direct and practically grounded comparison of k-means and hierarchical clustering for customer segmentation in the e-commerce context, addressing a gap in prior research by focusing on customer search behavior rather than the more commonly used purchase history data. The authors argue that search patterns provide a more forward-looking signal of customer intent and brand preference, making them a more proactive basis for segmentation than transactional records that only reflect past behavior. The study uses a dataset of 30,000 customer entries with 38 attributes, predominantly capturing the frequency with which customers searched for specific brands such as Jordan, Samsung, and H&M. After preprocessing the data through missing value imputation and min-max scaling to ensure equal feature contribution, the authors applied both algorithms and evaluated their outputs using silhouette scores and cluster visualizations. The optimal number of clusters for k-means was determined through a combination of the elbow method and silhouette analysis, both pointing to three as the ideal value. Hierarchical clustering was implemented using Ward's linkage and Euclidean distance, which minimizes within-cluster variance at each merging step and tends to produce compact, well-separated groups. In terms of performance, k-means produced a silhouette score of 0.293 compared to 0.254 for hierarchical clustering, indicating that it generated more distinctly separated segments, a finding also supported visually through cluster plots. The three k-means segments identified were high searchers with low purchase conversion, loyal customers with high order volumes and moderate engagers falling between the two extremes. Each segment was then mapped to a specific marketing response, with targeted promotions and personalized recommendations proposed

for the low-conversion group, loyalty programs for high-value customers and A/B testing strategies for moderate engagers. The study concludes that while both methods are viable tools for e-commerce segmentation, k-means is the more suitable choice when dealing with large-scale behavioral data, given its superior ability to produce well-separated, actionable segments at manageable computational cost.

In conclusion, the studies presented above seem to illustrate an overarching picture of clustering techniques as flexible tools within the marketing analyst's toolkit, with significant potential for generating relevant customer intelligence across a wide range of marketing applications, often by their combination with other techniques or with domain-specific knowledge.

Principal Component Analysis

Moving on with our overview of the unsupervised machine learning approaches used nowadays to solve marketing problems, it is now time to introduce principal component analysis (PCA). Similarly to what we have seen with k-means and hierarchical clustering, PCA does not involve making predictions about future events. Still, as a statistical method applied to unlabeled data without an external criterion that defines its desired outcome, principal component analysis belongs to the subset of unsupervised machine learning techniques. Formulated by Pearson (1901) as a way to fit lines and planes to a point cloud, PCA quickly emerged as one of the most popular statistical tools used to this day. PCA basically performs dimensionality reduction: in essence, given a high-dimensional data table with potentially correlated variables, a new low-dimensional table with new orthogonal variables known as principal components is created. Each of these principal components capture a certain proportion of the initial data's variability. PCA is

thus a dimensionality-reducing approach that increases the simplicity of the dataset while losing as little relevant information as possible by creating a new data table with new orthogonal variables that capture as much variance as possible. This is as Gewers et al. (2021) elaborate in their extensive overview of the underlying theory and application of PCA: such a dimension-reducing transformation enables one to capture most of the information contained in a data set, while operating in significantly fewer dimensions, putting practitioners in a condition which is very useful when working with the complex high-dimensional datasets typical of marketing.

From a marketing analytics perspective, the relevance of PCA is primarily due to two of its main applications that, once again, do not fall too far from what we saw in the clustering section: preprocessing and an exploratory analysis.

PCA application in marketing has been exemplified by many research papers dedicated to developing various algorithms that incorporate it as a core component of their analytical pipelines. To give a notable example, Alkhayrat et al. (2020) demonstrated the usefulness of PCA as a preprocessing step in marketing analytics by applying it to a dataset of a telecom company with information on 100,000 clients, who were characterized by 220 behavioral and demographic variables, and dividing them into clusters with the help of k-means clustering. Thus, the researchers stated that using PCA before performing any clustering procedures can be considered as a useful way of tackling high-dimensional data with linearly correlated features. The findings have shown that the use of PCA to reduce dimensionality positively impacted the quality of clustering, which indicated the usefulness of the procedure as a preprocessing algorithm that improves the efficiency of clustering.

Another notable example of PCA in marketing analytics is provided by Abdulhafedh (2021), who used it together with k-means and hierarchical clustering to segment 8,950 credit card owners in order to define a company's marketing strategy. The author proposed a two-stage solution that involved applying PCA and determining the appropriate number of clusters to use and then using the clustering techniques mentioned above in order to find out that PCA could be effectively applied during the clustering process alongside k-means and hierarchical clustering as a validation and optimization technique.

1.2 Natural Language Processing

After investigating machine learning approaches for structured data, we will now delve into another subset of Artificial Intelligence: that of Natural Language Processing (NLP). Although NLP itself is a separate area of study concerned with the relationship between computers and human language, it often uses machine learning to make sense of the ambiguities inherent in language. For example, if determining sentiment is an NLP problem, a neural network may be used as the underlying algorithm to determine whether a piece of text is positive or negative.

As defined in the popular introduction to the domain by Nadkarni et al. (2011), NLP represents the branch of computer science and linguistics focusing on the development of technologies that can interpret natural human language. In other words, NLP makes it possible for a computerized system to take a sentence composed in any natural human language like English, Spanish or French and perform different tasks such as understanding the meaning conveyed by the sentence, determine the sentiment expressed, translate the message or summarize the sentence .

A variety of operations need to be performed in order for a computer system to understand a sentence.

Khurana et al. (2023) have provided a systematic review of the field, outlining multiple levels in which languages could be processed by computers. NLP could be grouped into stages that correspond to various linguistic layers starting from syntactic parsing of sentences, going further into the process of lemmatization, named entity recognition, and so forth. The basic operation that needs to be carried out on any sentence in NLP is tokenization, i.e., splitting up a sentence into tokens such as words and phrases. Afterward, more advanced operations can follow depending on the particular purpose of the analysis, which includes, but is not limited to, filtering out the stop words (such as prepositions, articles, auxiliary verbs, etc.), applying lemmatization and stemming techniques, carrying out parts of speech tagging for each token or even implementing named entity recognition procedures. These operations can be viewed as the main components in the initial preprocessing pipeline that needs to be applied for any NLP-based analysis, which includes the aspect-based sentiment analysis carried out in our study.

The reason why NLP is considered to be highly relevant to the field of marketing has to do with the huge amount of textual data being generated by consumers daily. As already mentioned in the introduction, an impressive amount of digital text has emerged in recent years due to the expansion of the Internet, thus making available a wide range of resources from where the consumer researchers can obtain valuable data. To grasp the extent of this, it is enough to think about online consumer discussion threads, online reviews, archives of press release news articles and blog posts.

Indeed, Humphreys and Wang (2018) in their pioneering work dedicated to text

analysis for consumer research, perfectly encapsulate the main motivation for the growing importance of NLP for marketing researchers: traditional surveys based on direct questioning can provide an insight into what people want to say; however, the vast amount of text data available in terms of online reviews, forum posts, social media conversations, and complaints can give the researchers a picture of consumers' actual behavior and thoughts, free from researcher-induced bias.

Berger et al. (2020) argue along the same lines in their influential article on text analysis in marketing, highlighting the fact that the words are present in almost all market exchange situations, with online reviews, customer service call transcripts, press releases and communication messages creating a massive pool of text data available for analysis and interpretation through text analysis tools and thus serving as means both of prediction and understanding of consumer behavior. In this context, the relevance of NLP to marketing research does not lie in replacing traditional data collection techniques but rather in extending the researchers' ability to analyze the available information.

1.2.1 Latent Dirichlet Allocation

Latent Dirichlet allocation (LDA) is the first technique we analyze in terms of proper text analysis. It is an unsupervised learning technique applied to NLP tasks.

LDA has been proposed by Blei et al. (2003) in one of the most cited papers in machine learning research. According to the authors, LDA is a generative probabilistic model for a collection of discrete data items such as texts, where each document in a collection is represented as a finite mixture over an underlying set of topics, where each topic is further represented as a mixture over an underlying set of word probabilities. Simply put, the basic idea behind this approach is quite

obvious: it is rare that any particular document, whether it is a customer review or a message posted on social media, is "about" a single matter only. Much more realistically, it mixes together several themes in different proportions. For example, a review of a hotel may contain predominantly discussions about room quality, but still also mentions breakfast and location. LDA attempts to reverse engineer this sort of mixed thematic structure from a whole corpus of texts, without requiring humans to categorize or label any particular document beforehand.

In his concise overview of probabilistic topic models, Blei (2012) defines topic modeling algorithms as statistical approaches aimed to discover themes in a text, along with their connections and changes in these connections over time. Crucially, they do not require any labels, as the discovered topics emerge through analyzing the original texts themselves. This puts them inside the unsupervised learning category, despite supervised variants existing. The technical explanation behind this is quite simple: LDA assumes a fictitious process behind the creation of texts. Specifically, for each document in a corpus, some distribution over topics is first picked up at random. Then, each word in a given document is generated by choosing the appropriate topic from this distribution and generating a word accordingly to the latter. The model is then applied in the reverse order: based on the observed words, the most likely set of topics and corresponding distributions is recovered.

For a marketer, the advantage of this method is enormous: it allows one to quickly analyze the contents of hundreds or even thousands of textual documents and determine the predominant topics in the text, thus avoiding the impossibility of manual analyses of such a large number of documents. As pointed out by Hannigan et al. (2019) in a review of topic modeling in management studies, scholars increasingly apply topic modeling for uncovering phenomenon-based constructs

and concept-based relationships in textual data. These efforts result in new insights related to online communities, product and service review analysis, newspapers analysis and other fields.

So how are researchers and practitioners currently implementing latent Dirichlet allocation in their problem-solving efforts?

As is now clear from the ever-growing corpus of empirical research, the answer is: in an impressively diverse range of ways. For an insightful summary of this emerging research field, one may refer to the comprehensive systematic review carried out by Reisenbichler and Reutterer (2019), in which the authors analyze 61 studies using topic models to tackle marketing-related problems. It should be noted that their review unequivocally demonstrates the increasing attention that marketing scholars have been paying to employing topic models in marketing applications over the past decade, as well as the significant progress that has been achieved in several subfields of marketing, with LDA serving as a basis for analyzing latent topics in high-dimensional textual data.

One of the most prevalent and practically applicable fields of application of topic models is associated with the analysis of customer reviews to identify the most discussed features of a product or service.

The research by Guo et al. (2017) presents a particularly illustrative example of the applicability of LDA to sentiment detection. The authors use LDA to analyze reviews of 266,544 hotels in 16 countries, encompassing 25,670 individual properties. In their work, the authors manage to extract 19 dimensions that should be controlled by hotels in order to improve their interaction with customers and to identify variations in their relative importance among various demographic groups,

allowing for perceptual mapping based on hotel star ratings.

This type of in-depth analysis of what factors influence customers' perceptions of a particular company through a relatively straightforward process of automatically extracting the key variables from massive datasets is certainly a revolutionary breakthrough compared to traditional methods of questionnaire surveys that can only be designed based on pre-conceived assumptions.

A second highly valuable use of topic models in marketing involves the identification of comparative competitive strengths between two or more brands based on the analysis of reviews of their products or services. Wang et al. (2018) introduce a framework for applying LDA to the analysis of reviews of two competing products simultaneously, with the ultimate goal of identifying their competitive advantages. Specifically, their framework first applies LDA to capture the latent topics that correspond to the voices of customers reflected in the unstructured large-volume online reviews of each product, and then analyzes these latent topics through a series of statistical operations, such as topic difference analysis to detect unique topics for each product; relative importance analysis to identify competitive advantages of each product and topic heterogeneity analysis to detect differences in consumer preferences. The authors note that the results generated by their framework provide intuitive insights that help product designers and e-commerce companies to better compete with their rivals. Furthermore, the authors suggest extending their framework to the analysis of online communities and even news articles, in order to achieve more granular and comprehensive insights.

Finally, it is worth mentioning that the use of topic models has also found fertile grounds in the analysis of user-generated content regarding product launches. For

instance, Liang et al. (2022), whose study involved analyzing online reviews posted on JD.com, demonstrated how LDA could be applied to identify the key topics discussed by customers when evaluating newly launched products. Specifically, the authors found that the use of LDA allowed for efficient organization of text information, conversion of unstructured review content into computational topic structure and identification of hidden topic information within large-scale document collections. More precisely, the authors identified five topics reflecting different aspects of consumer concerns, including design factors, setup factors, logistics factors, after-sales factors and user experience factors. Additionally, they discovered that the impact of online reviews on customers' purchase decisions was significantly different for each of these topics.

1.2.2 The Bag-of-Words model

The next model we are about to analyze is the bag-of-words (BoW): a natural language processing technique used in machine learning pipelines. Conceptually, the bag-of-words goes back to Harris (1954) who showed in his study of distributional structures in linguistics that words could be identified and classified based on co-occurrence frequencies rather than in syntactic and grammatical terms. Building on this insight, the bag-of-words model translates a document into a simple list of words and how frequently they occur, representing the document as a set of word counts.

The idea behind BoW is typically referred to conceptually in literature, without researchers explicitly naming it. Instead, they highlight specific technical actions such as building a "document-term matrix" and word "frequency counts".

One of the ways BoW is implemented in marketing research was suggested by

Netzer et al. (2012). The researchers proposed a text mining methodology to "listen" to customers through analyzing online forums. The authors used BoW logic to quantify the co-occurrence of various brands or product characteristics, which made it possible to transform qualitative discussion into quantifiable perceptual maps. The maps helped to visually uncover the market dynamics and brand positions similar to those generated during standard survey analysis, but with a cheaper process that did not require direct interviewing of consumers.

Another example is provided by Tirunillai & Tellis (2014), who employed word-frequency representation of textual data about online discussions. Based on the logic that lies behind the BoW representation, the researchers used this method to analyze thousands of reviews in multiple product markets. By concentrating on word frequency rates, it became possible to detect satisfaction factors and evaluate the dynamics of brand position in the market. It is worth mentioning, however, that here the BoW logic is used to create word counts aimed at feeding an unsupervised LDA framework, so BoW itself does not constitute the whole data analysis pipeline.

There are also other useful frameworks proposed for extracting marketing intelligence from customer comments in the form of unsolicited feedbacks. According to Glance et al. (2005), a large-scale system was designed for analyzing consumer feedbacks found in weblogs and message boards, which used a BoW document modeling. This BoW system models the document content, regardless of word order but paying attention to its presence within each document. For example, while classifying topics mentioned in consumers' feedbacks, the system normalizes all references to certain keywords to be analyzed together. In particular, the system would consider all "Axim," "Dell Axim," and "Dell PDA" as the same keyword associated with a particular brand. With the use of machine learning

techniques based on the BoW model as well as linguistic analysis, the authors were able to extract facts about brands and particular product characteristics.

In more recent studies, the efficiency of the BoW model for social media analysis was further confirmed. In particular, in his study, Lee (2024) demonstrates that BoW models should be distinguished according to the sets of adjectives expressing positive, neutral, and negative emotions, outlining a first step towards modern sentiment analysis. Transforming texts from raw data into BoW-based numerical vectors allows to use the information about qualitative emotional expressions (e.g., "joy" and "laughter" for positive ones and "despair" or "bitter" for negative ones). In comparing the performance of several learning algorithms with BoW data, Lee concludes that when combined with the BoW model, the Random Forest algorithm produces the most accurate prediction results concerning users' emotional predispositions. It is therefore evident that despite the increase in complexity of the current social media data, BoW remains a powerful and useful tool for enterprises willing to automate text data processing.

Lastly, Büschken and Allenby (2016) assess the effectiveness of the bag-of-words model in contrast to its sentence-based representation counterpart. Leveraging hundreds of reviews on hotels and restaurants available on Expedia and We8there, the authors state that even though the bag-of-words model proves effective when applied to big data processing, the context-sensitive property of language (e.g., the word great meaning either a positive experience or disappointment) has been ignored in this approach. Restricting their model to represent words from each sentence only from the same topic proves to be highly effective, offering many more "distinguished and coherent" insights regarding customer satisfaction. It becomes clear that there is a very crucial evolutionary step here, since it is evident that not only word frequency but also the word relations matter, which justifies the

usage of context-aware techniques and architectures, such as the DeBERTa model, which represents the destination of this literature review.

1.2.3 Rule-based sentiment analysis

With the bag-of-words technique outlined in the previous section, the task of mapping each piece of text to an array becomes a simple affair. Still, the information about the appearance and frequency of certain words does not give any clue regarding the emotion that underpins the choice of words made by the author of the textual data. Reviews that contain words such as "fast," "sharp," and "reliable" are undoubtedly favorable towards a product, while reviews featuring such words as "slow," "broken," and "overpriced" tend to indicate a negative opinion.

Rule-based sentiment analysis, or lexicon-based sentiment analysis, implies the comparison between the words contained in the text and a pre-defined lexicon in which those words have certain polarities depending on their emotional load. The fundamental idea is that if there are many positive words in a text, it is safe to say that the text is favorable in its essence; if there are many negative words, it is a negative text.

The earliest example of sentiment analysis with the help of a lexicon-based approach dates back to one of the first papers dedicated to sentiment analysis. Turney (2002) demonstrated that it is possible to automatically detect the sentiment of a text (be it a positive or negative text) relying only on linguistic knowledge and association rules, without training any model. According to the researcher, the classification of reviews depended on the semantic orientation of the adjective or adverb-containing phrases of the text: if the phrase was associated with something

positive, the text was considered a positive text. Similarly, if the phrase was associated with something negative, the text was considered a negative text. The results of the algorithm's application were impressive: 74% of 410 reviews analyzed from Epinions (regarding cars, banks, movies, and destinations) turned out to be correctly classified. This achievement shows that there is much information in texts' linguistic structure.

After seeing rule-based algorithms prove effective for document-level sentiment classification, researchers soon extended them to analyze finer-grained sentiment about various customer opinion targets. In particular, the rule-based framework was adapted to automatically identify which features of a product received praise or criticism within the review. Hu and Liu (2004) offered one of the first rule-based frameworks to achieve this task and automatically extract product features along with the sentiment in the customers' opinion. Unlike existing approaches which operate at the review-document level, their solution identifies features and opinions on them at the sentence level, getting closer to the aspect-based sentiment analysis which represents the final destination of this work.

Their framework performs a number of tasks: identifying and extracting features, finding opinion-bearing sentences, extracting opinion-bearing words and assigning sentiment orientations to each word and to the overall sentence. Let's see how it actually works in practice.

First, product features are extracted, then the most frequent features are found and, afterwards, several more pruning steps are performed to remove the non-relevant ones. Once relevant features have been found, opinion-bearing sentences need to be located within the text and, inside those sentences, opinion words (typically adjectives) are extracted and assigned sentiment orientation. Lexical resources

such as WordNet may be used as additional information for determining sentiments. After the semantic orientation of individual opinion words is known, it is aggregated to calculate the overall sentiment orientation, which leads to building summaries listing features of the products along with the counts of positive and negative sentiments elicited towards them.

In contrast to DeBERTa-based methods such as the one just used in this analysis, the solution proposed by Hu and Liu uses manually engineered linguistic patterns to extract and analyze data. Rule-based algorithms rely on linguistic patterns such as specific syntactic constructions (like adjectives expressing opinions), while DeBERTa uses neural networks which implicitly capture more subtle dependencies within the text. While linguistic patterns provide an easier-to-interpret solution, they do not scale well to more subtle language phenomena and depend heavily on the quality of the rules and lexicons defined beforehand. Details of DeBERTa will be addressed in later sections.

It is evident that the business value of this method consists in providing brand managers with concrete data regarding the features that customers like or dislike about the product in a semi-automatized way.

While relatively simple, rule-based methods continued developing over the years. One of the most commercially important contributions to this field was made by VADER (Valence Aware Dictionary and sEntiment Reasoner). Unlike early rule-based systems, it focused on analyzing the unique features of sentiment expression in social media through capitalization, punctuation, emoticons and slang (Youvan, 2024). VADER is a simple but effective way of conducting sentiment analysis with the help of a set of rules. Moreover, the algorithm takes into account the list of features, carefully selected in such a way as to provide sentiment analysis of a

social media-type text, as well as five general rules representing grammatical and syntactical ways of expressing and emphasizing sentiment. When tested against eleven state-of-practice benchmarks, VADER demonstrated better results than individual human raters and surpassed other tested methods in terms of its performance in diverse conditions (Hutto and Gilbert, 2014). This tool, therefore, provides a powerful and understandable method for sentiment analysis of social media data, where rule-based models usually do not work due to non-conformity with assumptions about the language.

Moreover, VADER has been successfully utilized in marketing environments characterized by products. In her research, Sally (2022) utilized VADER to annotate the sentiment expressed in the YouTube comments section of three smartphones that had been released in 2021, namely Apple iPhone 13, Google Pixel 6, and Samsung Galaxy S21. Taking advantage of VADER's ability to process the informal language, the tool was applied to classify each comment as either positive, negative or neutral depending on its compound score. Consequently, it was found out that all three phones had garnered overwhelmingly positive sentiments from consumers, with the iPhone 13 getting the most positive comments. The VADER-generated labels were then employed to train and validate three different machine learning classifiers: Naïve Bayes, Decision Tree and Support Vector Machine. Among these models, the Support Vector Machine was found to perform the best when applied to all three products' reviews in terms of accuracy and F1-scores. Hence, this case study shows the utility of VADER for annotating consumer opinions in video-sharing websites, which helps marketers to measure the sentiment of their products without having to rely on manual operations.

Furthermore, Pandey et al. (2023) have established that sentiment analysis of social

media using NLP techniques yields actionable knowledge that can benefit marketing strategies. For instance, it may help marketing practitioners determine the features of their products that cause consumers to have positive or negative opinions about them.

After seeing the logic and applications underlying these last two techniques, we can think that bag-of-words transformations coupled with rule-based sentiment analysis should be regarded as a first level of automatic sentiment analysis of consumers' text data. Since our analysis goes further and relies on more advanced technologies, it is important to delve into a high-level overview of neural networks.

1.3 From Neural Networks to the Transformer architecture.

1.3.1 Neural Networks

In order to bypass the limitations imposed by rule-based and frequency-driven approaches, it is necessary to resort to complex machine learning tools that place artificial neural networks at the forefront.

Schmidhuber (2015) summarizes the research made around ANNs, which can be defined as a type of computer model that consists of numerous simple computing units called neurons. Every neuron receives certain signals coming from other neurons using the weights attached to the connections between neurons, which then produce some numeric values as output. In such a way, data passes through an input layer and several hidden layers, where it is processed and transformed and, finally, the output appears. Thus, learning involves solving the problem of determining the optimal value of weights between each connection, which makes the ANN display a certain behavior. For example, classifying product reviews as

positive or negative. Learning takes place with the so-called backpropagation algorithm, which estimates the contribution of each particular weight into the prediction error and tunes the value of that weight until all the outputs correspond to their targets. If there are many hidden layers, every next layer represents the abstracted information about the input that was not provided explicitly but discovered by ANN itself while working with data. According to Schmidhuber (2015), the crucial difference between shallow and deep learners lies in the chain of transformations of the data within ANN's hidden layers.

The term deep neural network (DNN) implies that the hidden layers of neurons within a network number more than one. Successive hidden layers develop more abstract representations of data, where lower layers typically recognize simple features of the input signals while higher layers are able to capture complex interactions between various features. Larger computation capabilities of recent computers have allowed researchers to increase the number of layers to enable accurate and precise calculations in larger data sets (Goodfellow, 2016). This capability along with increasing amounts of data mentioned in the introduction of this work has resulted in a massive development of neural networks and turned them into cutting-edge tools used in many data-driven industries.

In marketing specifically, ANNs found their use in two areas, which have been widely documented by recent research. The first application area is related to forecasting the sales and demand for a product. Time-series models like ARIMA require linearity and stationarity assumptions that are often unfulfilled in commercial data. Newer classes of machine learning algorithms such as ANNs outperform ARIMA because they can handle complicated nonlinear relations between causal factors. An example of superiority of neural networks in forecasting has been given by Mansur et al. (2025) that have shown superiority of

ANN over linear regression and ARIMA in forecasting sales in all conditions measured in terms of root mean square error. Therefore, the practical importance of ANNs for marketing executives can be estimated as better forecasting accuracy which translates to lower storage costs and optimal pricing decisions.

Another application area concerns predicting customer churn. As the marketing literature proves consistently, acquiring new clients requires much greater expenses than retaining existing ones. Therefore, the problem of predicting customer churn deserves close attention of marketing analysts as ANNs possess excellent abilities in building non-linear dependencies among multiple behavioral customer indicators and identifying the probability of churn. Tsai and Yu-Hsin (2024), in their paper studying ANN-based decision models through an American telecom company's dataset containing 34,761 churners and 16,545 non-churners, have proven that combining two ANNs performs best when making churn predictions. Thus, using ANNs for churn prediction, marketers can develop a proactive retention strategy for specific customers instead of reacting to customer churn.

While the power of deep neural networks is undeniable, sequential text processing by means of neural networks is associated with certain limitations: computational inefficiency and limited ability to understand great token contexts within the input sequence. Precisely tackling these inefficiencies, Vaswani et al. (2017) introduced the revolutionary Transformer architecture which relies solely on self-attention mechanisms.

1.3.2 The transformer architecture

The transformer architecture brings a crucial innovation that solves the limits of

simple artificial neural networks: parallelization and long-range dependency. In fact, ANNs process tokens one at a time, increasing computation time and cost and hindering true parallelization. Moreover, simple ANNs struggled to maintain precise information about tokens that appeared far earlier in a sequence, making the model “lose context”.

These limits made it particularly timely to ask the question explored by the authors in the seminal paper "Attention Is All You Need" (Vaswani et al., 2017). The question put into practice by the researchers working for Google Brain in their paper is whether it is possible to build a purely attention-based translation model that can outperform the existing architectures.

For the experiment, the authors designed and tested a new architecture. This architecture, the Transformer, was trained on two datasets of translations; from English to German and French. Comparisons in quality of the generated translations and the time required to complete the training process showed that the proposed Transformer surpassed all existing recurrent architectures in terms of effectiveness. Moreover, due to its parallelisable nature, the training period was drastically reduced compared to any existing approach. It is crucial to note here that this was done not due to improved quality per se, but thanks to the possibility of processing the whole sequence at once rather than sequentially.

Three main innovations introduced by the Transformer architecture can help explain the effectiveness of its application in language understanding, which is also why models such as the DeBERTa v3 came to existence.

As already mentioned, the core mechanism introduced by the transformer is self-attention. In any language sentence, words usually do not contain meaning in

themselves. Instead, they depend on each other.

For instance, in order to understand the meaning of the word “which” in the sentence "the product arrived damaged, which was not what the customer expected", one needs to know the meaning of the word 'damaged', located earlier.

Recurrent neural networks coped with this challenge in a somewhat inconvenient way: by sequentially transferring the word meaning from the beginning to the end of the sentence. Due to this, the more information the network has to pass until the end of the sentence, the greater the risk of inaccuracy.

Attention allows you to bypass this step: all words in the sentence are able to make comparisons with all other words and calculate their relevance to each other. That is, at each moment of time, a word receives its weight depending on how much other words contribute to its meaning. In the case of a self-attention layer in the encoder, keys, values, and queries come from the same position at once (Vaswani et al., 2017).

Another innovation brought by transformers is multi-head attention. Instead of applying self-attention once, the Transformer uses several instances in parallel. Each instance, also called a head, independently learns various linear projections for the query, key, and value matrices. (Vaswani et al., 2017) By doing so, the model becomes capable of taking into account various aspects of the inter-word dependencies at once. In practice, one head can analyse the grammar (for example, connecting verbs with their subjects), another can trace semantic dependencies (connecting the attribute and the noun, for example, "the car is fast"), and the third can resolve pronominal references, which are responsible for detecting the pronouns and which word they point to in a sentence.

Unlike recurrent networks, Transformers process words simultaneously, without a clear order. This brings up an issue that needs to be addressed: that of positional encoding. Thus, positional information has to be added to the representation of each word manually. In the first version of the transformer, the input was augmented with the so-called positional encodings, which consisted of sinusoidal mathematical functions (Ahmed et al., 2023). This permits to give the input coordinates and understand where the words stand in the sequence.

We just saw how the attention mechanism allows neural networks to cope with long-range token dependencies. But transformers fix the second limit we introduced earlier as well: in fact, they enable parallel operations of the network. So, if with neural networks that processes tokens sequentially the time required to train models increases proportionally with the length of the sequence being processed, with transformers this bottleneck ceases to exist (Vaswani et al. 2017).

1.3.3 The BERT architecture

This advantage of the Transformer architecture has had far-reaching implications. In two years, it became the foundation of most modern language models thanks to the BERT architecture introduced in the paper by Devlin et al. (2019) titled "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding".

The main question posed in the paper relates to the possibility of creating a universal Transformer model that would be trained on a huge number of texts and then fine-tuned specifically for each task. The methodology consisted in pretraining the model on the whole English Wikipedia along with a large corpus of books and using two kinds of self-supervised learning tasks: predicting

randomly masked words and determining whether two sentences follow each other in a paragraph. The authors, then, fine-tuned the resulting pre-trained model using eleven NLP tasks. The results of the study demonstrated a remarkable quality of the model: BERT surpassed existing recurrent networks in eleven standard benchmarks, reaching a GLUE benchmark score of 80.5, which set itself 7.7 points up compared to previous methods and getting 93.2 points at a question answering test, which exceeds human performance (Devlin et al. 2019).

The main innovation of BERT, which stands for Bidirectional Encoder Representations from Transformers, is that it utilizes a specific component of the transformer, the encoder, to process text in both directions simultaneously, allowing the model to capture deeper context than traditional left-to-right systems. Traditional transformers, on the other hand, are unidirectional, despite processing information in a parallel way rather than sequentially: it means that if we take the word “bank” in the sentence “I went to the bank to deposit money”, a regular transformer would only know be aware of the words “I went to the”, while BERT would look at the whole string of text in both directions.

1.3.4 The DeBERTa model

After showing the different machine learning techniques that marketing practitioners have been using in the past and present years to solve different marketing problems in new, accurate and efficient ways and exploring newer technologies such as transformers and their variants, we are equipped to delve into the last model of this literature review: the Decoding-enhanced BERT with disentangled attention (DeBERTa).

While the BERT architecture provided a foundation for transferrable skills with

regard to language, the emergence of the DeBERTa framework was specifically aimed at resolving specific complexities of human expression that often result in a failure for simpler models, such as irony or sarcasm.

As far as a marketing strategist is concerned, the most valuable part of the DeBERTa framework lies in its "disentangled attention" feature. In comparison with earlier versions of the algorithm that combine word meaning and word position into one entity, this new approach separates these two parameters, acknowledging the significant variation in importance that a word can have based on its position (He et al., 2021). In the context of sentiment analysis, the word “excellent”, for example, acquires a different overall meaning based on its position relative to specific characteristics of a product, such as battery life or sound quality.

As a consequence, the DeBERTa architecture enables capturing subtle semantic dependencies and sentiments through a higher degree of precision in comparison with its predecessor and other models like BERT or RoBERTa (Vallejo Vera & Timoneda, 2025).

On the practical side, recent studies have demonstrated that DeBERTa is especially powerful when applied to analyze large volumes of user-generated content characterized by low levels of formality, such as Amazon customer reviews. For example, DeBERTa has achieved extremely high accuracy scores in cross-domain sentiment analysis, with the best results amounting to 93.21% accuracy (Cao, 2025).

Lastly, the DeBERTa framework incorporates an “enhanced mask decoder,” which is instrumental in utilizing absolute positional information throughout the learning process (He et al., 2020). This mechanism informs the model about the syntactic function of specific words within a sentence, enabling it to better differentiate between subjects such

as “users” or “products”.

1.3.5 Transformer use-cases in Marketing

Since the transition from traditional machine learning to the transformer architecture has marked a significant shift in how marketing researchers and practitioners process consumer data, the analysis of some noteworthy studies about transformer-driven marketing research is the last step before going into the details of the methodology of this study itself.

The first use-case is an interesting empirical research (Guo, 2024) with the core objective of using state-of-the-art natural language processing techniques to generate accurate sentiment scores of a huge dataset of reviews from Amazon. Guo examines how the sentiment analysis findings relate to decision making and their alignment within the scope of behavioral economics principles such as electronic Words-of-Mouth (eWOM) and confirmation bias. First and foremost, the methodology involves the application of the RoBERTa model, which belongs to a new generation of BERT models. For the purpose of generating reliable sentiment scores, the researcher trained the model by feeding beauty products reviews into it and then evaluated its ability to categorize them as positive or negative.

As a result of the research, one can see that the RoBERTa model is satisfyingly effective for conducting marketing research as it reached 88.44% accuracy and 0.93 AUC for distinguishing between sentiments.

In particular, Guo (2024) discovered that sentiment scores vary depending on different stages of the product development, market penetration and on the intensity of eWOM. This means that, for example, if a random eWOM burst happened, more positive sentiment scores were observable. Another fact worth

noting is that the higher the purchase rate of a certain product is, the lower the average sentiment score as the customer's expectations were more diverse and harder to satisfy fully. From the point of view of a marketing or a product manager, one can conclude that the insights that this type of analysis can provide are critical to product management operations and to conduct marketing activities in real-time (Guo, 2024).

Ahlstrand et al. (2025), used transformers for B2B contractual churn prediction based on customer behavior data, aiming at solving the difficult problem of client retention, where even a small number of lost customers could bring severe financial losses due to high customer acquisition costs. The major goal was to create a scalable machine-learning algorithm using a modified transformer to identify customers that would no longer want to renew their contracts based on the patterns in their daily activity data. These data include mobile data usage statistics, call frequency, revenue trends and even the sentiment of support tickets. Furthermore, the authors applied the technique of conformal prediction to evaluate how confident the model was about each specific decision, enabling managers to understand whether the case required additional attention, as opposed to being considered a lost cause.

Based on the test results, the transformer appears to have good predictive power, with the ability to spot five times more churners in its top-ranked clients compared to a completely random choice and a moderately high AUC of 0.77. Nonetheless, one of the most significant findings is the difficulty in estimating certainty due to lack of clear indicators that a particular client would leave in the data provided. However, this is not true for the top 10% of customers at churn risk: for those ones, the model is particularly confident in its assessments (Ahlstrand et al. 2025).

Business-wise, management should consider the level of confidence of a transformer in making its decisions and use this indicator to decide which customers need to be offered expensive incentives to retain their loyalty. This is related to the complicated and intricate essence of the churn problem itself: as stated in the paper, “The underlying cause is likely a series of events where the company did not meet the customers’ expectations over time. In a B2B setting, contracts typically last for several years, which makes it difficult to pinpoint exactly when customers decide not to renew ” (Ahlstrand et al. 2025).

Xu et al. (2019) describe a new type of task named Review Reading Comprehension, which converts raw user reviews into a sort of knowledge base that should be used for answering user queries based on product reviews.

The first step the research takes is that of modifying and fine-tuning the BERT transformer to account for domain-specific knowledge and subtle nuances of consumer opinions. In their research, authors used a joint post-training process methodology, which involved learning how to use two different sources of knowledge: large-scale unsupervised domain-specific reviews for acquiring knowledge related to the language of certain product categories, and supervised out-of-domain knowledge for improving the mechanism of answering questions in general. It also implied using original BERT pretraining methods such as predicting missing words and identifying sentence relationships in relation to domain-specific reviews. The post-training process had an impressive effect on BERT performance on question answering and aspect extraction: domain knowledge impacted aspect extraction performance while training on question answering datasets helped to locate the beginning and end of the answer inside reviews.

On a practical level, these results imply that businesses may use this technology to create more effective and fast customer service systems by using all the user experience accumulated by organizations. Moreover, this method enables reducing time spent on getting answers to customer questions, avoiding both delays in receiving answers via community support and expenses associated with building custom knowledge bases. Marketers could take advantage of improved automatic sentiment analysis related to particular product aspects, which helps them to find out consumer feelings toward specific aspects. It can become a basis for further product improvement and effective marketing strategies based on data.

2. Research Methodology

In this section, the rationale behind the empirical analysis will be discussed. Specifically, we will shed light on the ways in which the data was collected, why the DeBERTa model was chosen as the foundation of the analysis and finally how the methodology was designed to yield valuable business insights from customer reviews.

Before diving into the essence of this research from a methodological perspective, it is essential to understand what we mean by enhancing marketing strategy.

In the context of Aspect-Based Sentiment Analysis, we can define a marketing strategy enhancement as an increase in the availability of information related to a decision revolving around marketing-centered areas (such as pricing, communication and advertisements, product development or customer experience) that comes from the aggregated observation and exploration of the output of a machine learning model. Such decisions are defined as “marketing strategy-enhancing” as without the possibility of looking at these data, no action would have been performed.

2.1 Model and product choice

The product chosen for sentiment mining in this work is the Meta Quest 3, a virtual reality headset recently released by Meta. The choice was driven by multiple motivations.

The first reason relates to the commercial sphere, as the Meta Quest 3 falls into the space where a new product category intersects with a well-established brand, providing the object of study with substantial sales volume and a significant user

base thanks to the higher perceived trust towards the brand itself.

Second, the model in question has been pretrained on a large number of consumer electronics reviews. Therefore, it is suited to analyze the specialized language that appears in headset reviews. However, we have a reasonable degree of certainty that the training data had only a fraction, if any, of entries related to virtual reality products. This characteristic is a valuable strength, as it prevents the model's output from overadapting to the scraped data. As a result, a model trained on consumer electronics becomes a more neutral tool for sentiment detection in consumer reviews.

Lastly, the DeBERTa model represent a cutting-edge innovation when it comes to machine learning since, as explained in the literature review, allows to capture the relative position of words thanks to the disentangled self-attention mechanism, using the absolute position only as a complementary information. The disentangled attention can be represented by the following equation (He et al., 2020):

$$\tilde{A}_{i,j} = Q_i^C K_j^{CT} + Q_i^C K_{\delta(i,j)}^T + K_j^C Q_{\delta(i,j)}^T$$

Where $\tilde{A}_{i,j}$ is the disentangled attention score between token i and token j , representing how much attention token i should pay to token j once both tokens' content and relative position are taken into account; Q_i^C is the content query vector of token i ; K_j^{CT} is the content key vector of token j ; $K_{\delta(i,j)}^T$ is the relative position key vector corresponding to $\delta(i,j)^T$, which is the functions that returns relative distance between token i and token j and $Q_{\delta(i,j)}^T$ is the relative position query vector corresponding to $\delta(i,j)^T$, the relative distance between token j and token i ; and $\delta(i,j)$ is the function that returns the relative distance between the two tokens' positions in the sequence. The three terms summed in the equation correspond

respectively to the content-to-content interaction between the two tokens, the content-to-position interaction capturing how token i 's content relates to token j 's relative position, and the position-to-content interaction capturing the reverse relationship, together forming the full disentangled attention score.

While this equation shows how the model builds a contextual representation of each review-aspect pair, the core equation underlying the model's prediction that explain how that representation is converted into the sentiment label and confidence score is a SoftMax function:

$$H_o = softmax(\frac{\tilde{A}}{\sqrt{3d}})V_c$$

Where H_o is the output representation produced by the self-attention layer, that is the new, contextualized vector obtained after each token has gathered information from the rest of the sequence; softmax is the function that normalizes the disentangled attention scores into a set of weights that sum to one, so that they can be interpreted as the proportion of attention paid to each token; $\tilde{A}$ is the disentangled attention score matrix computed in the previous equation; $\sqrt{3d}$ is the scaling factor applied to stabilize training, where d is the dimensionality of the hidden representations and the multiplication by 3 accounts for the three components summed within $\tilde{A}$; while V_c is the matrix of content value vectors, which are weighted according to the normalized attention scores to produce the final output.

2.2 Data collection and processing

The raw data collected in the current study includes 426 customer reviews of the Meta Quest 3. The source used for the data collection was the official Meta-

branded Amazon storefront. Specifically, the reviews section of the product detail page of the Meta Quest 3 sold by Meta Platforms, Inc. for each the six markets analyzed in this paper.

The choice of Amazon reviews, as opposed to other feedback platforms such as Trustpilot or the product manufacturer's own website, was made for several reasons. First, Amazon is one of the primary retailers for consumer electronics across all six markets covered by the study. Second, Amazon's review system implies that only verified customers can write product reviews, thus increasing the likelihood of users' having hands-on experience with the product and therefore writing truthful, high-quality reviews. Third, Amazon offers a rich amount of user-generated content about the product in question, thus enabling the researcher to obtain a meaningful number of observations even within the limitations of computational power available for the study.

Data collection was done through a web-scraping browser extension that automatically extracts structured information from online sources and exports it in a comma-separated values (csv) format. Unlike more sophisticated tools, the use of the said browser extension avoided the creation of custom-made code specifically to target the web pages selection, significantly speeding up the data collection process.

Although web scraping of product information on Amazon is against the terms of service, the collected data was used strictly for noncommercial, academic purposes. No personal information of reviewers was obtained beyond their reviews themselves and the country label. Thus, there was no violation of Amazon's terms of service as the collected data is fully anonymized and does not allow for the identification of individual reviewers.

In total, 426 reviews have been obtained in the course of this study. None of the reviews were removed or filtered out during the data cleaning stage: all observations provided by the extension were kept intact in order to reflect the unbiased distribution of the reviews on the platform. The decision to not filter the reviews based on such criteria as the length, the star rating range or recency was dictated by the necessity to ensure the absence of selection bias and, therefore, keep the real opinion distribution untouched. Applying filters would skew the opinion distribution in an artificial way that is not reflective of the actual situation in the market.

Once collected, the data is presented in the form of a flat table with three columns per observation: the unique numerical identifier of the review, the full body of the review and the country label representing the particular Amazon store where the review originated from.

Table 1: description of the variables found in the source data

Variable name	Variable description
Id	Unique identifier of the review
review_text	The unaltered, raw review text as extracted from the Amazon product reviews section
country	A label identifying the market linked to the customer review

This sparse table structure was chosen due to the specifics of the downstream ABSA task, which uses review texts only without metadata such as star ratings or review title.

Regarding country selection, the six markets were chosen on the basis of two

complementary criteria. The first was the desire to construct a cross-market dataset representative of two distinct linguistic contexts: a European, non-natively-English-speaking group, comprising Italy, Germany, France and Spain (some of the biggest European economies, where a high-tech product can find a fertile market) and a natively English-speaking group, represented by the United States and Australia. The second one is to maintain a balance between the number of reviews of each cluster, so that no linguistic context is overrepresented and the model can be observed, if needed, across both subset of data.

This two-cluster design allows the analysis to be relevant to the operational reality of international marketing teams, who frequently need to synthesize consumer feedback across linguistically heterogeneous markets. It should be noted, however, that the distribution of reviews across countries is not perfectly uniform. Within the European cluster, Italy and Germany contribute substantially more observations (97 and 81 respectively) than France and Spain (30 and 20), a disparity that reflects the relative review volumes available on each national storefront (due to Amazon's e-commerce policies) rather than a deliberate sampling choice. The resulting dataset is nonetheless balanced at the cluster level, with the European group accounting for 53.5% of total observations and the English-speaking group for the remaining 46.5%.

Table 2: count of countries for each market represented in the source data and share of each country compared to the total observations in the source data

Country	Count of Country	Share of Total
Italy	97	22,8%
United States	98	23,0%

Country	Count of Country	Share of Total
Germany	81	19,0%
France	30	7,0%
Australia	100	23,5%
Spain	20	4,7%
Total	426	100,0%

An important preprocessing decision concerns the language of the review texts. Although the pre-trained DeBERTa model used for this experiment was developed based on multi-language corpora, the model was proven to perform poorly in terms of aspect extraction when working with non-English languages.

In order to overcome this problem, all the non-English reviews were translated into English by using the “GOOGLETRANSLATE” function in Google Sheets. The rationale behind this choice lies in the ease of use of the translation function, reproducibility and homogeneity of the preprocessing algorithm as well as the need for translation to English that is inherent to the task at hand. Original text of each review and its English translation were included into the dataset in order to be able to trace back each result of sentiment extraction to its source as well as verify the quality of the translation process in a few selected cases.

Empirically, the impact of this preprocessing step can be shown through comparing the aspect extraction rate of the model for the original-language and translated reviews. Thus, when run on the translated dataset, the battery aspect was

found to have sentiments in 131 reviews, whereas for the original-language version only 45 reviews contained the respective sentiment. Similar numbers of sentiments were obtained for the comfort (120 vs 44) and price aspects (90 vs 32). The difference is close to three times the rate of aspect extraction. Therefore, the gain in coverage is quite significant and serves as strong empirical evidence for this preprocessing choice. While the possibility of a subtle sentiment distortion due to machine translation can't be ruled out, especially for the figurative and idiomatic language often used in consumer reviews, this distortion is outweighed by the advantage in model performance obtained through this preprocessing step.

Finally, regarding the model processing itself, the analysis was run through Python 3.14 along with the “transformers” library and the pre-trained model on Hugging-Face (Yang et al., 2023).

2.3 Aspect selection

The choice of aspects for analysis is, arguably, one of the key methodological considerations to make as it defines the scope of the practical analysis and therefore the richness of insights yielded by the model. The aspects selected for this study were based on an exploratory assessment of the product reviews.

The seven aspects selected for analysis include battery, comfort, resolution, price, passthrough, controllers and strap, where “passthrough” is a special product feature that allows users to see their physical surroundings without removing the headset. These were identified on the basis of two perspectives. On one hand, product specifications available on the official website of the Meta Quest 3 headset were analyzed to determine the attributes emphasized by the company to create the value proposition of its product. On the other hand, an initial, unsystematic exploration of customer reviews revealed that these same attributes occurred often

enough in the corpus, making the selection process straightforward.

Of the seven selected aspects, some are technical (resolution, battery and passthrough), while others are experiential (comfort, strap, and controllers). Price is a special case of an attribute because it cuts across both spheres of analysis. It is interesting to note the difference between technical and experiential attributes because they may well correspond to different areas in the field of marketing: while the former are more relevant to product development and competition, the latter are more relevant to the post-purchase stage.

One more thing needs to be emphasized. Namely, the selected aspects cannot be said to form a comprehensive list of all the possible factors that influence customer satisfaction with a product like the Meta Quest 3. Other attributes that could have been included in the corpus include, for instance, sound quality, software or installation. However, due to computational limitations outlined previously in the methodology section, this study deliberately narrowed down the number of aspects.

2.4 Advantages and limits of ABSA

Prior to delving into the findings, it is appropriate to place the technique within a critical business lens. Understanding its practical capabilities and limitations from the perspective of a commercial enterprise, rather than a technical expert, is essential for drawing meaningful insights and making informed managerial decisions based on the outputs.

Advantages

To begin with, one notable strength of using Aspect-Based Sentiment Analysis (ABSA) for market insight generation is its ability to provide information on a

highly granular level. While aggregate measures are interesting in their way, they fail to give actionable signals to individual departments in a product-oriented organization: it means that while indicators such as the average star rating for a product acts as a guiding light, the out of the algorithm in analysis can, when properly inspected, suggest a concrete actions to take in order to change things.

Another advantage, closely aligned to the previous one, is that decomposing overall sentiment into individual product attributes aligns the output directly with the information that each internal stakeholder needs to have. To exemplify: a high number of negative sentiment reviews linked to the “battery” aspect is a direct information for the internal function that deals with that specific part of the product (in this case, it could be either procurement or production). In the end, this still improves marketing strategy as they will be responsible, for example, to effectively communicate an improvement in battery life through advertisement.

Moreover, since attribute-based sentiment scores are generated automatically at the point of extraction from raw data, rather than through manual labor, the results can be shared broadly and quickly across an organization through different tools such as dashboards.

Second, the data underlying ABSA insights captures the topics a consumer deemed worth discussing in his or her own words and without prompting or nudging (which happens with survey data, instead), so it contains valuable insights into the aspects of the product that matter to them. Thus, the consumer who chose to write a lengthy description of the poor ergonomics of the included strap did so unprompted, which is noteworthy and extremely valuable for a product company.

Another benefit, also of a structural nature, is the technique's scalability. Processing hundreds or thousands of consumer reviews per market manually and

drawing valid insights from them takes a lot of human capital and may produce inconsistent results. In contrast, ABSA insights take minimal additional effort to generate once an ABSA model has been trained and tested and the necessary pipeline implemented. This enables companies to continuously monitor changes in consumer sentiment over time, detecting potentially harmful shifts early enough in the product life cycle for proactive intervention to occur rather than having to discover them through lagging indicators such as declining sales or growing returns.

Limitations

As beneficial as it might seem, this kind of analysis also brings some important limitations worth considering.

First, it is essential to point out that in order to implement this technique a fixed list of aspects whose sentiments need to be analyzed is needed. Consequently, the technique will leave out any insights about aspects of the consumer experience that were not included in the initial specification. This characteristic poses a limitation to the scope of applications and requires a business to be clear about what aspects it wishes to monitor. In particular, ABSA is best used for focused diagnostics of a specific well-understood product rather than open-ended exploration of customer sentiment, unless used in combination with topic-modeling algorithms.

Second, while ABSA provides sentiment labels along with the corresponding confidence scores, it does not reveal how exactly the model came up with those scores. For instance, a negative sentiment towards comfort will be flagged by the tool but not described with regards to specific ergonomic features or discomfort-causing situations, such as a hot climate or excessive wearing time. This inability to understand the rationale behind a prediction is an inherent property of machine

learning models called the explainability problem: the model is essentially a black box that generates scores a manager has to believe in without being able to check. Even if nowadays some techniques are currently being explored to enhance models outputs' explainability, this remains an important limitation to keep in mind when using ABSA in a high-stakes decision context (Perikos and Athanasios, 2024).

The third limitation, which is intrinsic to any machine learning system, is the fact that ABSA is bound to make mistakes. The two broad categories of errors the technique might commit are false positives and false negatives. The first occurs when a model detects sentiment where there is none: for instance, when classifying a review as expressing negative sentiment towards a product's battery life despite the neutral or comparative tone of the reviewer's comments. False negatives occur when a sentiment is present but overlooked. Which of the two types is more severe in a given case will depend on the application context of ABSA. For instance, if a company is using ABSA to find problems in its products that need to be solved, false negatives pose a danger of overlooking those problems altogether, while false positives pose no risk in this scenario. Conversely, if ABSA is used to uncover product strengths that can be highlighted in communication with investors or customers, false positives are likely to be more problematic. Depending on the application, then, the error profile of the model should be specified upfront, with the sentiment scores adjusted as needed.

Finally, although rarely discussed in technical literature on the topic, there is another practical constraint related to the review population. Reviewers are clearly not a random sample of consumers and tend to give quite a polarized opinion about the product. As writing a review is an activity with a certain degree of friction, consumers are likely to leave one only when they feel something strong to express (Schoenmueller, Netzer and Stahl, 2020). This does not invalidate the insights the

study uncovered; however, it is an extra consideration that should be taken into account in order to make informed decisions and avoid the risk of overreliance towards these techniques.

Comparing Aspect Based Sentiment Analysis (ABSA) with other traditional text analysis techniques

After looking closely at ABSA's strengths and limits, it helps to compare it to two other main text analytics techniques that can be performed by implementing the other models seen in chapter 1.2, such as Latent Dirichlet Allocation or Rule-based Sentiment Analysis. Doing so clarifies the choice of ABSA for this study highlighting what sets it aside: the higher level of granularity provided through the same input. Table 7 illustrates how traditional sentiment analysis can hide mixed feelings in a single sentence. Topic modeling shares ABSA's focus on finding themes, yet it never judges them as good or bad. It reports only the topics that appear in a corpus. ABSA combines these: it adds the sentiment score to the thematic detail of topic modeling. By zooming in on each aspect, it gives a separate rating for every feature noted in a review. The summary table lists the three methods, their main goal, the granularity level of the information provided as output and a typical input and output examples generated by the models.

Table 3: Comparing Aspect based sentiment analysis and other traditional text analytics techniques

Technique	Primary Objective	Granularity Level	Input Example	Output Example
ABSA	Identify specific features and determine the distinct sentiment polarity for each.	Aspect or Sub-sentence level	“The screen is gorgeous, but the battery life is terrible.”	Screen: Positive (+1) Battery life: Negative (-1)
Sentiment Analysis	Classify the overall polarity (positive, neutral, negative) of an entire text or sentence.	Document / Sentence level	“The screen is gorgeous, but the battery life is terrible.”	Overall Polarity: Mixed / Neutral (single aggregate score)
Topic Modeling	Discover latent themes and cluster texts by statistical word patterns without evaluating sentiment.	Corpus / Collection level	Dataset of 5,000 customer reviews.	Topic 1: [delivery, shipping, delay, package]

3. Empirical Findings

3.1 Structure and interpretation of the model output

At the end of the data processing, the model output was exported in the format of an excel file. The empirical results consist therefore in a spreadsheet with 426 rows, corresponding to a different customer review, and 17 columns. The first three columns are purely descriptive: the *id* column is the numerical identifier allocated to each review at extraction time and allows tracing all model outputs back to their source text; the *review_text* column is the actual text of the review in English and the *country* column holds the marketplace information from which the review was extracted.

The other fourteen columns represent the core of the model's output and can be grouped into seven pairs of features, one for each aspect analyzed. Each aspect has an associated *aspect_sentiment* column reporting the polarity label predicted by the DeBERTa model (either Positive, Negative or Neutral) and an associated *aspect_score* column holding the internal confidence of the model in the classification as a number between 0 and 1, indicating the likelihood of belonging to the indicated class. A value close to 1 is thus indicative of certainty, whereas a value in the range of 0.4 or 0.5 shows that the model had to commit to a classification despite considering the other polarities as likely to be correct.

There are two important considerations regarding the dataset's missing entries.

First, the dataset is relatively sparse, as a polarity is recorded only if there is an indication of the presence of the aspect under consideration in the review text. This means that empty values for each aspect do not always represent missing

information. Instead, some of these missing values can be treated as a meaningful absence: the reviewer did not comment on the given attribute. Second, some missing records are linked to a coverage problem at the review level. Of the 426 reviews, 180, i.e., 42.3% of the corpus, have no extractions for any of the seven considered aspects. This silence within some parts of the dataset comes from either very short and non-specific reviews ("excellent product", "product fully matches my needs"), where no discussion about the predefined attributes can be extracted, or from very specific reviews which refer to attributes that fall out of the scope of the analysis. This predefined aspect constraint is discussed in Section 2.5: all the analysis and findings presented here will hold for the 57.7% of the corpus commenting on the aspects considered.

3.2 Confidence scores and the reliability of the classifications

Before the sentiment patterns themselves can be interpreted, it is necessary to establish how much trust the underlying classifications deserve. Table 2 reports, for each aspect, the mean confidence score over all extracted observations, the mean score conditional on the assigned polarity, and the share of classifications falling below a confidence of 0.6: a conventional threshold below which a label should be treated as a classification tentative rather than a confident prediction.

Table 4: Mean classification confidence in total, by aspect, by polarity and share of observations whose classification confidence score was below the 60% threshold.

Aspect	Mean score (all)	Mean score (positive)	Mean score (negative)	Share below 0.6
Resolution	0.859	0.893	0.712	11.1%
Controllers	0.835	0.861	0.777	9.1%
Comfort	0.834	0.876	0.754	11.8%
Passthrough	0.814	0.829	0.777	14.7%
Price	0.806	0.832	0.767	21.3%
Battery	0.787	0.788	0.788	20.0%
Strap	0.787	0.772	0.814	15.6%

By looking at Table 3, we can understand how the confidence score performed from a high-level perspective across the different product features. For instance,

we witness quite high overall confidence levels, with average values for each of the seven aspects' classification scores clustering around 0.8. Plus, in none of the seven predefined aspects the percentage of tentative classifications is lower than the 0.6 threshold. This, as explained before, is important for the model's prediction to be trustworthy in the first place.

3.3 Aspect salience: what customers spontaneously discuss

The next layer of findings concerns not what customers think, but what they choose to talk about. Since reviews are unsolicited, the frequency with which an aspect surfaces is a revealed measure of its importance in the customer's own hierarchy of concerns. Table 4 reports the extraction coverage of each aspect with respect to the 426 reviews.

Table 5: aspect salience (share of reviews in which the aspect was detected by the model)

Aspect	Reviews with aspect detected	Coverage
Battery	130	30.5%
Strap	122	28.6%
Comfort	119	27.9%
Price	89	20.9%
Controllers	66	15.5%
Resolution	63	14.8%
Passthrough	34	8.0%

The frequency table represents a clear, yet strategically challenging hierarchy. In fact, the top three elements of the hierarchy (battery, strap and comfort), which each appear in almost a third of all reviews, consist in very standard, non-

innovative characteristics, but their bold presence in the data determine hints at high levels of importance for the physical experience of the customers using the product. On the other hand, visual and technical characteristics, explicitly resolution and passthrough, are at the bottom of the list of attributes. Passthrough, which represents an innovative feature that may act as a product differentiator, is barely mentioned, being extracted in only one every twelve reviews. When looking at these results it is important to critically evaluate the possible reasons beneath this phenomenon: at this point in our analysis, the passthrough aspect is not necessarily a bad implementation, rather one that has not become popular yet. In order to make any claim about the success of such a new feature, further exploration of the sentiment will be carried out in the subsequent sections.

3.4 Sentiment polarity and the salience-sentiment matrix

In order to drive further analysis and comparisons among features, two metrics have been calculated and employed. Firstly, *net sentiment* is defined as the proportion of the observations for a certain aspect that are identified as being positive, minus the proportion of observations of the same aspect identified as being negative, expressed in percentages between -100 and $+100$, where zero acts the baseline. For example, the “battery” aspect was extracted in 130 reviews, of which 73 were positive (56.2%), 48 negative (36.9%) and 9 neutral (6.9%). So:

$$\text{Net Sentiment} = 56.2 - 36.9 = +19.2$$

The second complementary metric overcomes the first one’s biggest limitation: not taking into account confidence scores. This means that a review that the model classified as positive for a certain aspect with two distinct confidence scores of 0.5 and 0.9 would have the same weight when calculating the aspect net sentiment of

the same aspect. The confidence-weighted index solves this problem by computing the confidence-weighted average of the scores assigned to each review, multiplied by a factor of +1 if the observation is positive, -1 if the observation is negative, 0 if the observation is neutral and finally divided by the number of reviews in which the aspect was detected, enabling comparison across aspects with different detection frequencies. The final calculation for every aspect, therefore, would be:

$$\text{Confidence weighted index} = \frac{\sum \text{Positive score} - \sum \text{Negative score}}{\text{Number of reviews}}$$

Where *Confidence weighted index* is the resulting per-aspect measure that expresses the average sentiment intensity on a scale from -1 to +1, accounting for how confident the model was in each classification; $\sum \text{Positive score}$ is the sum of the confidence scores belonging to all reviews classified as positive for that aspect; $\sum \text{Negative score}$ is the sum of the confidence scores belonging to all reviews classified as negative for that aspect and, finally, *Number of reviews* is the total count of reviews in which the aspect was extracted, that is, the reviews classified as positive or negative for that aspect combined. The numerator represents therefore the net, confidence-weighted indicator for the aspect, obtained by subtracting the total negative confidence mass from the total positive confidence mass, while the division by the number of allows aspects mentioned with different frequencies to be compared on the same scale.

Table 6: Sentiment polarity distribution by aspect, net sentiment score (sum of positive and negative sentiment proportions) and confidence-weighted index.

Aspect	% Positive	% Negative	Net sentiment	Confidence-weighted index
Resolution	81.0%	19.1%	+61.9	+0.587
Passthrough	70.6%	29.4%	+41.2	+0.356
Controllers	69.7%	28.8%	+40.9	+0.377
Comfort	66.4%	32.8%	+33.6	+0.335
Price	60.7%	37.1%	+23.6	+0.220
Battery	56.1%	36.9%	+19.2	+0.152
Strap	53.3%	42.6%	+10.7	+0.064

The two measures provide precisely the same hierarchy, suggesting that the ordering of aspects might be a property of the underlying data distribution and not

of the scale of the data. All aspects receive a positive score, indicating that, overall, the product fares well on all seven attributes, even if there is substantial variation from an aspect that scores almost unanimously positively, “resolution”, to one that scores almost equally negatively, “strap”.

The intrinsic value of the above numbers can be exposed only by combining them with the data on sentiment salience presented in Section 3.3. The aspects of the product are mapped on a two-dimensional plane according to their scores for salience and net sentiment. Horizontal and vertical lines passing through the averages of cross-aspect scores serve as references. The outcome is a two-by-two matrix that identifies four strategic areas within the product. The primary insight generated by this map is that the aspects that customers talk about the most are also the aspects that they like least, whereas the aspect they appreciate the most gets very little discussion.

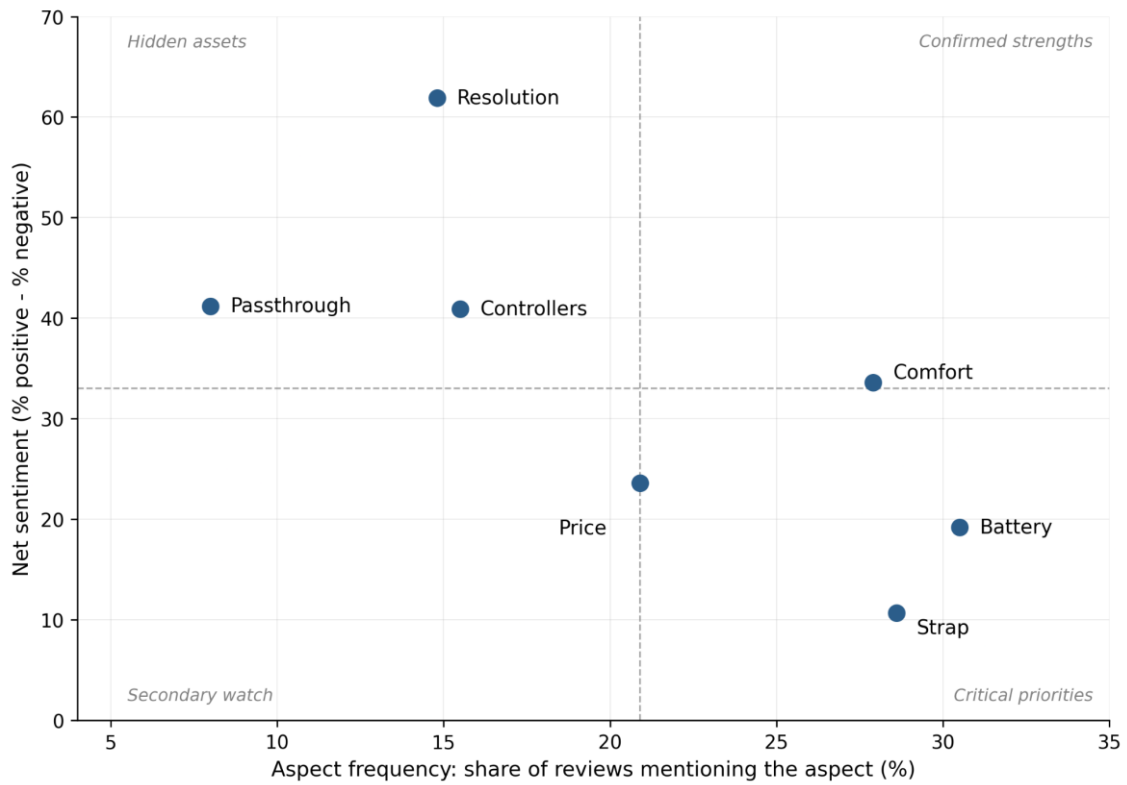


Figure 1¹: Aspect-salience matrix. Dashed lines on the axis indicate cross-aspect means

3.5 Implications for marketing strategy

The salience-sentiment matrix in figure 1 demonstrates that there are several key points.

To start, battery and strap are the two most common aspects with the lowest net sentiment, which makes them major drivers of the online product reputation, potentially influencing a customer's decision to purchase the product. In this case, the net sentiment is still positive, but as the lowest scoring aspects, this insight should be addressed with different possible recommendations to explore, such as developing internally a more comfortable head strap or a better performing battery,

¹ Image elaborated through python. Find the code in the appendix for the code.

as well as seeking such things through partnerships with new supplier. The marketing team should be responsible of communicating clearly the product feature through their advertising and distribution channels, so that purchasing customers' expectations are always met.

Another insight is that the resolution aspect scores the highest both in net sentiment and confidence-weighted index, making it the most valuable asset among the product's features. Strategically speaking, this information is a clear indicator that the marketing team should at least test pushing this aspect more when it comes to advertisements. A similar recommendation can be made for the passthrough and controller aspect, whose net sentiment and salience values are slightly lower.

The comfort aspect, instead, sits just above strap and battery, suggesting that beside the head strap, the product is perceived as comfortable.

Lastly, the price aspect deserves its own comment as it lies around the middle of the matrix on both dimensions (salience 20.9%, net sentiment +23.6). In the previous tables, it is considered as an aggregate of markets that show a greater divergence of views on prices than anything else. In fact, price is a very peculiar dimension for a product, especially when marketed with an international perspective.

So, price serves as perhaps the best example in this study of how a drill-down approach to data can help achieve better insights.

Table 7: price sentiment by market

Market	Positive	Negative	Net sentiment	Mean positive score	Mean negative score
Italy	20	6	+53.8	0.841	0.691
Germany	8	13	-21.7	0.906	0.788
France	2	3	-20.0	0.759	0.844
Spain	4	0	+100.0	0.793	-
United States	12	10	+9.1	0.872	0.738
Australia	8	1	+77.8	0.714	0.994
Overall	54	33	+23.6	0.832	0.767

The numbers for Spain and France have only been presented for the sake of completeness and no conclusions can be made from them based on their sample

size. The true contrasts in this regard exist in the two markets where a sufficient number of participants can be analyzed. While the Italian market perceives the price of the product positively, Germany perceives the price negatively. The net-sentiment discrepancy between Germany and Italy is over seventy-five.

Possible explanations may be related to several factors, such as the presence of local competitive alternatives, a low perceived product value compared to the price and market-based adoption barriers. In fact, it is possible that the German market's consumers may not be technologically mature enough for such products and therefore consider it too expensive to buy. Establishing which factor is more important or accurate would require a separate market study, which is clearly an important outcome for the marketing strategy and intelligence team.

Finally, the divided American sentiment implies that value communication should be experimentally tested as a marketing tactic, possibly through strategic pricing initiatives such as personalized discounts, up-sells and cross-sells. For instance, the company could run an A/B test in which one group of American customers sees an email offering a discounted battery head strap bundled with their purchase, while another group sees no offer at all, in order to check whether the bundle actually makes customers feel the price is more justified. Another example could involve testing two different advertisement messages, one selling the product by itself and another selling it with a free trial period of a software subscription, like how video or music streaming services such as Netflix or Spotify offer a trial before charging, to see which message makes American customers view the price more positively.

Conclusions

This work began with a clear goal: understating whether Aspect Based Sentiment Analysis of customer reviews could enhance the marketing strategy of the virtual reality headset Meta Quest 3. This goal was pursued in three different steps.

In first place, we defined where this technique sits in the world of machine learning, describing the most widely used methods in marketing, such as decision trees; then transitioning into natural language processing tasks and the other text analytics techniques. The literature review finally narrowed down to the model used in the analysis: the DeBERTa, which is essentially a pre-trained model meant to understand text built on top of a specific neural network architecture called transformer.

Secondly, we traced the methodological boundaries of this work, explaining the choices that guided this thesis in terms of data collection and processing and highlighting potential limitations.

Finally, we presented the empirical findings that stem from the data. The evidence can be summarized through four key takeaways.

Firstly, the product is overall well perceived under every predefined aspect, with net sentiment scores and confidence-weighted scores always positive.

The second takeaway concerns the aspects “strap” and “battery”. They represent product priorities that appear to be, for a selected portion of users, either miscommunicated and / or poorly designed. The emerging suggestions to explore

are improving strap design, battery duration and communication strategy accordingly.

Another important point is that there is a credible price perception discrepancy between Italy and Germany, two countries with apparently similar customer buyer persona.

Lastly, the resolution and passthrough aspects are under-communicated, yet low-risk assets for future campaigns.

In none of these cases was the piece of information known prior to the analysis being conducted and each finding can be mapped clearly to the various marketing functions involved in most organizations: product development, pricing and commercial strategy, communication and campaign design, meeting the anticipated multi-domain nature of the impact of the analysis.

In none of these cases was the piece of information known prior to the analysis being conducted and each finding can be mapped clearly to the various marketing functions involved in most organizations: product development, pricing and commercial strategy, communication and campaign design, meeting the anticipated multi-domain nature of the impact of the analysis.

Therefore, the answer provided by the empirical evidence must be affirmative: aspect-based sentiment analysis of customer reviews is an enhancer of marketing strategy as it provides structured information that can be aggregated at different levels in order make more informed decision within marketing-related business units.

Managerial implications

Beyond its contribution to academic literature, this study shows empirical value for marketing practitioners operating within product organizations, especially those in the consumer electronics industry. It shows, for instance, that Aspect Based Sentiment Analysis can convert an otherwise unread volume of customer reviews into a standardized and structured information at a fraction of the time and cost of traditional market research methods such as focus groups, commissioned surveys and manual customer review analysis.

Second, with their existing knowledge base, companies can leverage ABSA to go further than merely identifying what product features are going well and which ones are perform badly: they can start from the model output to build new marketing initiatives and to collaborate with internal stakeholders to improve their product, as shown in the previous chapter.

Third, because the method relies on a pre-trained model applied out of the box, it is realistically replicable by marketing or product teams that have access to customer reviews data and no need for in-house data-intensive infrastructure, which lowers the practical barrier to adoption to the acquisition of a set of data analytics skills for the company employee.

Taken together, these implications suggest that ABSA is not merely an academic curiosity but a deployable tool for continuous, low-cost customer listening that can inform decisions across product, pricing and communication functions simultaneously.

Theoretical contribution and future research prospects

To this date, very little has been done regarding the application of the neural network architecture used to power Large Language Models such as the DeBERTa (i.e. transformers), to marketing-related tasks. According to the systematic review carried out by Petratos and Giannoula (2024) via PRISMA screening of 481 studies, only eight publications employed transformers for the joint purpose of sentiment analysis and marketing. Notably, according to the same source, transformer models clearly outperform more traditional methods like Convolutional Neural Networks (CNNs), Long Short-Term Memories (LSTMs) and Recurrent Neural Networks (RNNs) in their capacity to classify sentiments, implying that the limited adoption rate is likely attributable to the novelty of transformers, rather than any deficiencies on their part. At the same time, an important fact remains: there is a gap within the research field in question and this study will contribute to addressing it.

Firstly, the existing literature on transformer-based sentiment analysis has a predominantly technical nature. The focus of most papers in the selected pool of studies is on choosing models that best fit a particular task according to specific evaluation metrics. Moreover, Petratos and Giannoula (2024) explicitly note that transformer-based AI tools “has not been extensively applied yet to the core of marketing and business strategy” and encourage scholars to explore areas such as pricing and promotions, scopes in which transformer-based research is still scarce, at best. To that end, this study responds to that encouragement by applying an exploratory approach, in which DeBERTa serves as a tool for obtaining business insights rather than being the subject of performance evaluation itself. Thus, the goal of this study is not to benchmark algorithmic performance, but rather assess its applicability and suggest some concrete marketing strategies.

The second gap to be addressed lies in the scope of segments to which the technology can be applied. As Petratos and Giannoula (2024) show, five of the eight papers deal exclusively with product management in the form of review analysis, leaving only three papers to link that review analysis to marketing strategy and brand management or advertising. Marketing segments that have not been covered yet include pricing strategy, communication plan development and customer experience design. The latter aspect becomes possible because the analyzed segments are represented by certain aspects (such as price, comfort or resolution) that can be attributed to specific marketing decision-making domains, hence creating opportunities for simultaneous analysis of several marketing functions.

A third gap concerns the scope of product categories covered by previous research. As the review shows, most studies on transformers in marketing deal with consumer electronics and specifically with mature high-volume categories such as televisions and general assortment from e-commerce websites. Virtual reality headsets make up a relatively new product category in which hardware performance and experiential characteristics are intertwined. Relevant aspects of VR satisfaction (like passthrough quality, spatial comfort or controller responsiveness) cannot be easily mapped onto the taxonomies used for other types of electronics. Therefore, the proposed research constitutes both a necessity and an innovation for this purpose.

Lastly, it should be noted that none of the eight studies identified by Petratos and Giannoula (2024) have a cross-market component to them; they all use data from a single language and market context. In that regard, this study makes a unique contribution by analyzing sentiment data from six different countries that belong to two linguistic clusters, increasing the similarity to a concrete, real world, use

case.

To conclude, there are several future research prospects that may complement this study and address the limitations mentioned in the second chapter. These limitations regard mainly this specific study rather than ABSA in general and they are related to the selection bias and to the poor data availability.

One of the possible future studies could include pairing ABSA with unsupervised aspect discovery methods, which would allow to craft a better aspect list before diving into the data through ABSA. It would also be interesting to test such models' performance on the natively multilingual datasets to evaluate how much performance degradation is caused by the fact that the majority of non-English reviews had to be translated prior to processing, especially since, with the current model, the aspect extraction rates are noticeably lower for translated texts (as demonstrated in Section 2.2, indicating that multilingual ABSA enabled by translation may be an area worth exploring in more detail. Most importantly, further research should be conducted to evaluate the marketing or product KPIs in a context with ABSA against similar contexts without ABSA, to see if the companies using this type of technique actually scores better in pre-defined benchmarks.

References

Roszko-Wójtowicz, Elżbieta, et al. "Innovation-driven e-commerce growth in the EU: An empirical study of the propensity for online purchases and sustainable consumption." *Sustainability* 16.4 (2024): 1563.

Rydning, David Reinsel-John Gantz-John, John Reinsel, and John Gantz. "The digitization of the world from edge to core." Framingham: International Data Corporation 16 (2018): 1-28.

Halevy, Alon, Peter Norvig, and Fernando Pereira. "The unreasonable effectiveness of data." *IEEE intelligent systems* 24.2 (2009): 8-12.

Mao, Yanying, Qun Liu, and Yu Zhang. "Sentiment analysis methods, applications, and challenges: A systematic literature review." *Journal of King Saud University-Computer and Information Sciences* 36.4 (2024): 102048.

Kühl, Niklas, et al. "Machine learning in artificial intelligence: Towards a common understanding." *arXiv preprint arXiv:2004.04686* (2020).

Simeone, Osvaldo. "A very brief introduction to machine learning with applications to communication systems." *IEEE Transactions on Cognitive Communications and Networking* 4.4 (2018): 648-664.

Samuel, Arthur L. "Some studies in machine learning using the game of checkers." *IBM Journal of research and development* 3.3 (1959): 210-229.

Mitchell, Tom M. "Does machine learning really work?." *AI magazine* 18.3

(1997): 11-11.

Goodfellow, Ian, et al. Deep learning. Vol. 1. No. 2. Cambridge: MIT press, 2016.

Bzdok, Danilo. "Classical statistics and statistical learning in imaging neuroscience." *Frontiers in neuroscience* 11 (2017): 543.

Mazzieri, Marcos, and Eduardo Dantas Soares. "Conceptualization and theorization of the Big Data." *International Journal of Innovation: IJI Journal* 4.2 (2016): 23-41.

McAfee, Andrew, et al. "Big data." *The management revolution. Harvard Bus Rev* 90.10 (2012): 61-67.

Hastie, Trevor, et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. New York: springer, 2009.

James, Gareth, et al. *An introduction to statistical learning: with applications in R*. Vol. 103. New York: springer, 2013.

Sutton, Richard S., and Andrew G. Barto. *Reinforcement learning: An introduction*. Vol. 1. No. 1. Cambridge: MIT press, 1998.

Huang, Ming-Hui, and Roland T. Rust. "A strategic framework for artificial intelligence in marketing." *Journal of the academy of marketing science* 49.1 (2021): 30-50.

Geiler, Louis, Séverine Affeldt, and Mohamed Nadif. "A survey on machine learning methods for churn prediction." *International Journal of Data Science and Analytics* 14.3 (2022): 217-242.

Yoshi, Areyfin Mohammed, et al. "Real-Time Dynamic Pricing Using Machine Learning: Integrating Customer Sentiment and Predictive Models for E-Commerce." *International Journal of Advanced Computer Science & Applications* 16.9 (2025).

Hicham, Nouri, and Sabri Karim. "Analysis of unsupervised machine learning techniques for an efficient customer segmentation using clustering ensemble and spectral clustering." *International Journal of Advanced Computer Science and Applications* 13.10 (2022).

Linden, Greg, Brent Smith, and Jeremy York. "Amazon. com recommendations: Item-to-item collaborative filtering." *IEEE Internet computing* 7.1 (2003): 76-80.

Salmerón-Ríos, Alejandro, et al. "Fine grain emotion analysis in Spanish using linguistic features and transformers." *PeerJ Computer Science* 10 (2024): e1992.

Hartmann, Jochen, and Oded Netzer. "Natural language processing in marketing." (2023).

Herhausen, Dennis, et al. "Machine learning in marketing: Recent progress and future research directions." *Journal of Business Research* 170 (2024): 114254.

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1 (1996): 267-288.

Nelder, John Ashworth, and Robert WM Wedderburn. "Generalized linear models." *Journal of the Royal Statistical Society Series A: Statistics in Society* 135.3 (1972): 370-384.

Salih, A. M., and Y. Wang. "Are Linear Regression Models White Box and

Interpretable? arXiv 2024." arXiv preprint arXiv:2407.12177.

Dervovic, Danial, et al. "Are logistic models really interpretable?." arXiv preprint arXiv:2406.13427 (2024).

Qader, Khowanas Saeed, et al. "Analyzing different types of advertising and its influence on customer choice." *International Journal of Humanities and Education Development (IJHED)* 4.6 (2022): 8-21.

Jemal, Muluka Kedir, and Kassegn Berhanu Melese. "The impact of advertising on consumers' buying behavior: the case of Safaricom Ethiopia, Addis Ababa, Ethiopia." *Cogent Social Sciences* 11.1 (2025): 2454355.

Ali, Nafez, and Osama Samih Shabn. "Customer lifetime value (CLV) insights for strategic marketing success and its impact on organizational financial performance." *Cogent Business & Management* 11.1 (2024): 2361321.

Effendi, Muhammad Junaid, and Syed Abbas Ali. "Click through rate prediction for contextual advertisement using linear regression." arXiv preprint arXiv:1701.08744 (2017).

Dong, Jian, et al. "Potential trend for online shopping data based on the linear regression and sentiment analysis." *Mathematical Problems in Engineering* 2020.1 (2020): 4591260.

Constantin, Cristinel. "Using the Logistic Regression model in supporting decisions of establishing marketing strategies." *Bulletin of the Transilvania University of Brasov. Series V: Economic Sciences* (2015): 43-50.

Chajia, Meryem, and El Habib Nfaoui. "Customer churn prediction approach

based on llm embeddings and logistic regression." *Future Internet* 16.12 (2024): 453.

Nie, Guangli, et al. "Credit card churn forecasting by logistic regression and decision tree." *Expert Systems with Applications* 38.12 (2011): 15273-15285.

Li, ZhuoJun. "Accurate digital marketing communication based on intelligent data analysis." *Scientific Programming* 2022.1 (2022): 8294891.

Chen, Nie. "Research on E-Commerce Database Marketing Based on Machine Learning Algorithm." *Computational Intelligence and Neuroscience* 2022.1 (2022): 7973446.

Kingsford, Carl, and Steven L. Salzberg. "What are decision trees?." *Nature biotechnology* 26.9 (2008): 1011-1013.

Blockeel, Hendrik, et al. "Decision trees: from efficient prediction to responsible AI." *Frontiers in artificial intelligence* 6 (2023): 1124553.

Ying, L. U. "Decision tree methods: applications for classification and prediction." *Shanghai archives of psychiatry* 27.2 (2015): 130.

Amro, Asma, et al. "Instance reduction for avoiding overfitting in decision trees." *Journal of Intelligent Systems* 30.1 (2021): 438-459.

Opitz, David, and Richard Maclin. "Popular ensemble methods: An empirical study." *Journal of artificial intelligence research* 11 (1999): 169-198.

Breiman, Leo. "Bagging predictors." *Machine learning* 24.2 (1996): 123-140.

Yanni, Meri Hari, Khairil Anwar Notodiputro, and Bagus Sartono. "Performance

Comparison of Random Forest, Bagging, and CART Methods in Classifying Recipients of the Family Program in North Aceh." *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika* 11.1 (2025): 50-58.

Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.

Freund, Yoav, Robert Schapire, and Naoki Abe. "A short introduction to boosting." *Journal-Japanese Society For Artificial Intelligence* 14.771-780 (1999): 1612.

Friedman, Jerome H. "Greedy function approximation: a gradient boosting machine." *Annals of statistics* (2001): 1189-1232.

Tirenni, Giuliano, Christian Kaiser, and Andreas Herrmann. "Applying decision trees for value-based customer relations management: Predicting airline customers' future values." *Journal of Database Marketing & Customer Strategy Management* 14.2 (2007): 130-142.

Liu, Yi, and Shuo Yang. "Application of Decision Tree-Based Classification Algorithm on Content Marketing." *Journal of Mathematics* 2022.1 (2022): 6469054.

Hayadi, B. Herawan, and Ibrahiem MM El Emary. "Predicting campaign ROI using decision trees and random forests in digital marketing." *Journal of Digital Market and Digital Currency* 1.1 (2024): 1-20.

Gülter, Erkan, and Muhammed Fatih Cevher. "Evolution of digital marketing campaigns with artificial intelligence and machine learning: Analysing success prediction capabilities." *Business & Management Studies: An International Journal* 13.2 (2025): 478-493.

Singh, Ravinder. "Predicting Airline Ticket Prices Using Decision Trees and Random Forests: A Comparative Analysis." 2025 7th International Conference on Energy, Power and Environment (ICEPE). IEEE, 2025.

Saravanan, K., et al. "Support vector machines: Unveiling the power and versatility of svms in modern machine learning." 2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 2023.

Guido, Rosita, et al. "An overview on the advancements of support vector machine models in healthcare applications: a review." Information 15.4 (2024): 235.

Tanveer, M., et al. "Enhancing imbalance learning: A novel slack-factor fuzzy SVM approach." IEEE Transactions on Emerging Topics in Computational Intelligence 9.4 (2025): 3112-3121.

Tanveer, Mohammad, et al. "Comprehensive review on twin support vector machines." Annals of Operations Research 339.3 (2024): 1223-1268.

Rezvani, Salim, Xizhao Wang, and Farhad Pourpanah. "Intuitionistic fuzzy twin support vector machines." IEEE Transactions on Fuzzy Systems 27.11 (2019): 2140-2151.

Dhar, Sauprik, and Vladimir Cherkassky. "Single class universum-svm." arXiv preprint arXiv:1909.09862 (2019).

Sanz, Hector, et al. "SVM-RFE: selection and visualization of the most relevant features through non-linear kernels." BMC bioinformatics 19.1 (2018): 432.

Zhao, BenYu, and Jingqian Hu. "Precision Consumer Segmentation and Marketing Strategy Optimization Using Support Vector Machines (SVM) Combined with

Particle Swarm Optimization (PSO)." Proceedings of the 2025 2nd International Conference on Big Data and Digital Management. 2025.

Xiahou, Xiancheng, and Yoshio Harada. "B2C E-commerce customer churn prediction based on K-means and SVM." Journal of Theoretical and Applied Electronic Commerce Research 17.2 (2022): 458-475.

Yuan, Yuan. "Analysis and Research of Retail Enterprise Customer Consumption Behavior Based on Support Vector Machine Algorithm." Proceedings of the 2025 International Conference on Economic Management and Big Data Application. 2025.

Rogić, Sunčica, Ljiljana Kaščelan, and Mirjana Pejić Bach. "Customer response model in direct marketing: solving the problem of unbalanced dataset with a balanced support vector machine." Journal of Theoretical and Applied Electronic Commerce Research 17.3 (2022): 1003-1018.

Lian-Ying, Zhou, et al. "Combined appetency and upselling prediction scheme in telecommunication sector using support vector machines." International Journal of Modern Education and Computer Science 10.6 (2019): 1.

Jain, Anil K. "Data clustering: 50 years beyond K-means." Pattern recognition letters 31.8 (2010): 651-666.

Wedel, Michel, and Wagner A. Kamakura. Market segmentation: Conceptual and methodological foundations. Springer Science & Business Media, 2000.

MacQueen, J. "Multivariate observations." Proceedings of the 5th Berkeley symposium on mathematical statistics and probability. Vol. 1. Oakland, CA, USA: University of California press, 1967.

Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the royal statistical society. series c (applied statistics)* 28.1 (1979): 100-108.

Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of computational and applied mathematics* 20 (1987): 53-65.

Murtagh, Fionn, and Pedro Contreras. "Algorithms for hierarchical clustering: an overview." *Wiley interdisciplinary reviews: data mining and knowledge discovery* 2.1 (2012): 86-97.

Ward Jr, Joe H. "Hierarchical grouping to optimize an objective function." *Journal of the American statistical association* 58.301 (1963): 236-244.

Murtagh, Fionn, and Pierre Legendre. "Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?." *Journal of classification* 31.3 (2014): 274-295.

Müller, Andreas C., Sebastian Nowozin, and Christoph H. Lampert. "Information theoretic clustering using minimum spanning trees." *Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

Bouguettaya, Athman, et al. "Efficient agglomerative hierarchical clustering." *Expert Systems with Applications* 42.5 (2015): 2785-2797.

Raymaekers, Jakob, and Ruben H. Zamar. "Pooled variable scaling for cluster analysis." *Bioinformatics* 36.12 (2020): 3849-3855.

Punj, Girish, and David W. Stewart. "Cluster analysis in marketing research: Review and suggestions for application." *Journal of marketing research* 20.2 (1983): 134-148.

Kansal, Tushar, et al. "Customer segmentation using K-means clustering." 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS). IEEE, 2018.

Kumar, Sumit, et al. "Customer segmentation in e-commerce: K-means vs hierarchical clustering." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 23.1 (2025): 119-128.

Pearson, Karl. "LIII. On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11 (1901): 559-572.

Gewers, Felipe L., et al. "Principal component analysis: A natural approach to data exploration." *ACM Computing Surveys (CSUR)* 54.4 (2021): 1-34.

Alkhayrat, Maha, Mohamad Aljnidi, and Kadan Aljoumaa. "A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA." *Journal of Big Data* 7.1 (2020): 9.

Abdulhafedh, Azad. "Incorporating k-means, hierarchical clustering and pca in customer segmentation." *Journal of City and Development* 3.1 (2021): 12-30.

Liu, Hongping. "Big data precision marketing and consumer behavior analysis based on fuzzy clustering and PCA model." *Journal of Intelligent & Fuzzy Systems* 40.4 (2021): 6529-6539.

Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. "Natural language processing: an introduction." *Journal of the American Medical Informatics Association* 18.5 (2011): 544-551.

Khurana, Diksha, et al. "Natural language processing: state of the art, current trends and challenges." *Multimedia tools and applications* 82.3 (2023): 3713-3744.

Humphreys, Ashlee, and Rebecca Jen-Hui Wang. "Automated text analysis for consumer research." *Journal of Consumer Research* 44.6 (2018): 1274-1306.

Berger, Jonah, et al. "Uniting the tribes: Using text for marketing insight." *Journal of marketing* 84.1 (2020): 1-25.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3.Jan (2003): 993-1022.

Blei, David M. "Probabilistic topic models." *Communications of the ACM* 55.4 (2012): 77-84.

Hannigan, Timothy R., et al. "Topic modeling in management research: Rendering new theory from textual data." *Academy of Management Annals* 13.2 (2019): 586-632.

Reisenbichler, Martin, and Thomas Reutterer. "Topic modeling in marketing: recent advances and research opportunities: M. Reisenbichler, T. Reutterer." *Journal of business economics* 89.3 (2019): 327-356.

Guo, Yue, Stuart J. Barnes, and Qiong Jia. "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation." *Tourism management* 59 (2017): 467-483.

Wang, Wenxin, Yi Feng, and Wenqiang Dai. "Topic analysis of online reviews for two competitive products using latent Dirichlet allocation." *Electronic Commerce Research and Applications* 29 (2018): 142-156.

Kang, Min, et al. "A study on the influence of online reviews of new products on consumers' purchase decisions: An empirical study on JD. com." *Frontiers in Psychology* 13 (2022): 983060.

Harris, Zellig S. "Distributional structure." *Word* 10.2-3 (1954): 146-162.

Netzer, Oded, et al. "Mine your own business: Market-structure surveillance through text mining." *Marketing Science* 31.3 (2012): 521-543.

Tirunillai, Seshadri, and Gerard J. Tellis. "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation." *Journal of marketing research* 51.4 (2014): 463-479.

Glance, Natalie, et al. "Deriving marketing intelligence from online discussion." *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. 2005.

Lee, Siyuan. "Social media text sentiment classification based on bag-of-words models with multiple machine learning algorithms." *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*. IEEE, 2024.

Büschken, Joachim, and Greg M. Allenby. "Sentence-based text analysis for customer reviews." *Marketing Science* 35.6 (2016): 953-975.

Turney, Peter. "Thumbs up or thumbs down? Semantic orientation applied to

unsupervised classification of reviews." Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002.

Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004.

Youvan, Douglas C. "Understanding sentiment analysis with VADER: a comprehensive overview and application." AI and Data Science Journal (2024).

Hutto, Clayton, and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Proceedings of the international AAAI conference on web and social media. Vol. 8. No. 1. 2014.

Sally, Sherina. "Sentiment analysis on YouTube smart phone unboxing video reviews in Sri Lanka." arXiv preprint arXiv:2302.03496 (2023).

Pandey, Kamal Kishor, et al. "Natural language processing for sentiment analysis in social media marketing." 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2023.

Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 (2015): 85-117.

Goodfellow, I., Yoshua, B., & Aaron, C. 2016. Introduction to Deep Learning. MIT Press. <https://www.deeplearningbook.org/>

Mansur, Saad, et al. "Sales forecasting for retail stores using hybrid neural networks and sales-affecting variables." PeerJ Computer Science 11 (2025): e3058.

Tsai, Chih-Fong, and Yu-Hsin Lu. "Customer churn prediction by hybrid neural networks." *Expert Systems with Applications* 36.10 (2009): 12547-12553.

Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

Ahmed, Sabeen, et al. "Transformers in time-series analysis: A tutorial." *Circuits, systems, and signal processing* 42.12 (2023): 7433-7466.

Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 2019.

He, Pengcheng, et al. "Deberta: Decoding-enhanced bert with disentangled attention." *arXiv preprint arXiv:2006.03654* (2020).

Timoneda, Joan C., and Sebastián Vallejo Vera. "BERT, RoBERTa, or DeBERTa? Comparing performance across transformers models in political science text." *The Journal of Politics* 87.1 (2025): 347-364.

Cao, Zijian. (2025). Enhancing Sentiment Analysis with Fine-Tuned DeBERTa in a Cross-Domain Study on IMDb and Amazon Polarity Datasets. *Applied and Computational Engineering*. 149. 198-201. 10.54254/2755-2721/2025.KL22625.

Guo, Xinli. "Sentiment analysis based on RoBERTa for amazon review: An empirical study on decision making." *arXiv preprint arXiv:2411.00796* (2024).

Ahlstrand, Jim, et al. "Using Transformers for B2B Contractual Churn Prediction Based on Customer Behavior Data." *ICEIS* (1). 2025.

Xu, Hu, et al. "BERT post-training for review reading comprehension and aspect-based sentiment analysis." Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019.

Perikos, Isidoros, and Athanasios Diamantopoulos. "Explainable aspect-based sentiment analysis using transformer models." Big data and cognitive computing 8.11 (2024): 141.

He, Pengcheng, et al. "Deberta: Decoding-enhanced bert with disentangled attention." arXiv preprint arXiv:2006.03654 (2020).

Schoenmueller, Verena, Oded Netzer, and Florian Stahl. "The polarity of online reviews: Prevalence, drivers and implications." Journal of Marketing Research 57.5 (2020): 853-877.

Yang, Heng, Chen Zhang, and Ke Li. "Pyabsa: A modularized framework for reproducible aspect-based sentiment analysis." Proceedings of the 32nd ACM international conference on information and knowledge management. 2023.

Petratos, Pythagoras, and Mina Giannoula. "Transformer-based AI for Sentiment Analysis in Marketing." Proceedings of the 2024 6th Asia Conference on Machine Learning and Computing. 2024.

Appendix

```
# Set up libraries
!pip install -q transformers pandas torch openpyxl
import pandas as pd
import json
from transformers import pipeline

# Configure the exact model
print("Loading the model... (This relies on your internet connection)")
absa_classifier = pipeline(
    "text-classification",
    model="yangheng/deberta-v3-base-absa-v1.1",
    device=-1 # 0 to use dedicated GPU, otherwise -1
)

aspects_to_analyze = [
    "battery", "comfort", "resolution", "price",
    "passthrough", "controllers", "strap"
]

# Load data
file_name = "reviews_eng.xlsx"
print(f"Attempting to read {file_name}...")

# Safety check: load excel and print the first 2 reviews
try:
    df = pd.read_excel(file_name)
    print("File loaded successfully!")
    print("Preview of first 2 reviews:")
    print(df['review_text'].head(2))
except FileNotFoundError:
    print(f"ERROR: Could not find '{file_name}'.")
    df = None
except Exception as e:
    print(f"An error occurred: {e}")
```

```

df = None

# Stop if file didn't load
if df is None:
    raise SystemExit("Stopping execution: No file loaded.")

# Convert column to text, handling any empty cells gracefully
reviews = df['review_text'].fillna("").astype(str).tolist()

# Running the analysis
results = []
print(f"\nStarting analysis on {len(reviews)} reviews...")

for i, review in enumerate(reviews):
    row_result = {"review_text": review}

    # Skip empty reviews to save time
    if len(review) < 5:
        results.append(row_result)
        continue

    for aspect in aspects_to_analyze:

        # Check if the aspect exists in the review text
        if aspect in review.lower():

            try:
                # Ask the model for sentiment
                output = absa_classifier(review, text_pair=aspect)
                row_result[f"{aspect}_sentiment"] = output[0]['label']
                row_result[f"{aspect}_score"] = output[0]['score']

            except Exception as e:
                # If text is too long, just skip this aspect
                print(f"Skipped an aspect due to error: {e}")
                row_result[f"{aspect}_sentiment"] = None
                row_result[f"{aspect}_score"] = None

```

```

        else:
            row_result[f"{aspect}_sentiment"] = None
            row_result[f"{aspect}_score"] = None

        results.append(row_result)

    # Progress tracker
    if (i + 1) % 10 == 0:
        print(f"Processed {i + 1}/{len(reviews)} reviews...")

# Save back to Excel
output_filename = "absa_results_quest3_FINAL.xlsx"

df_results = pd.DataFrame(results)

df_results.to_excel(output_filename, index=False)

print(f"DONE! Results saved to: {output_filename}")
print("Check the Files tab on the left to download it.")

```

```

import sys
import pandas as pd
import numpy as np
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt

# configuration

ASPECTS = ["battery", "comfort", "resolution",
           "price", "passthrough", "controllers", "strap"]
INPUT_FILE = sys.argv[1] if len(sys.argv) > 1 else "absa_results.xlsx"
SHEET = "sentiment"
OUTPUT_FIG = "frequency_sentiment_matrix.png"

# Label-nudge offsets so point labels don't overlap markers (x, y in axis units)
LABEL_OFFSETS = {
    "Battery": (0.6, 0.0), "Strap": (0.6, -1.5), "Comfort": (0.6, 1.0),
    "Price": (-1.0, -4.5), "Controllers": (0.6, 0.0),
    "Resolution": (0.6, 0.0), "Passthrough": (0.6, 0.0),
}

```

```

# metric computation
def compute_metrics(df):
    """Return a DataFrame with frequency, net sentiment and the
    confidence-weighted index for each aspect."""
    n_reviews = len(df)
    rows = []

    for aspect in ASPECTS:

        sentiment = df[f"{aspect}_sentiment"]
        score = df[f"{aspect}_score"]
        extracted = sentiment.notna()
        n_extracted = int(extracted.sum())
        pos = int((sentiment == "Positive").sum())
        neg = int((sentiment == "Negative").sum())
        neu = int((sentiment == "Neutral").sum())

        # Frequency: share of ALL reviews that mention the aspect
        frequency = n_extracted / n_reviews * 100

        # Net sentiment: (% positive - % negative) among extracted reviews
        net_sentiment = (pos - neg) / n_extracted * 100

        # Confidence-weighted index
        signed = np.where(sentiment == "Positive", score,
                          np.where(sentiment == "Negative", -score, 0.0))
        weighted_index = signed[extracted].mean()
        rows.append({
            "aspect": aspect.capitalize(),
            "n_extracted": n_extracted,
            "frequency_pct": round(frequency, 1),
            "pct_positive": round(pos / n_extracted * 100, 1),
            "pct_negative": round(neg / n_extracted * 100, 1),
            "pct_neutral": round(neu / n_extracted * 100, 1),
            "net_sentiment": round(net_sentiment, 1),
            "weighted_index": round(weighted_index, 3),
        })
    return pd.DataFrame(rows)

# plotting
def plot_matrix(metrics, output_path):
    """Scatter aspects by frequency (x) and net sentiment (y), with quadrant
    reference lines at the cross-aspect means."""
    x = metrics["frequency_pct"].values
    y = metrics["net_sentiment"].values
    names = metrics["aspect"].values
    mean_x = x.mean()
    mean_y = y.mean()
    fig, ax = plt.subplots(figsize=(8.5, 6), dpi=300)

    # Quadrant reference lines
    ax.axvline(mean_x, color="gray", linestyle="--", linewidth=0.9, alpha=0.7)
    ax.axhline(mean_y, color="gray", linestyle="--", linewidth=0.9, alpha=0.7)

    # Points and labels

```

```

for name, xi, yi in zip(names, x, y):
    ax.scatter(xi, yi, s=70, color="#2b5d8a", zorder=3)
    dx, dy = LABEL_OFFSETS.get(name, (0.6, 0.0))
    ha = "right" if name == "Price" else "left"
    ax.annotate(name, (xi, yi), xytext=(xi + dx, yi + dy),
                fontsize=10.5, ha=ha, va="center")

# Quadrant captions (positioned relative to axis limits)
ax.text(34.5, 67, "Confirmed strengths", fontsize=9, style="italic",
        color="gray", ha="right")
ax.text(34.5, 2, "Critical priorities", fontsize=9, style="italic",
        color="gray", ha="right")
ax.text(5.5, 67, "Hidden assets", fontsize=9, style="italic",
        color="gray", ha="left")
ax.text(5.5, 2, "Secondary watch", fontsize=9, style="italic",
        color="gray", ha="left")
ax.set_xlabel("Aspect frequency: share of reviews mentioning the aspect
(%)", fontsize=11)
ax.set_ylabel("Net sentiment (% positive - % negative)", fontsize=11)
ax.set_xlim(4, 35)
ax.set_ylim(0, 70)
ax.grid(True, linewidth=0.4, alpha=0.35)

for spine in ["top", "right"]:
    ax.spines[spine].set_visible(False)
plt.tight_layout()
plt.savefig(output_path, dpi=300, bbox_inches="tight")
print(f"Figure saved to {output_path}")

# main
def main():
    df = pd.read_excel(INPUT_FILE, sheet_name=SHEET)
    metrics = compute_metrics(df)
    pd.set_option("display.width", 120)
    print(metrics.to_string(index=False))
    print()
    plot_matrix(metrics, OUTPUT_FIG)

if __name__ == "__main__":
    main()

```