PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Acta Informatica Pragensia
2024, Volume 13, Issue 1, pp. 100–113
https://doi.org/10.18267/j.aip.231

Article Open Access

Understanding Consumer Perceptions About Smartwatches: Feature Extraction and Opinion Mining Using Supervised Learning Algorithm

Dhanya Manayath [1], Sanju Kaladharan [1], Nikita Venal Soman [1], Abith Vijayakumaran [2]

[1] Amrita School of Business, Amrita Vishwa Vidyapeetham (Deemed to be University), Kochi, India
[2] Conestoga College Institute of Technology and Advanced Learning, Kitchener, Canada

Corresponding author: Dhanya Manayath (dhanyamrajeev@gmail.com)

Abstract

Against the backdrop of increasing smartwatch usage and the dynamic landscape of evolving features, a nuanced understanding of consumer opinions and preferences is vital for tailoring features and crafting effective marketing strategies. This study addresses this imperative by conducting a comprehensive analysis of customer reviews on smartwatches, aiming to determine the pivotal factors guiding consumer purchasing decisions. By employing word clouds to visually represent sentiments, the study uncovers notable trends. Positive reviews prominently highlight the term "quality", suggesting a strong emphasis on product excellence. In contrast, negative reviews were characterized by the prevalence of the term "fake", indicating concerns related to authenticity. Additionally, a comparative assessment of two machine learning algorithms, namely support vector machines and Naive Bayes, demonstrates that support vector machines exhibit superior accuracy in classification. These findings offer valuable insights for industry practitioners navigating the competitive landscape of the smartwatch market, providing actionable information for optimizing product features and refining marketing strategies to meet consumer expectations.

Keywords

Healthcare 5.0; Sentiment analysis; Opinion mining; Supervised learning algorithms; Wearable technology.

Citation: Manayath, D., Kaladharan. S., Venal, N. S., & Vijayakumaran, A. (2024). Understanding Consumer Perceptions About Smartwatches: Feature Extraction and Opinion Mining Using Supervised Learning Algorithm. *Acta Informatica Pragensia*, 13(1), 100–113. https://doi.org/10.18267/j.aip.231

Academic Editor: Zdenek Smutny, Prague University of Economics and Business, Czech Republic

1 Introduction

Healthcare 5.0 is gaining momentum in the integration of various technologies of industry 5.0 to deliver more personalized healthcare. Integration of novel technologies such as artificial intelligence, the internet of things (IoT) and blockchain ushers healthcare to emerge into various paradigms of patient-centric care in contrast to the traditional hospital-centric approach. These enabling technologies help in effortless information navigation and interconnect organizations, healthcare professionals and other resources to react effectively with intelligent solutions. Various scenarios of Healthcare 5.0 include remote monitoring, assisted living, telemedicine, smart self-management, etc. (Mbunge et al., 2021; Saraswat et al., 2022).

Wearable technology advancements are evolving rapidly in recent years. Wearable devices increase the effectiveness of tracking risk patients at hospitals by allowing doctors to access information anytime and anywhere. The real-time data generated, along with the intelligent data analysis tools, facilitate precise and collective decision-making. The majority of wearable devices constantly remain close to human bodies, providing direct and automatic interaction with the human body without requiring much manual input. Wearables such as smartwatches, smartglasses and wristbands have been used massively for health monitoring and in daily life. The wearable shipment volume is projected to reach 637.1 million units in 2024 as per the reports of the International Data Corporation, and it predicts a five-year compound annual growth rate of 12.4%, indicating the increasing usage trend of smart wearables (Jin et al., 2022). The amount of data produced and processed by each wearable device is huge. By analysing large numbers of real-time and historical data points, these devices and connected networks can produce valuable health analytics and offer feedback personalized to users.

A smartwatch is a personal computing device that has been used by an enormous number of people. It is a combination of a wristwatch and the features of a smartphone. The smartwatch market is projected to reach USD 198.90 million by 2026. Smartwatches have evolved to become a key player in the biomedical sector, offering a range of features, including internet access, calling and health and fitness tracking (Chandel et al., 2022). They are equipped with sensors that can detect various activities, including sleep and can be used for non-obtrusive interaction in ambient environments (Bieber et al., 2013). In the health and wellness domain, smartwatches have been used for activity and heart rate monitoring, speech therapy adherence, diabetes self-management and detection of seizures, tremors, scratching, eating and medication-taking behaviour (Reeder & David, 2016).

Against the backdrop of increasing smartwatch usage and ever-evolving features, it is important to understand consumer opinions and preferences. This can be of great importance for smartwatch companies to customize the features and frame marketing strategies. Although studies demonstrate an increasing adoption of smartwatches in the health and fitness domains, little research has focused on customers' opinions and preferences. It creates a blind spot for smartwatch companies in terms of understanding what features resonate with customers and how to frame the marketing campaign effectively.

Consumer review analysis is one of the fastest and most efficient ways of understanding customers' satisfaction and dissatisfaction with the product. Classification of reviews into positive and negative feelings is required to obtain valuable insights from a large set of reviews. Sentiment analysis using a Naive Bayes and support vector machine (SVM) classifier is used to classify the text into positive, negative or neutral, and a comparison is made to find which classifier performs better. This study mainly focuses on performing text mining using a supervised learning algorithm for analysing customer reviews. Sentiment classification is supervised learning carried out on text-based data to extract the feelings of a given keyword in the text (Basha & Rajput, 2019). Using powerful techniques such as support vector machines, researchers have achieved impressive consistency in extracting emotional intent from textual data (Arpaci et al., 2021). This research focuses on developing models for sentiment classification using machine learning algorithms. Furthermore, to unveil recurring themes and key aspects driving customer sentiment, this study utilizes word clouds, visually showcasing the most frequently used words in smartwatch reviews, leading to the following research questions:

RQ1: What is the overall sentiment of customer reviews for smartwatches?

RQ2: What are the most important features for customers when purchasing a smartwatch?

RQ3: What is the effectiveness of different machine learning algorithms in classifying sentiment (positive, negative, neutral) in textual data?

The paper is organized as follows. The next section provides a comprehensive review of relevant research papers, which discusses literature in three themes: consumer perceptions, the concept of sentiment analysis and Naive Bayes algorithms. The subsequent section discusses methodology, followed by sections presenting results and discussion. The paper is concluded by explaining the contributions and implications.

2 Literature Review

The literature survey was done in three dimensions, namely on works relating to analysis of consumer perceptions about smartwatches, sentiment analysis and Naive Bayes algorithms. The search was carried out in the Scopus database and Google Scholar using specific keywords such as "smartwatch adoption", "consumer perceptions", or "sentiment analysis".

2.1 Consumer perceptions about smartwatches

Given that the smartwatch represents a novel technological advancement, numerous research endeavours have been undertaken to explore consumer perceptions of smartwatches using various theoretical frameworks, including the technology acceptance model (TAM), the task technology fit model, the unified theory of acceptance and use of technology, the expectation confirmation model (ECM) and remediation theory (Bölen, 2020). Investigations have concentrated on factors such as display features, motion state and gender distinctions in smartwatch interactions (Zhang & Rau, 2015). Additionally, a subset of studies has explored the sustained use of smartwatches from a behavioural standpoint. Furthermore, there is a body of research that scrutinizes smartwatch adoption through the lens of technical attributes such as display size or shape (Jung et al., 2016). Consumer perceptions of smartwatches are generally positive, with a strong purchase intention among those who are aware of the product (Kumar, 2017). The value of smartwatches is primarily determined by their functions, particularly display shape and standalone communication (Jung et al., 2016). The categorization of smartwatches as either wristwatches or wearable devices can influence consumer evaluations, with a preference for the latter (Park et al., 2016). However, there are limitations to the current design of smartwatches, with users perceiving them more as functional sensors than as watches or smartphones (Ha et al., 2017). Table 1 summarizes a few studies which have analysed consumer perceptions about smartwatches.

Table 1. *Summary of literature on consumer perceptions about smartwatches.*

Reference	Study topic	Method	Key findings
(Almheiri et al., 2024)	Drivers of smartwatch use	Survey	Factors influencing smartwatch use: performance expectancy, social influence, price value, habit and perceived security. Use of smartwatches has a significant influence on environmental sustainability.
(Xu et al., 2024)	Smartwatch comfort	In-depth interviews	Physical attributes such as size and material and functionalities such as interoperability and automation

Reference	Study topic	Method	Key findings
			capabilities can directly influence smartwatch comfort experience.
(Misra et al., 2023)	Intention to adopt wearables	Survey	Fashionability, ease of use, connectivity and healthcare influence the intention to adopt wearables.
(Almarzouqi et al., 2022)	Intention to use	Survey	Factors affecting intention to use: perceived ease of use, user satisfaction, perceived usefulness, flow experience.
(Elnagar et al., 2022)	Intention to use	Survey	Innovativeness and content richness can enhance perceived usefulness.
(Salahuddin & Lee, 2021)	Key quality features for wearables	Survey	Key quality features: performance, level of durability, product safety, long battery life, usability, comfortability and reasonable price.
(Siepmann & Kowalczuk, 2021)	Continued smartwatch usage	Survey	Device annoyance is an important barrier to continuous smartwatch use.
(Al-Maroof et al., 2021)	User acceptance for medical purposes	Survey	Content richness positively affects smartwatch adoption.
(Jung et al., 2016)	Consumer valuation	Survey-based conjoint analysis	Display shape and standalone communication outweigh brand and price in smartwatch choices. The preference leans towards a curved display shape. Users prioritize smartwatch functions over considerations of brand and price.
(Ha et al., 2017)	User perceptions	Dynamic topic modelling	Users view the smartwatch as a collection of functional sensors rather than a traditional watch or smartphone.
(Park et al., 2016)	Consumers' perceived fit	Experimental design	Frame category and manufacturer play crucial roles in determining consumers' perceived fit.
(Chuah et al., 2016)	Adoption	Survey	Most important factors affecting adoption intention: perceived usefulness and visibility.
(Kim & Shin, 2015)	Adoption	Survey	Affective quality and relative advantage of smartwatches were associated with perceived usefulness.

2.2 Sentiment analysis

Sentiment analysis, otherwise known as opinion mining (Srinivasan & Subalalitha, 2023) or emotion AI (Tandon & Mehra, 2023), identifies the emotions present in a given text. While Nasukawa & Yi (2003)are credited with coining the term "sentiment analysis", research into sentiment actually predates the year 2000 within the field of linguistics (Liu, 2012; Tandon & Mehra, 2023). Initially, sentiment analysis was considered a highly challenging task in natural language processing (NLP) (Tandon & Mehra, 2023), but has later expanded its applications such as automatically gathering product reviews, predicting prices and analysing product feedbacks. The main objective of performing sentiment analysis is to extract opinions and present information, which is summarized based on the polarity of opinions (negative, positive, neutral) (Cambria & White, 2014). An opinion can be defined as a positive or negative sentiment, view, attitude, emotion or appraisal about an entity (product, person, event, organization or topic) or an aspect of that entity from a user or group of users. Generally, there are three types of opinions: (i) regular opinions, which demonstrate good or bad views about a particular product and pertain to a single entity; (ii) comparative opinions, which compare two or more entities and can be understood with respect to some common features; and (iii) suggestive opinions, which provide suggestions about single or multiple entities (Qazi et al., 2017).

Consumer sentiment analysis (CSA) has recently gained significant traction in social media applications, spanning diverse fields such as healthcare, finance, travel and even academia (Jain et al., 2021). It utilizes online reviews, including product/service reviews, tweets and online surveys, to extract the underlying sentiment — positive, negative or neutral — from the text. Organizations can harness the power of CSA to collect valuable insights from previous customer experiences, allowing them to identify areas for improvement in their services or products (Chaturvedi et al., 2018). Additionally, CSA helps identify potential issues in newly released products based on real-time consumer feedback. This enables organizations to gauge customer satisfaction with their offerings and compare themselves to competitors, ultimately leading to better-informed decision-making. For new consumers, CSA serves as a valuable tool to gain insights and knowledge about products or services before making purchase decisions. Despite its wide applications, sentiment analysis has challenges such as context dependence, computational costs, the need for subjectivity detection and accuracy detection (Jain et al., 2021).

2.3 Naive Bayes algorithm

The approaches to performing sentiment analysis can be classified broadly into two. The first group is lexicon-based, comprising an unsupervised learning approach that relies on pre-defined sentiment lexicons to determine the polarity of a text (positive, negative or neutral). One of the key advantages of lexicon-based approaches is their independence of training data and reference corpora. They primarily capture generic sentiment polarity and often ignore the context and domain-specific nuances of language. The other group is machine learning approaches, belonging to the category of supervised learning. This means that they require a training process where a model is "taught" to classify input data (text) into specific output categories (sentiment labels). This training involves feeding the model a large amount of labelled data, where each piece of text is associated with a known sentiment value (positive, negative or neutral). By learning from such labelled data, the model develops the ability to identify patterns and relationships between words and their associated sentiments (Chamlertwat et al., 2012; Jain & Dandannavar, 2016; Srinivasan & Subalalitha, 2023) The ever-growing volume, subjectivity and heterogeneity pose significant challenges to manual processing. To address this, machine learning (ML) techniques have emerged as powerful tools, enabling efficient and effective analysis of this vast and complex data landscape (Jain et al., 2021). The popular machine learning classification algorithms include Naive Bayes, maximum entropy and support vector machines (Chamlertwat et al., 2012).

Classification algorithms fall broadly into two categories: probabilistic and non-probabilistic. Probabilistic data classification approaches rely on approximating the underlying distribution of features within a class. This reliance aligns well with the observed tendency of many real-world features to exhibit a probabilistic nature in their distribution. As a result, probabilistic approaches often demonstrate robust performance in classification tasks (Wickramasinghe & Kalutarage, 2021). Naive Bayes is a simple probabilistic classifier based on applying Bayes’ theorem and is particularly suitable when the dimensionality of the inputs is high (Chamlertwat et al., 2012). A number of studies have successfully utilized the Naive Bayes algorithm to assess customer perceptions. Sánchez-Franco et al. (2019) applied this method to classify hotel reviews, achieving high precision and recall. Putri et al. (2020) compared Naive Bayes with a support vector machine for sentiment analysis of e-wallet reviews, finding that both algorithms performed well, with Naive Bayes achieving a fair classification. Laksono et al. (2019) used Naive Bayes to classify restaurant customer satisfaction, reporting high accuracy. Ernawati et al. (2018) implemented the Naive Bayes algorithm with feature selection using a genetic algorithm for sentiment review analysis of online fashion companies, resulting in improved accuracy. Collectively, these studies demonstrate the effectiveness of the Naive Bayes algorithm in assessing customer perceptions.

3 Data and Methodology

3.1 Data collection

The data were collected from customer reviews along with the overall ratings of five popular smartwatch brands from the Amazon website. An Amazon user review consists of four important aspects: summary (review title), review text (review text content), rating (user rating of the product on a scale of 1 to 5) and helpfulness (the number of people who found the review useful). These aspects will help us understand and analyse the reviews to derive insights. We collected 1866 reviews, along with ratings such as 1 for very poor and 5 for excellent (2 - poor, 3 - good, 4 - very good), and the text was scraped using Python programming (scrappy package). Thus, a customer with good experience while using the smartwatch would give very good or excellent and vice versa. Matching HTML tags are pointed out and customer text reviews are extracted based on matched tags. The steps for the semi-automated scraping of data are as follows:

- In URL parsing, hardcoded URLs are fetched and it also navigates to the webpage depending on the URL provided.
- In HTML parsing of the relevant tags, which mainly contain the customer-generated text, the overall satisfaction rating is found.
- Extracting the data involves fetching the text reviews and also the overall satisfaction rating from the relevant tags.
- Saving the data into a comma-separated variable format for better understanding.

3.2 Sentiment analysis

Sentiment analysis is performed on the text reviews using the Python NLTK package (Bird, 2006; Loper & Bird, 2002). NLTK is a Python package designed for both teaching and research in computational linguistics and NLP. It has undergone significant development to stay up to date with Python advancements. Figure 1 shows a representation of the steps performed.

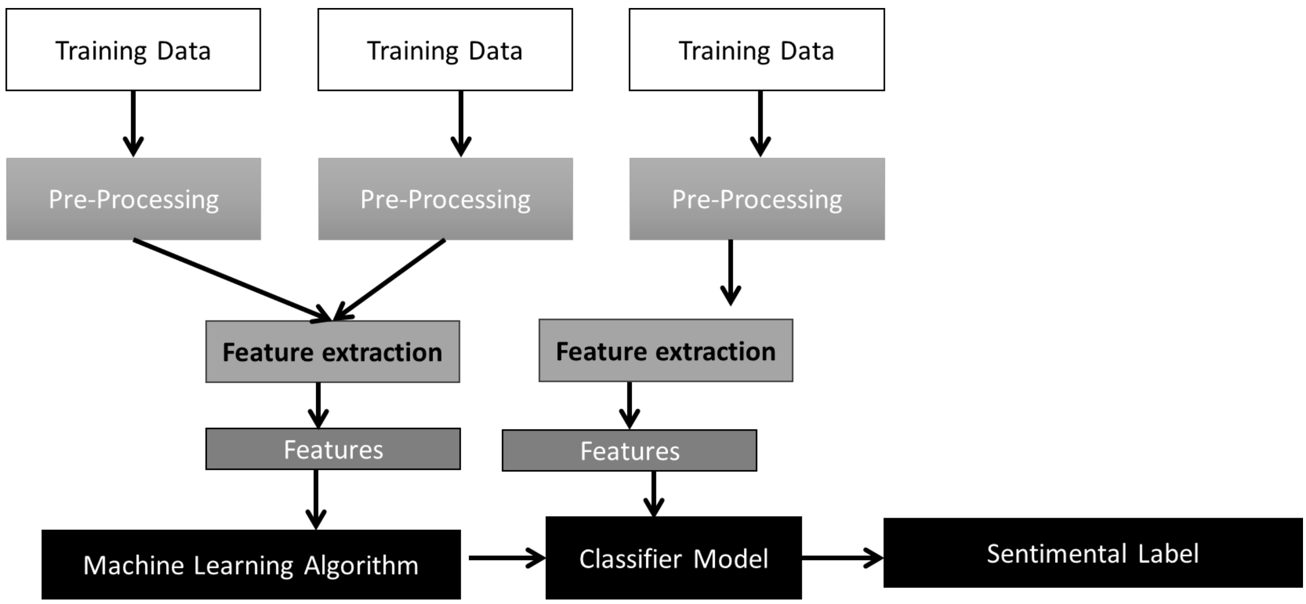

Figure 1. *Text analysis flowchart.*

3.2.1 Data pre-processing

Data segmentation

The method of breaking a text into a list of words based on white space and punctuation marks is called tokenization. A token is a series of characters that are grouped to be used for further processing.

Data normalization

This method transforms a document's entire word token into either lowercase or uppercase since most reviews consist of both case characters, i.e., lowercase and uppercase characters. It is easy to use the intent of moving all tokens into a single format to predict.

Stop word removal

The method of stop word elimination decreases the dimensionality of the datasets and the automated function extraction techniques can then easily classify the remaining keywords in the analysis corpus. Stop words ″a″, "of", "the", "I", "he", "she", "it", "at", "about", "and", etc. are of high frequency.

3.2.2 Feature extraction from text

The first step in the text classifier for machine learning is text extraction, and the bag of words is the classical approach to their frequency. The feature extraction methods are applied based on the word embedding. The feature extractor is primarily used in the prediction process to transform unseen text inputs into a feature vector. In the model that generates predicted tags such as positive, negative and neutral, the feature vectors are typically given.

Bag of words

The bag of words plays a crucial role in information retrieval within natural language processing, particularly in documentation. In sentiment analysis, it is regarded as a set of words conveying specific meanings, sometimes with dual interpretations. To address this ambiguity, a bag of words is often refined by incorporating part-of-speech (POS) tagging, ensuring a singular and precise meaning. This enhancement eliminates the potential confusion caused by the dual direction of meanings associated with word frequency. The POS tagger manages and provides a clear direction for the singular interpretation of a given word frequency. The bag of words encompasses two aspects: a vocabulary of known words and the assessment of the presence of these known words (Shroff et al., 2021). The word bag is accomplished by considering two positive and negative classes in which a bag of positive words is included in the positive class and a bag of negative words is included in the negative class. The data are predominantly collected as text and the frequency in the documents is counted. There are thus positive and negative results. After the extraction of features, the input is transferred mainly into a function vector by the feature extractor. The supervised learning algorithms, such as the Naive Bayes algorithm and SVM models, are applied to forecast the user reviews. The function vector then feeds into the model to produce tags such as positive, negative and neutral, which is expected. A comparison of the two models is performed. Also, a word cloud is generated to find the most frequently repeated terms in the positive and negative reviews.

3.2.3 Naive Bayes and support vector machine

One of the prominent supervised classifications is the Bayesian network classifier and the well-known classifier is the Naive Bayes, which is a probabilistic classifier that mainly relies on Bayes' theorem for the process. The approach provides a probability for a certain word, which has to be considered positive or negative. It calculates the probability of each word in a specific sentence, and this is considered output after the word with the highest probability is calculated. The probability of the text is calculated while considering the two classes of positive or negative, without analysing the input document, which mainly makes the test results positive or negative. The calculation is done by counting the relative frequencies contained in a corpus for each class.

Support vector machines are a method for finding the possible surface for separating the positive and negative training samples. The support vector machine algorithm is used in the sentiment polarity model to train and classify the text by primarily taking a step above the x/y prediction, such as positive and negative tags with two x and y data features. It trains the classifier to output either positive or negative as an x/y coordinate. The SVM is primarily defined as an input and output format, with a vector space in the input and a positive or negative output of 0 or 1.

For the comparison of the two models, we used a confusion matrix for identifying the complete performance of a particular model and classification accuracy such as precision, recall and F1 scores (Gamal et al., 2018; Tripathy et al., 2016) as indicated below.

The ratio between all recovered cases of true positives is precision, which can be calculated as:

$$Precision = \frac{\text{True positive}}{(\text{True positive + False positive})} \tag{1}$$

The ratio between all positive cases of true positives is recalled, which can be calculated as:

$$Recall = \frac{\text{True positive}}{(\text{True positive + False Negative})} \tag{2}$$

The combination of recall and precision is the F1 score, which can be calculated as:

$$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{(\text{Recall + Precision})} \tag{3}$$

To evaluate the capacity of a model to predict the true positives of each category available, we calculated specificity:

$$Specificity = \frac{\text{True Negative}}{(\text{True Negative + False Positive})} \tag{4}$$

3.2.4 Word cloud generation

A word cloud is an image of words. Common terms and phrases are highlighted using cloud creators based on frequency and relevance. They offer short visual perspectives that can lead to more in-depth analysis. Here, the word cloud is generated using the NLTK package to find the most frequently repeated term in the positive and negative reviews.

4 Results

Data generated are pre-processed with different steps such as cleansing, tokenizing, stemming (Porter's), lemmatizing, part-of-speech tagging, etc., using the NLTK package. After pre-processing, the reviews are classified as positive and negative ones. In total, 70% of the reviews are positive, 10% are neutral and 20% are negative, indicating that the overall sentiment is positive.

4.1 Word cloud from text analysis

Feature extraction is done using a bag of words and a word cloud is formed for both positive and negative reviews, as shown in Figures 2 and 3. In the positive word cloud, “quality” has the greatest frequency and in the negative word cloud, “fake” obtained the greatest frequency.

Figure 2. Positive word cloud from text analysis.

Figure 3. Negative word cloud from text analysis.

4.2 Classification report for bag of words

According to the analysis, the Naive Bayes achieved an accuracy of 69%. The classification report for the Naive Bayes is shown in Table 1. The SVM classifier achieved an accuracy of 71%. The classification report for the SVM classifier is shown in Table 2.

Comparing the classification report for both Naive Bayes and SVM, the F1 score with respect to the positive class is higher for SVM than for Naive Bayes, whereas the F1 score for the negative class for both the classifiers attains similar levels of accuracy.

Table 1. Classification report for Naive Bayes.

	Precision	**Recall**	**F1-score**	**Support**
Positive	0.90	0.04	0.07	250
Negative	0.69	1.00	0.82	536
Macro avg.	0.79	0.52	0.44	786

Table 2. *Classification report for SVM.*

	Precision	Recall	F1-score	Support
Positive	0.75	0.12	0.21	250
Negative	0.71	0.98	0.82	536
Macro avg.	0.84	0.5	0.41	

4.3 Confusion matrix for bag of words

The analysis shows that for Naive Bayes, the sensitivity rate is 69% and the specificity rate is 90%. From Table 3, the confusion matrix shows 535 true positives, one false positive, 241 false negatives and nine true negatives. the true positive value is the highest, which means that 535 positive class data points are classified correctly by the model. In the SVM classifier, the sensitivity rate is 71% and the specificity rate is 75%. From Table 4, the confusion matrix shows 526 true positives, 10 false positives, 220 false negatives and 30 true negatives. The true positive value is again the highest, which means that 526 positive class data points are classified correctly by the model.

Table 3. *Confusion matrix for Naive Bayes.*

TP 535	**FP** 1
FN 241	**TN** 9

Table 4. *Confusion matrix for SVM.*

TP 526	**FP** 10
FN 220	**TN** 30

5 Discussion

In this research, we presented an approach to perform a sentiment analysis of smartwatch customer reviews using supervised machine learning algorithms for classification. We collected textual reviews from customer reviews on the Amazon website, which also included the overall ratings for five popular smartwatch brands. The sentiment analysis of the smartwatch reviews was performed using two classifiers, namely the support vector machine and Naive Bayes after pre-processing and creation of word clouds. The overall sentiment of the reviews is skewed in favour of positivity.

According to the word cloud, which is formed for both positive and negative reviews, "quality" is the word that appeared high in the positive reviews, which means that smartwatch customers are more focused on quality due to several features such as battery life, fitness features, display, connectivity, calling and notification. However, the word that appeared with the highest frequency in the negative review was "fake", which represents that even though the quality of the smartwatch had a positive response, there were concerns regarding whether the product was fake since it synchronizes with other devices and people are worried about their personal data breach. While purchasing the smartwatch, customers prefer quality and also check the genuineness of the product. This study shows that consumers emphasize quality while purchasing a smartwatch and also that the decision to purchase a smartwatch depends on the quality and distinguishing whether the product is fake or original. As implicated by Salahuddin & Lee (2021), customers may look for quality features such as performance level of durability, product safety, long battery life, usability, comfortability and reasonable price. Other research has established that smartwatch adoption is influenced by factors such as technical attributes (Jung et al., 2016) and information quality. In accordance with the findings of Park et al. (2016), the perceived quality of smartwatches appears to be intricately linked to features emphasized in advertising campaigns. This underscores the importance of not only technical aspects and information presentation but also the strategic portrayal of features in marketing efforts, collectively shaping consumer perceptions and influencing adoption decisions.

According to the analysis, the Naive Bayes classifier provided an accuracy of 69% and the SVM classifier provided 71%. The SVM had a higher accuracy compared with Naive Bayes, so it is a good option to provide better accuracy and a faster prediction. This is in line with the findings of Rahat et al. (2019). While comparing the sensitivity and specificity of the classifier, Naive Bayes had a sensitivity of 69% and a specificity of 90%. A higher specificity value means that samples correctly classified as negative are likely to be truly negative, and a lower sensitivity value means that the test does not correctly classify a sample as positive. The SVM classifier has a sensitivity of 71% and a specificity of 75%, which means that the specificity value is only slightly higher than the sensitivity value.

Addressing the research questions, RQ1 is confirmed by the overall sentiment in customer reviews leaning towards positive sentiments, emphasizing the importance of quality. RQ2 reveals that the most critical features for customers include aspects related to quality, echoing the findings from the word cloud analysis. Lastly, RQ3 demonstrates the effectiveness of support vector machines over Naive Bayes in classifying sentiment in textual data, providing insights into the suitability of machine learning algorithms for sentiment analysis in this context. This holistic understanding contributes to discussions on consumer preferences, quality considerations and the relevance of machine learning algorithms in interpreting sentiment in smartwatch reviews.

Based on the study's findings, industry practitioners should emphasize the superior quality of smartwatches in marketing, addressing authenticity concerns and highlighting features such as extended battery life and advanced fitness capabilities. Strategic advertising campaigns should align with customer priorities, focusing on educational content to explain technical aspects. Emphasize authenticity in advertising, reassuring customers about the genuine nature of the product. Highlight features that distinguish genuine smartwatches, addressing concerns about counterfeit products. Continuous communication on the product's originality will enhance consumer trust and contribute to a positive market presence.

5.1 Implications and Future Directions

The paper makes significant theoretical contributions by delving into consumer behaviour related to smartwatches by means of an analysis of customer reviews. Businesses can make use of this insight to discern the specific features and attributes prioritized by customers during their purchasing decisions. This knowledge not only facilitates strategic product development, aligning offerings with consumer preferences and market demands, but also empowers businesses to refine marketing strategies. Armed with insights derived from customer reviews, companies can tailor promotional efforts to highlight and emphasize key features, thereby enhancing overall effectiveness. In addition to the current contributions, future directions for this study could encompass a deeper exploration of contextual differences and consumer characteristics. Investigating how regional or cultural variations influence smartwatch preferences can provide valuable insights for businesses operating on diverse markets.

6 Conclusion

Against the backdrop of increasing smartwatch usage and ever-evolving features, it is important to understand consumer opinions and preferences to customize features and frame marketing strategies. In this regard, this study analysed customer reviews of smartwatches to identify the factors that the customers prefer most while purchasing the products. Word clouds for positive and negative reviews were formulated and the words "quality" and "fake" appeared high in positive and negative reviews, respectively. We also compared two machine learning algorithms and SVM was found to have higher accuracy than Naive Bayes.

Additional Information and Declarations

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: D.M.: Conceptualization, Methodology, Software, Data curation, validation supervision, Writing – Reviewing and Editing. S.K.: Conceptualization, Methodology, Writing – Reviewing and Editing. N.V.S.: Writing – Original draft preparation, Data curation, Investigation. A.V.: Conceptualization.

Data Availability: The data that support the findings of this study are available from the corresponding author.

References

Al-Maroof, R. S., Alhumaid, K., Alhamad, A. Q., Aburayya, A., & Salloum, S. (2021). User Acceptance of Smart Watch for Medical Purposes: An Empirical Study. *Future Internet*, 13(5), Article 127. https://doi.org/10.3390/FI13050127

Almarzouqi, A., Aburayya, A., & Salloum, S. A. (2022). Determinants of intention to use medical smartwatch-based dual-stage SEM-ANN analysis. *Informatics in Medicine Unlocked*, 28, 100859. https://doi.org/10.1016/J.IMU.2022.100859

Almheiri, E., Al-Emran, M., Al-Sharafi, M. A., & Arpaci, I. (2024). Drivers of smartwatch use and its effect on environmental sustainability: evidence from SEM-ANN approach. *Asia-Pacific Journal of Business Administration*, (ahead-of-print). https://doi.org/10.1108/APJBA-10-2023-0490

Basha, S. M., & Rajput, D. S. (2019). An innovative topic-based customer complaints sentiment classification system. *International Journal of Business Innovation and Research*, 20(3), 375–391. https://doi.org/10.1504/IJBIR.2019.102718

Bieber, G., Haescher, M., & Vahl, M. (2013). Sensor requirements for activity recognition on smart watches. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments* (paper no. 67). ACM. https://doi.org/10.1145/2504335.2504407

Bird, S. (2006). NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, (pp. 69–72). ACM. https://doi.org/10.3115/1225403.1225421

Bölen, M. C. (2020). From traditional wristwatch to smartwatch: Understanding the relationship between innovation attributes, switching costs and consumers' switching intention. *Technology in Society*, 63, 101439. https://doi.org/10.1016/J.TECHSOC.2020.101439

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48–57. https://doi.org/10.1109/MCI.2014.2307227

Chamlertwat, W., Bhattarakosol, P., Rungkasiri, T., & Haruechaiyasak, C. (2012). Discovering Consumer Insight from Twitter via Sentiment Analysis. *Journal of Universal Computer Science*, 18(8), 973–992.

Chandel, R. S., Sharma, S., Kaur, S., Singh, S., & Kumar, R. (2022). Smart watches: A review of evolution in bio-medical sector. *Materials Today: Proceedings*, 50, 1053–1066. https://doi.org/10.1016/J.MATPR.2021.07.460

Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, 44, 65–77. https://doi.org/10.1016/J.INFFUS.2017.12.006

Chuah, S. H. W., Rauschnabel, P. A., Krey, N., Nguyen, B., Ramayah, T., & Lade, S. (2016). Wearable technologies: The role of usefulness and visibility in smartwatch adoption. *Computers in Human Behavior*, 65, 276–284. https://doi.org/10.1016/J.CHB.2016.07.047

Elnagar, A., Alnazzawi, N., Afyouni, I., Shahin, I., Bou Nassif, A., & Salloum, S. A. (2022). Prediction of the intention to use a smartwatch: A comparative approach using machine learning and partial least squares structural equation modeling. *Informatics in Medicine Unlocked*, 29, 100913. https://doi.org/10.1016/J.IMU.2022.100913

Gamal, D., Alfonse, M., M. El-Horbaty, E.-S., & M. Salem, A.-B. (2018). Analysis of Machine Learning Algorithms for Opinion Mining in Different Domains. *Machine Learning and Knowledge Extraction*, 1(1), 224–234. https://doi.org/10.3390/make1010014

Ha, T., Beijnon, B., Kim, S., Lee, S., & Kim, J. H. (2017). Examining user perceptions of smartwatch through dynamic topic modeling. *Telematics and Informatics*, 34(7), 1262–1273. https://doi.org/10.1016/J.TELE.2017.05.011

Jain, A. P., & Dandannavar, P. (2016). Application of Machine Learning Techniques to Sentiment Analysis. In *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology*. IEEE. https://doi.org/10.1109/ICATCCT.2016.7912076

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41, 100413. https://doi.org/10.1016/J.COSREV.2021.100413

Jin, X., Li, L., Dang, F., Chen, X., & Liu, Y. (2022). A survey on edge computing for wearable technology. *Digital Signal Processing*, 125, 103146. https://doi.org/10.1016/J.DSP.2021.103146

Jung, Y., Kim, S., & Choi, B. (2016). Consumer valuation of the wearables: The case of smartwatches. *Computers in Human Behavior*, 63, 899–905. https://doi.org/10.1016/J.CHB.2016.06.040

Kim, K. J., & Shin, D. H. (2015). An acceptance model for smart watches: Implications for the adoption of future wearable technology. *Internet Research*, 25(4), 527–541. https://doi.org/10.1108/INTR-05-2014-0126

Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naive Bayes. In *Proceedings of 2019 International Conference on Information & Communication Technology and Systems*. IEEE. https://doi.org/10.1109/ICTS.2019.8850982

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Springer. https://doi.org/10.1007/978-3-031-02145-9

Loper, E., & Bird, S. (2002). NLTK: The Natural Language Toolkit. arXiv preprint cs/0205028.

Mbunge, E., Muchemwa, B., Jiyane, S., & Batani, J. (2021). Sensors and healthcare 5.0: transformative shift in virtual care through emerging digital health technologies. *Global Health Journal*, 5(4), 169–177. https://doi.org/10.1016/J.GLOHJ.2021.11.008

Misra, S., Adtani, R., Singh, Y., Singh, S., & Thakkar, D. (2023). Exploring the factors affecting behavioral intention to adopt wearable devices. *Clinical Epidemiology and Global Health*, 24, 101428. https://doi.org/10.1016/j.cegh.2023.101428

Nasukawa, T., & Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In *Proceedings of the 2nd international conference on Knowledge capture* (pp. 70–77). ACM. https://doi.org/10.1145/945645.945658

Park, Y. W., Son, S., & Kim, B. (2016). Is smartwatch perceived as a wristwatch or a wearable device? The experimental study for examining the categorization and the perceived fit with manufacturer on consumer evaluations. In *Proceedings of the 18th Annual International Conference on Electronic Commerce: e-Commerce in Smart connected World*. ACM. https://doi.org/10.1145/2971603.2971624

Putri, D. A., Kristiyanti, D. A., Indrayuni, E., Nurhadi, A., & Hadinata, D. R. (2020). Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review. *Journal of Physics: Conference Series*, 1641(1). https://doi.org/10.1088/1742-6596/1641/1/012085

Qazi, A., Raj, R. G., Hardaker, G., & Standing, C. (2017). A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges. *Internet Research*, 27(3), 608–630. https://doi.org/10.1108/INTR-04-2016-0086

Rahat, A. M., Kahir, A., & Masum, A. K. M. (2019). Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. In *2019 8th International Conference System Modeling and Advancement in Research Trends*, (pp. 266–270). IEEE. https://doi.org/10.1109/SMART46866.2019.9117512

Reeder, B., & David, A. (2016). Health at hand: A systematic review of smart watch uses for health and wellness. *Journal of Biomedical Informatics*, 63, 269–276. https://doi.org/10.1016/J.JBI.2016.09.001

Salahuddin, M., & Lee, Y. A. (2021). Identifying key quality features for wearable technology embedded products using the Kano model. *International Journal of Clothing Science and Technology*, 33(1), 93–105. https://doi.org/10.1108/IJCST-08-2019-0130

Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2019). A Naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research*, 101, 499–506. https://doi.org/10.1016/j.jbusres.2018.12.051

Saraswat, D., Bhattacharya, P., Verma, A., Prasad, V. K., Tanwar, S., Sharma, G., Bokoro, P. N., & Sharma, R. (2022). Explainable AI for Healthcare 5.0: Opportunities and Challenges. *IEEE Access*, 10, 84486–84517. https://doi.org/10.1109/ACCESS.2022.3197671

Shroff, A., Shah, B. K., Jha, A., Jaiswal, A. K., Sapra, P., & Kumar, M. (2021). Multiplex Regulation System With Personalised Recommendation Using ML. In *2021 5th International Conference on Trends in Electronics and Informatics* (pp. 1567–1573). IEEE. https://doi.org/10.1109/ICOEI51242.2021.9453005

Siepmann, C., & Kowalczuk, P. (2021). Understanding continued smartwatch usage: the role of emotional as well as health and fitness factors. *Electronic Markets*, 31(4), 795–809. https://doi.org/10.1007/S12525-021-00458-3

Srinivasan, R., & Subalalitha, C. N. (2023). Sentimental analysis from imbalanced code-mixed data using machine learning approaches. *Distributed and Parallel Databases*, 41(1–2), 37–52. https://doi.org/10.1007/S10619-021-07331-4

Tandon, V., & Mehra, R. (2023). An Integrated Approach for Analysing Sentiments on Social Media. *Informatica (Slovenia),* 47(2), 213–220. https://doi.org/10.31449/inf.v47i2.4390

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, 57, 117–126. https://doi.org/10.1016/j.eswa.2016.03.028

Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, 25(3), 2277–2293. https://doi.org/10.1007/S00500-020-05297-6

Xu, L., Pan, C., Xu, B., Sun, Z., Zhou, K., Yan, R., Weng, K., Wu, Y., & Li, H. (2024). A Qualitative Exploration of a User-Centered Model for Smartwatch Comfort Using Grounded Theory. *International Journal of Human–Computer Interaction*, (ahead-of-print), 1–16. https://doi.org/10.1080/10447318.2024.2313273

Zhang, Y., & Rau, P. L. P. (2015). Playing with multiple wearable devices: Exploring the influence of display, motion and gender. *Computers in Human Behavior*, 50, 148–158. https://doi.org/10.1016/J.CHB.2015.04.004

Editorial record: The article has been peer-reviewed. First submission received on 16 November 2023. Revisions received on 23 January 2024 and 28 February 2024. Accepted for publication on 2 March 2024. The editor in charge of coordinating the peer-review of this manuscript and approving it for publication was Zdenek Smutny.

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951