INSTITUTO
UNIVERSITÁRIO
DE LISBOA

A Text Mining Approach to Portuguese Terroir: Analysing Online Wine Reviews

Mariana Patrícia Silva Ferreira

Master in Data Science

Supervisor:
Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor, Iscte – Instituto Universitário de Lisboa

Co-Supervisor:
Doctor Fernando Manuel Marques Batista, Associate Professor, Iscte – Instituto Universitário de Lisboa

September, 2024

BUSINESS SCHOOL

TECHNOLOGY AND ARCHITECTURE

Department of Quantitative Methods for Management and Economics

Department of Information Science and Technology

A Text Mining Approach to Portuguese Terroir: Analysing Online Wine Reviews

Mariana Patrícia Silva Ferreira

Master in Data Science

Supervisor:
Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor, Iscte – Instituto Universitário de Lisboa

Co-Supervisor:
Doctor Fernando Manuel Marques Batista, Associate Professor, Iscte – Instituto Universitário de Lisboa

September, 2024

To my mom, Ana

Acknowledgment

While I was writing this dissertation, there were a few individuals whose support was particularly significant, and to whom I owe a special note of thanks. First and foremost, I would like to express my deepest appreciation to my supervisors, Professors Ricardo Ribeiro and Fernando Batista. Your constant support and insightful guidance have been the backbone of this work. I could not have asked for better, more inspiring mentors. It has been a privilege to be your student.

To my dear friends Mariana Casanova, André Forte, António Costa, and Edna Neves, your care and encouragement meant the world. I am deeply grateful to have you in my life. To Afonso Caniço, your deep interest in this work and constant motivation are too greatly appreciated.

To Vinicius Campbell, Mariana Fonseca, and André Amado, my trusted partners in assignments throughout this Master's, thank you for the collaboration, shared laughter, and unwavering support. I have truly enjoyed working with you all.

Balancing a full-time job while pursuing a Master's degree has been challenging, but I was fortunate to work with colleagues who recognised and supported my ambitions. Remembering my time at Milestone, I extend my heartfelt thanks to my colleagues there, especially Maria Vidal, Diogo Almeida, and Jorge Afonso – you are all very dear to me.

Transitioning to a new job brought some uncertainty. However, I luckily found an incredibly supportive and flexible work environment at Deloitte, for which I am immensely grateful. I would like to thank the entire EPM & Analytics team for taking an interest and offer valued advice, but particularly to Inês Dang, always celebrating even the smallest achievements, and Luís Tomás, whose investment in my studies and continuous support were invaluable.

Reflecting on the start of my career, I must acknowledge Paulo Valadares, who first hired me during my graduate studies and consistently encouraged me to pursue excellence. Thank you for your belief in me, your support, and your friendship.

Finally, I express my deepest gratitude to my father, Mário Ferreira, who has always supported and celebrated my academic achievements; and to the most important person in my life, my mother, Ana Ferreira. Had it not been for your nurturing of my curiosity, demonstrated perseverance and resourcefulness, I might not have developed the drive to pursue academic excellence and continually strive for greater heights. I am truly fortunate to have grown up with you as my role model. Thank you for your endless patience, encouragement, and unconditional love.

Resumo

A presente dissertação aplica técnicas de mineração de texto para analisar mais de 120.000 avaliações de vinhos portugueses, escritas em inglês, da plataforma Vivino. O principal objetivo é obter informações sobre as preferências dos consumidores, que podem ajudar os produtores e profissionais de marketing a alinhar as suas ofertas com as exigências do mercado. A dissertação segue uma metodologia inspirada no Cross-Industry Standard Process for Data Mining (CRISP-DM). O estudo utilizou análise de sentimento, revelando uma predominância de avaliações positivas; modelação de tópicos, que identificou três temas principais (características sensoriais, perceções de valor e preferências de estilo de vinho); deteção de emoções, que demonstrou que alegria é a emoção mais comum; e classificação de avaliações, onde foram usados modelos Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM) e Support Vector Machine (SVM) para prever as classificações dos vinhos com base no conteúdo das avaliações textuais. Os modelos alcançaram uma taxa de acerto moderada, sugerindo baixa consistência entre os utilizadores da plataforma. Estes resultados demonstram o potencial da mineração de texto para extrair informação sobre os consumidores a partir de avaliações online. Ao mesmo tempo, esta dissertação expandiu o campo de aplicação da área de Wineinformatics, sendo este o primeiro estudo a ter por base conteúdo gerado por utilizadores e não profissionais.

Palavras Chave: *Mineração de Texto*, *Wineinformatics*, *Análise de Sentimento*, *Modelação de Tópicos*, *Deteção de Emoções*, *Classificação*

Abstract

The present dissertation applies text mining techniques to analyse over 120,000 English-written reviews of Portuguese wines from Vivino. The main objective is to retrieve insights regarding consumer preferences, which can provide winemakers and marketers a way to align their offerings with market demands. The dissertation follows a Cross-Industry Standard Process for Data Mining (CRISP-DM) inspired methodology. The study employed sentiment analysis, revealing prevalence of positive reviews; topic modelling, which uncovered three main themes (sensory characteristics, value perceptions, and wine style preferences); emotion detection, that showed joy as the most common emotion; and review classification, which was based on Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Support Vector Machine (SVM) models to predict wine ratings from textual review content. The models achieved a moderate accuracy, suggesting low consistency among reviewers. All these findings demonstrated the potential of using text mining to extract consumer insights from online reviews. At the same time, this dissertation expands Wineinformatics as a field, given this is the first to utilise User-Generated Content (UGC) as its foundation, instead of professional reviews.

KEYWORDS: *Text Mining*, *Wineinformatics*, *Sentiment Analysis*, *Topic Modelling*, *Emotion Detection*, *Classification*

Contents

Acknowledgment iii

Resumo v

Abstract vii

List of Figures xi

List of Tables xiii

List of Acronyms xv

Chapter 1. Introduction 1
1.1. Motivation 2
1.2. Research Purpose 3
1.3. Research Questions 3
1.4. Methodology 4
1.5. Document Structure 5

Chapter 2. Literature Review 7
2.1. Theoretical Background 8
2.2. Methodology 10
2.3. Related Literature 11
2.4. Summary 21

Chapter 3. Data 23
3.1. Data Analysis 24
3.2. Data Preparation 26
3.2.1. Basic Preprocessing 27
3.2.2. Lemmatisation 27

Chapter 4. Extracting Knowledge from User Reviews 29
4.1. Methods Employed 29
4.1.1. Sentiment Analysis 29
4.1.2. Topic Modelling 30
4.1.3. Emotion Detection 31
4.1.4. Review Classification 31
4.2. Sentiment Analysis 33
4.3. Topic Modelling 35

4.4. Topics by Sentiment Polarity 37
4.5. Emotion Detection 40
4.6. Review Classification 43

Chapter 5. Conclusions and Recommendations 47
5.1. Implications for Wine Industry 47
5.2. Implications for Wineinformatics 48
5.3. Limitations 49
5.4. Future Work 49

References 51

List of Figures

Figure 2.1	Sentiment analysis approaches [33]	9
Figure 2.2	PRISMA diagram retrieved from PRISMA Statement online page	11
Figure 2.3	Publications per year – Number of reports published each year	12
Figure 2.4	Citations count per year	12
Figure 2.5	Documents per type – Percentage of documents reviewed per type	13
Figure 3.1	Review count over the years	24
Figure 3.2	Review distribution per Rating bin, in percentage ($n = 510,018$)	25
Figure 3.3	Review distribution per Price bin, in percentage ($n = 510,018$)	25
Figure 3.4	Top 10 languages with most reviews, in percentage ($n = 510,018$)	25
Figure 3.5	Review character length distribution per language	26
Figure 3.6	Word count distribution per language	26
Figure 4.1	Illustration of the research process adopted	29
Figure 4.2	Review distribution by assessed polarity ($n = 129,673$)	34
Figure 4.3	Sentiment polarity by Rating bins	35
Figure 4.4	Inter-topic distance map	36
Figure 4.5	Coherence scores by the number of topics	36
Figure 4.6	Topic 1: Wine Tasting Experience and Sensory Characteristics	37
Figure 4.7	Topic 2: Wine Value and Consumer Satisfaction Characteristics	38
Figure 4.8	Topic 3: Wine Style and Freshness	39
Figure 4.9	Topic distribution by sentiment polarity	39
Figure 4.10	Top 20 bigrams for each rating bin	40
Figure 4.11	Top 25 aromatic compounds correlated with joy and disgust	42
Figure 4.12	Top 20 unigrams for the emotion of joy (filtered)	42

List of Tables

Table 2.1	Summary of research results	13
Table 3.1	Collected data for each record	23
Table 4.1	Polarity Score Interpretation for Stanza and VADER Libraries	34
Table 4.2	Emotion Counts	41
Table 4.3	Proportion of classes for the Two-Class and Four-Class Classifications	43
Table 4.4	Accuracy, Precision, Recall, and F-1 Score for each model variation	45

List of Acronyms

API: Application Programming Interface

BERT: Bidirectional Encoder Representations from Transformers

BiLSTM: Bidirectional Long Short-Term Memory

CNN: Convolutional Neural Network

CWW: Computational Wine Wheel

KNN: K-Nearest Neighbors

LDA: Latent Dirichlet Allocation

LLM: Large Language Model

LSTM: Long Short-Term Memory

MA: Macro Average

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

ME: Mean Error

ML: Machine Learning

MSE: Mean Squared Error

mhl: million hectoliters

NLP: Natural Language Processing

NMAE: Normalized Mean Absolute Error

NN: Neural Network

NRMSE: Normalized Root Mean Squared Error

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

RMSE: Root Mean Squared Error

SMOTE: Synthetic Minority Oversampling Technique

SVM: Support Vector Machine

TF-IDF: Term Frequency-Inverse Document Frequency

UGC: User-Generated Content

VADER: Valence Aware Dictionary for sEntiment Reasoning

WA: Weighted Average

WOM: Word of Mouth

CHAPTER 1

Introduction

Nowadays, the wine sector is facing rising competition and changing consumer preferences. In particular, it has been observed a shift towards the consumption of less alcoholic wines and more sustainable production and packaging ways [1]. Although changing in its form, wine consumption has been registering a global decrease, making competitiveness and costumer retention much more challenging.

Since 2020, the wine industry has been significantly and steadily growing at a global level. In 2023, it had already crossed the 300-billion-euro mark of combined revenue. In 2024, an approximate 8% increase in revenue is expected and the industry is estimated to reach the value of 410 billion euros in 2028 [2].

This robust growth is mostly driven by the rising interest in premium wines, but also some changes in consumer preferences as organic wines take their space in the global market. At the same time, e-Commerce platforms and online presence boost sales and interest in the industry [3].

Regarding global wine production, in 2023 it registered 237 million hectoliters (mhl), a decrease of 9.6% when compared to 2022. This volume is the lowest since 1961, when it was recorded 214 mhl. This sharp decrease is credited to several adverse climate events and widespread fungal diseases that impacted harvest in the main players of both hemispheres [4].

When it comes to global wine consumption, in 2023 it is estimated to be set at 221 mhl, reflecting a reduction of 2.6% compared to the previous year. This is attributed to China's decrease in consumption, but still exacerbated by the downward trend set during the COVID-19 pandemic in 2020, due to the lockdown measures that greatly affected the wine markets around the world [4].

If we analyse international trade, though, we find one interesting fact. Global wine export volume, which decreased by 6.3% to 99.3 mhl, was the lowest recorded since 2010. Despite this, global wine export value in 2023 reached 36 billion euros, the second highest ever recorded. This reflects the high average export price of 3.62 EUR/l, representing an increase of 2% faced to 2022. This is the highest price ever recorded. Even if other aspects may influence registered prices, this is also a direct consequence of the inflationary pressures registered across the globe. This rise in prices stems higher costs incurred, consequently driving up bottle prices [4].

1.1. Motivation

Faced with their challenging and ever changing reality, wineries might want to address their obstacles through professional management approaches. Resorting to Marketing, producers can seduce new types of consumers and markets, build lasting relationships with current buyers, while also upgrading the experience proposed by the wine sold, creating value through branding [5].

For that, it is imperative to know which characteristics are consumers considering when they look for their next bottle. One way to become informed on these matters may be through social media, which marketing practitioners already recognise as a valuable source for expanding businesses [6].

However, one particular effective way to drive up wine sales is through Word of Mouth (WOM). The interaction among social media users allows buyers to share their experiences and encourage more users to make similar choices (or otherwise) [7]. Thus, social media arises as a vital and influential channel for wine purchases [6].

Websites like Vivino[1] are great sources of User-Generated Content (UGC) with a focus on wine tasting experiences and ratings. Forums like this are becoming crucial to the wine sector as they contribute to stimulate consumers' interest in a wine-producing destination, a specific winery, wine variety, or brand [8]. On the other hand, analysing their contents can reveal profiles, motives, perceptions and attitudes of wine consumers, resulting in a rich source of information to the ones behind the wine selling counter [9].

These reviews are usually comprised of a free-text field and a numerical rating. Although they are proved to influence customers' decision-making processes, their enormous volume and unstructured nature make it difficult for customers and analysts to derive valid and reliable insights from them [10].

As technology advances, text analysis becomes more useful for a myriad of applications. Indeed, Text Mining refers to the information retrieval, extraction and Natural Language Processing (NLP) techniques to uncover patterns and knowledge from textual data [11].

Within that field, we are able to use techniques such as Sentiment Analysis to further investigate wine reviews and retrieve their polarity (positive, negative, or neutral). Additionally, Topic Modelling enables us to uncover the main themes discussed by consumers, providing deeper insights into their preferences and concerns. This allows stakeholders to assess consumers' satisfaction and what are they stating when expressing such feelings towards wine they taste [12].

[1] www.vivino.com

1.2. Research Purpose

The wine market is witnessing a rise of competitiveness and is no longer owned by the Southern European countries. Often referred to as "Old World", the traditional wine-producing countries have been disputing more and more market share with "New World" countries [13].

One of the oldest and most famous producers of quality wine is Portugal. The country was the 10th largest wine producer and consumer in the world, in 2023 [4]. Being no exception to the trend, the country not only has been losing business but its vineyard area has been dwindling in recent years [13].

To course correct trajectories such as the one Portugal is drawing, new and creative tools must be used to overcome these adversities. In this perspective, Wineinformatics combines data science and wine-related datasets, producing useful information for wine producers, distributors and consumers [14].

Within this field, a plethora of studies have already been performed focussing always on expert notes. These have assessed wine aging capability [15], wine quality [16–18], consistency of wine judges [19–22], wine characteristics [23], amongst other goals.

However, to the present date and to the best of our knowledge, none have focused on reviews from the everyday consumer. Within this framework, the present study targets consumers' opinions on wine-domain social media forums while highlighting the different aspects pointed in their notes.

This focus on the common buyer's opinion is highly relevant as it provides direct insights on their preferences and perceptions. At a large scale, it can be used to understand market dynamics or, at least, get a hint of it. As a result, these reviews can provide producers with valuable insights to align their production with market demands, ultimately improving customer satisfaction.

In summation, Wineinformatics earns as well a new perspective, and wine-producing countries, such as Portugal, may trace strategies from this information to overcome their obstacles and regain market share.

1.3. Research Questions

The rise of UGC on online platforms has created new opportunities for industries to better understand consumer preferences through data-driven approaches.

One of the main objectives of this research is to analyse Portuguese wine reviews aiming at extracting meaningful insights about consumer sentiments and popular topics in their shared opinions. Additionally, this work also pretends to explore the possibilities that Text Mining offers to support wine producers and marketers in aligning business strategies with market demands, providing a way to evaluate the effectiveness of several Machine Learning (ML) models in assessing reviewers consistency.

In that sense, the research questions guiding the present dissertation are:

- RQ1: What are the most common sentiments and emotions expressed towards the reviewed wines?
- RQ2: What are the main topics influencing consumer preferences, regarding Portuguese wines?
- RQ3: How effectively can ML models predict wine ratings based on the textual data from reviews?

These questions are thoroughly explored and answered through the following chapters.

1.4. Methodology

The present dissertation revolves around the use of text mining techniques to extract meaningful information from Portuguese wine reviews, with the ultimate goal of providing stakeholders with valuable insights for their practice. By analysing user reviews, these players can adapt their production processes and/or marketing strategies to better align with consumer preferences, thus enhancing product commercialisation while overcoming the difficulties the sector is facing.

The data for this study was sourced from the Vivino website, a popular platform where everyday buyers can post reviews and rate wines. This dataset consists of UGC, including textual reviews and the corresponding ratings. Varying in length and content, these reviews ultimately reflect consumer perceptions of Portuguese wines. The challenge, then, lies in analysing the unstructured textual data to uncover useful, actionable insights.

The methodology followed was inspired in CRISP-DM (Cross-Industry Standard Process for Data Mining), which is widely adopted in data science projects, given its structure and iterations. The CRISP-DM methodology consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

Firstly, in the Business Understanding phase, we understood established consumer preferences, wine sector trends and ultimately resorted to what was already available in the literature to understand the consolidated research and what was missing regarding the subject.

Once that was achieved, we entered Data Understanding. This was the phase in which we explored the data structure, understood the various distributions of important variables, and identified key patterns within the corpora (see Section 3.1).

Data Preparation was comprised of several operations, in order to increase data quality and get better results in the following phase. Next, in the Modelling phase various experiments were undertaken: Sentiment Analysis, Topic Modelling, Emotion Detection and Review Classification, which was based on two Neural Network (NN) models plus a Support Vector Machine (SVM), having been evaluated using metrics such as accuracy, precision, recall, and F1 score.

Finally, although full deployment is beyond the scope of the present dissertation, the findings provide a foundation for practical applications in the industry and contribute with a novel approach to Wineinformatics as a field.

In summation, the experiments conducted in this dissertation focused on four main tasks. Sentiment analysis aimed to classify the overall sentiment expressed in the reviews as negative, neutral, or positive, with Stanza and Valence Aware Dictionary for sEntiment Reasoning (VADER); Topic modelling was used to uncover common themes and topics discussed in these reviews, through Latent Dirichlet Allocation (LDA). In addition to general sentiment, emotion detection focused then on identifying specific emotions expressed, providing further granularity to the first task, with the help of GPT-3.5. Lastly, review classification involved predicting the rating associated with each review, based solely on the textual content, having been used Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM) and SVM models.

1.5. Document Structure

The present work is structured as follows: Chapter 2 presents a review of the literature relating to the subject of the present dissertation. Chapter 3 explains all the matters related to Data Collection and subsequential Data Analysis and Preparation. Results and discussion are presented in Chapter 4, containing as well an introduction on the employed methods. Finally, Chapter 5 showcases the conclusions of the present work, its implications, and future work recommendations.

CHAPTER 2

Literature Review

While one may be acquainted with professional critiques and their distinct vernacular regarding trending wines, the present work, on the other hand, pretends to shed light on what the real, everyday consumers are saying online. This study plays a very relevant role for the wine industry in revealing customer's preferences and sentiments towards specific product characteristics.

In fact, positive reviews from experts generate a transitory, but quantitatively important, increase in demand for a given product. The mere fact of it being reviewed has a small positive effect on demand. However, these expert notes might not be so relatable for the customer, reducing the impact of the reviews. For its part, customer-generated information has been proven to affect demand for experience goods [24].

We live in an era of information overload; everyone can be an opinion maker. Within seconds, one can access social platforms where certain items are rated and discussed by whoever wants to express their feelings. These opinions are commonly summarized in rating statistics or available in individual comments, frequently in a concise format. Being accessible and convenient, online reviews can be a rich source of information for buyers.

The online WOM literature suggests these rating statistics commonly highlighted on the product page are the primary drivers of purchase decisions and product sales [25]. Yet, it was found that although consumers do not read reviews consistently, they do rely on them for products that are expensive or of uncertain quality [26].

Back in 2008, it was suggested that for lower-cost products with sufficiently informative reviews, the product attribute information provided by sellers and the reviews created by buyers would complement each other. In such cases, the optimal strategy for sellers is to increase the amount of information available about product attributes. Conversely, for higher-cost products, particularly when there is a significant number of less knowledgeable consumers, the two types of information act as substitutes. In these situations, the most effective response for sellers is to reduce the amount of product attribute information they provide [27].

Taking all this into account, it is known that until 2022, wine trade and international wine consumption by volume have not been significantly growing since 2011. However, in value terms, a sharp growth has been registered, which has been attributed to the product premiumisation [28].

Also, Del Rey and Loose [28] specified that this overall tendency of steady volume and increasing value does not equally apply to all product categories. There is currently a major global shift towards lighter and refreshing wines. This trend might reflect climate changes, younger consumers' likings, new geographical markets' preferences, among other explanations.

This way, we begin to discern some links. Reviews from experts are important, but emerging evidence suggests that individual reviews also affect the decision-making of consumers [25]. Given that wines are registering increasingly higher prices, and their qualities are difficult to assure, we can suspect that UGC may play a pertinent role when a buyer visits the internet to scout for information on a particular wine.

Building on this, we now know that different types of wine are becoming more popular, hence receiving more reviews. Thus, it might affect sales and overall perception. Considering then the different scenarios and wine characteristics, sellers might want to adapt the information they provide in response to how many reviews are issued and what do they say.

For that purpose, we can use NLP to extract features and summarize relevant and meaningful information. With that, it is possible to perform, for example, sentiment analysis on the unstructured textual data and carry investigations in applied domains, such as wine reviews, to incorporate in further analysis [29].

In light of this, Text Mining arises as a resourceful tool to measure the pulse of wine consumers. With it, we can discover what is being said or how much the product is being discussed and sellers might want to make changes on their production or the way they market their products. Additionally, these reviews are a good proxy to assess popular types of wine on the market, more specifically which wine characteristics are making a positive impact in the general opinion, revealing consumer trends.

All in all, this literature review aims to investigate how Text Mining techniques, such as Sentiment Analysis, have been applied to the field of online reviews. Afterwards, with an emphasis on corpora of wines reviews, this knowledge will be used to infer consumer trends and derive consequent strategies for winemakers to adopt.

Further on this chapter, we investigate the following Research Question: Are there studies on Text Mining techniques applied to Wine Reviews?

This is relevant given the growing standing of the Wine Industry and above elaborated considerations on the helpful insights Text Mining can bring to the market. Moreover, admitting the regional differences across the vintage offering world-wide, we intend to investigate the specific case of Portugal, since the country is one of the largest wine producers and consumers in the world [30].

2.1. Theoretical Background

In the present work, various tasks are performed. This section provides a brief introduction to each of those subjects.

With the popularity of social media, people were enabled to post their views on a diversity of topics online, be it in formal or informal settings. These arise in the form of individual reviews, forums, social media posts, blogs, or discussion boards [31].

Also referred to as Opinion Mining, Sentiment Analysis aims to discover positive and negative sentiments about a given subject and its attributes using text processing. In this sense, discovering "useful opinions from raw text requires a nontrivial amount of natural language processing". In this context, Sentiment Analysis resorts to various methodologies to discover the polarity of a given sentence. Examples of those are feature engineering, entity extraction or classification [31].

Although the study of Subjectivity Analysis has been around for some time, the research in Sentiment Analysis gained strength with the introduction of methods based on Machine Learning and the use of annotated datasets. Other approaches have also been introduced, such as Dictionary, Statistical, and Semantic, but Machine Learning was established as the predilect tool for tackling these problems [32].

Considering this, there are various approaches to Sentiment Analysis. The three mainly used ones are Lexicon Based Approach, Machine Learning Approach and Hybrid Approach [33]. Tied to these, there are several models available, which are presented in Figure 2.1 by Wankhade, Rao, and Kulkarni [33].

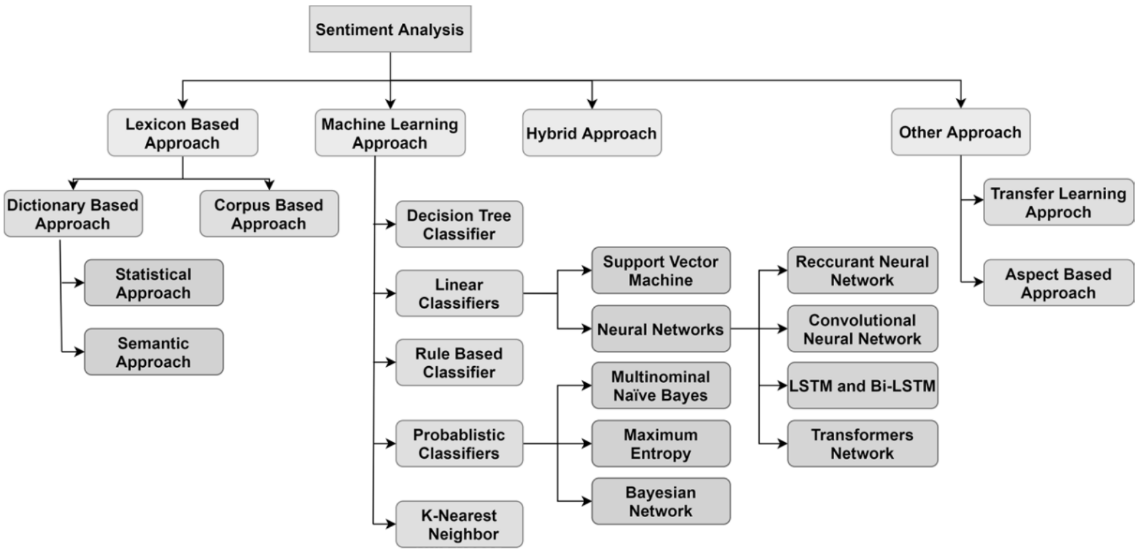

FIGURE 2.1. Sentiment analysis approaches [33]

According to Blei [34], Topic Modelling refers to a set of algorithms designed to identify and extract the main topics within large, unstructured collections of documents. These algorithms organise such collections based on the themes they uncover, allowing for a clearer understanding of the underlying patterns. Notably, Topic Modelling has proven effective for handling vast amounts of data, including streaming sources from web Application Programming Interfaces (APIs).

Still referring to the work of Blei [34], the author has suggested the New York Times complete history as an example. If one was to analysing its contents, at the top level, he could notice some of the themes would correspond with the newspaper sections. But if zoomed in, the section of foreign policy, for instance, could reveal various themes, such as the Chinese policy, conflicts over the Middle East, among others. But adding to that, it would even be possible to analyse these texts across the years, offering views regarding the way these subjects have evolved.

However, given the volume of data we face nowadays, this would be humanely impossible. To overcome this issue, ML emerges as a solution with the *probabilistic topic modelling*, which discovers and annotates corpora with its corresponding topic information [34].

More specifically, Topic Modelling algorithms such as LDA, used in the present work, are based on the proportions of different topics in the textual data [10]. Each one of these topics is characterised by specific vocabulary. So, in summation LDA uncovers the topics and associated words to describe the texts and annotates each document with probabilistic topic labels [10].

Finally, Review Classification is a task which employs ML to predict predefined categories based on the textual content reviews. In the framework of this work, such exercise comes as very valuable to assess UGC consistency, contrasting attributed ratings with the words written.

In similar pursuits, several models are commonly used, including SVM, CNN and BiLSTM. Each model has its distinct characteristics and degree of complexity.

SVM is a well established ML algorithm. It works by identifying a hyperplane that maximises the division of the data points into distinct classes. In the case of textual data, features such as word frequencies, in this work represented by Term Frequency-Inverse Document Frequency (TF-IDF), are used to convert text into numerical vectors [35].

CNNs were originally developed for image recognition tasks but were adapted for text classification. The model uses convolutional layers to detect patterns within the corpus, such as common word combinations, and local dependencies in text which are then utilised to predict the rating category [36].

Lastly, BiLSTM are networks designed to capture both past and future dependencies in sequential data by processing the input in two directions. This is translated by two Long Short-Term Memory (LSTM) layers, one for each direction, functioning as usual LSTM models, having the ability to maintain and update a memory cell over long sequences. The network retains relevant information while discards irrelevant data through its gating mechanisms (input, forget, and output) [37]. In this sense, the dual direction aspect allows BiLSTM models to consider the context of each word within the sequence, which is particularly important in contexts such as the one in this work, where the meaning of words depends on their surrounding.

2.2. Methodology

The methodology applied to this review was inspired by Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA); following, more specifically, the 2020 flow diagram for new systematic reviews, shown in Figure 2.2, which includes searches of databases and registers and other sources. Moreover, to structure this research, we defined the previously presented research question, following the choice of the databases and keywords, and finally the extraction and selection took place.

From now on, we will be orbiting the question of how sentiment analysis is being applied to wine reviews. For the search, we used the Scopus database. Regarding the query, we used the following:

TITLE-ABS-KEY (("Sentiment Analysis" OR "Opinion Mining" OR "Emotion Detection" OR "Text Mining" OR "Data Science" OR "Data Mining") AND ("Wine") AND ("Review"))

This query returned 35 results, during October 2023, and they were all screened. Given their title, keywords, and abstract, 18 of them were assessed for their eligibility. Adding to this, other sources (such as: Google Search, citation search, and recommendations) were used, and from these resulted another 7 studies, from which only 5 passed to eligibility assessment.

After a thorough reading of the articles, 17 from Scopus and 5 from other sources were chosen to be the basis for the present state-of-the-art review.

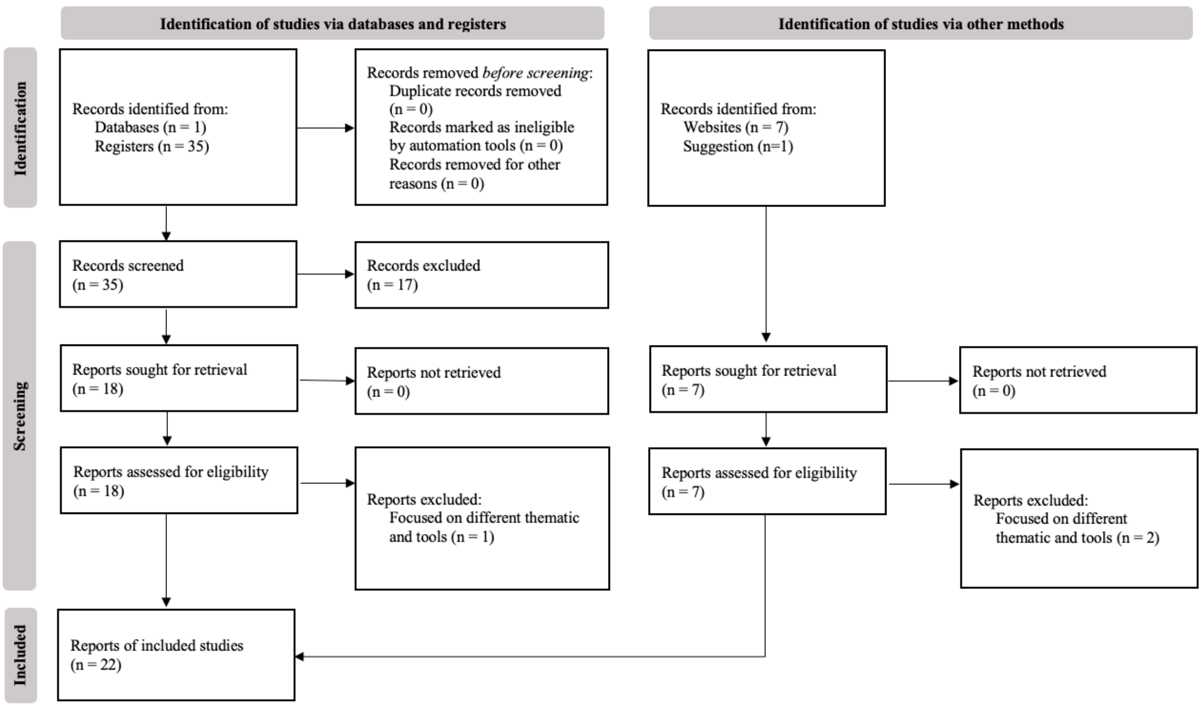

Figure 2.2. PRISMA diagram retrieved from PRISMA Statement online page

2.3. Related Literature

Firstly, we can say that this is a fairly recent topic. From the Scopus articles, the oldest article found is dated 2014. To the best of our knowledge, this was the date when Wineinformatics, a novel Data Science area applied to wine-related datasets, is originally introduced [38]. As can be seen in Figure 2.3, to this day, dedicated publications maintained modest levels, with citations following these numbers in some proportion.

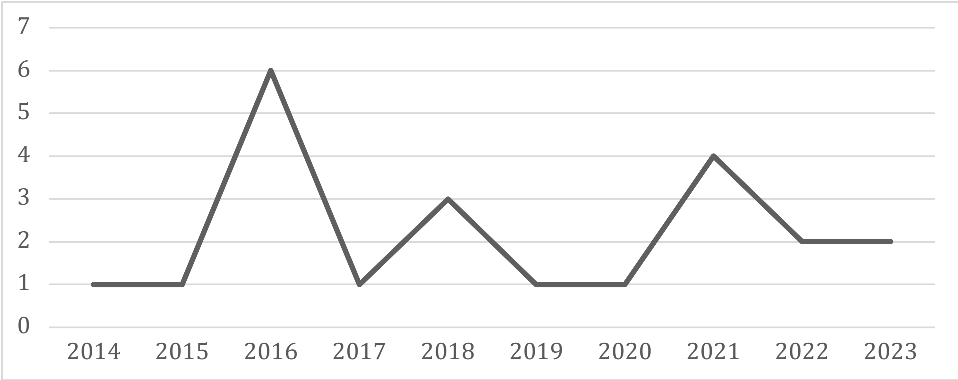

FIGURE 2.3. Publications per year – Number of reports published each year

Regarding citations, following what Figure 2.4 suggests, there is a clear difference between the first four years since Wineinformatics was introduced and the following period. This reflects the maturing of the field as well as the obvious consequence of the numbers above, their accumulated time available, and the popularity of the subject.

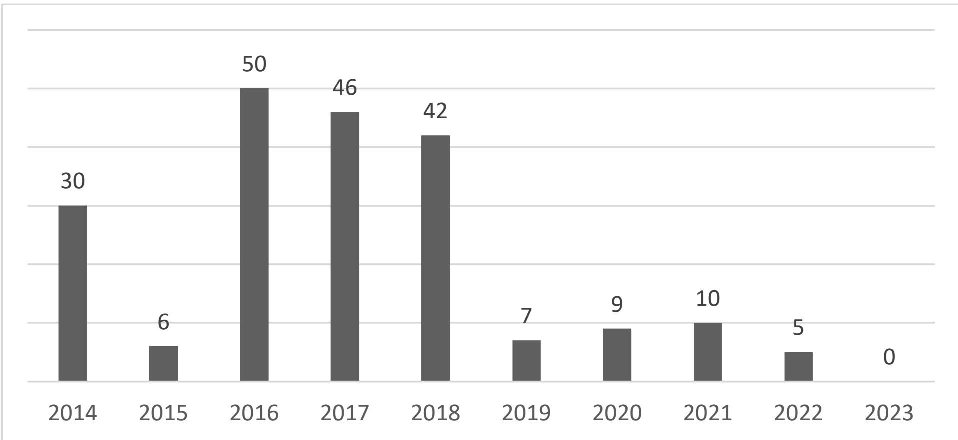

FIGURE 2.4. Citations count of the publications for a given year

Of all the considered studies, 13 are co-authored by Bernard Chen, one of the founder researchers of Wineinformatics. Regarding the nature of the results, the majority are journal articles and conference papers, as demonstrated by the percentage in Figure 2.5.

Regarding the methods applied, there is a prominence of algorithms such as SVM, Naïve Bayes, and K-Nearest Neighbors (KNN) and some interesting applications of BiLSTM, CNN and Bidirectional Encoder Representations from Transformers (BERT). These were applied to solve a variety of problems: assessing wine aging capability [15] predicting wine quality [16–18], evaluating the consistency of wine judges [19–22], among other goals.

In terms of performance measurement, metrics such as accuracy, sensitivity, and specificity are popular in classification problems, but the usage of metrics such as Mean Error (ME), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) was also observed in the respective regression problems.

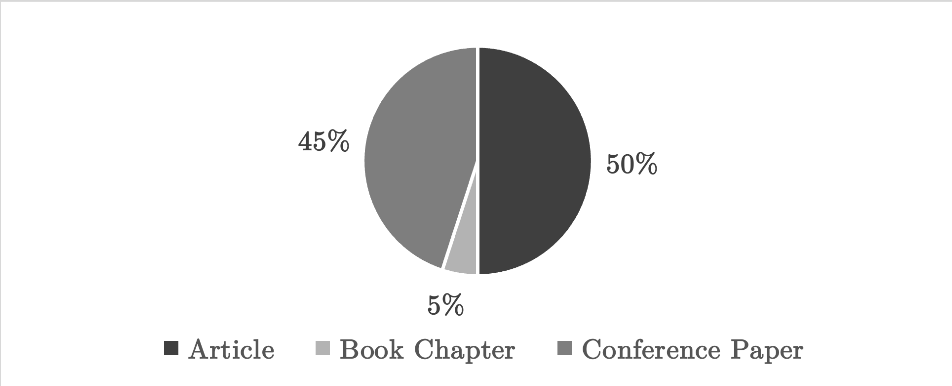

FIGURE 2.5. Documents per type – Percentage of documents reviewed per type

In addition, in all this literature, there is one very important highlight to be mentioned: the Computational Wine Wheel (CWW). As previously presented, reviews are classified as unstructured data and this requires processing of human language to make computers understand the format. In that sense, we are able to capture keywords from these unstructured datasets with the CWW, which utilises a dictionary-based approach, also known as the bag-of-words [39].

The results are summarised in Table 2.1.

Table 2.1: Summary of research results

Document	Dataset	Research goal	Method and Evaluation	Results
Applying Neural Networks in Wineinformatics with the New Computational Wine Wheel (2023) [39]	Wine Spectator and Robert Parker reviews	Propose the 3.0 Computational Wine Wheel and predict if a wine rates 95+ points or 94- points.	Methods: Naïve Bayes, SVM and NN. Evaluation: Accuracy	The CWW 3.0 combined with NN brought the better results than other algorithms.
Using Neural Network Models for Wine Review Classification (2022) [29]	Wine Spectator	Extract features from wine reviews and to classify those into different rating classes	Methods: CNN, BiLSTM, and BERT. Evaluation: Accuracy, precision, recall	All models demonstrated to be good predictors with no less than 80% accuracy, being BERT one of the top achievers, followed by BiLSTM.

Continues in the next page

Table 2.1 – Continuation from previous page

Document	Dataset	Research goal	Method and Evaluation	Results
Wineinformatics: Comparing and Combining SVM Models Built by Wine Reviews from Robert Parker and Wine Spectator for 95 + Point Wine Prediction (2022) [40]	Wine Spectator and Robert Parker reviews	Predict if a wine rates 95+ points or 94- points.	Method: SVM. Evaluation: Accuracy, sensitivity, specificity, precision	The combination of normalized values, categorical values and sub-categorical values predicted the class of the wine best.
Consistency of expert product reviews: an application to wine guides (2021) [19]	Wine Enthusiast	Examine wine expert reviewers in wine guides.	Method: Clustering. Evaluation: ANOVA	Given model performance, the conclusion turned out to be that experts are not very consistent with their reviews.
Wineinformatics: Can wine reviews in Bordeaux reveal wine aging capability? (2021) [15]	Wine Spectator	Assess if expert reviews on olfactive and gustative perceptions of wine reveal the suggesting holding time for a bottle of wine.	Methods: Naïve Bayes, SVM and KNN. Evaluation: Accuracy, Recall, Precision, and F-score	KNN and Naïve Bayes were capable of modelling the relationship between these variables with a +70% accuracy.

Continues in the next page

Table 2.1 – Continuation from previous page

Document	Dataset	Research goal	Method and Evaluation	Results
Sentiment analysis based on food e-commerce reviews (2021) [41]	NA – costumer information of multiple online wine shops	Propose a model to capture relevant words to a certain degree, to perform emotion detection on user reviews.	Methods: TF-IDF, Word2Vec, Attention mechanism, BiLSTM, CNN. Evaluation: Accuracy, Recall, F-Score	A model comprised of several techniques proposed on the paper yield better results than BiLSTM and CNN and is able to classify sentiment on the corpus.
Wineinformatics: Using the full power of the computational wine wheel to understand 21st century bordeaux wines from the reviews (2021) [42]	Wine Spectator	Improve previous research with the inclusion of continuous variables instead of dummy, with the help of CWW, to predict grade category.	Methods: SVM, Naïve Bayes. Evaluation: MAE, Accuracy, Recall, Precision, F-score	It was possible with both models to successfully predict 90+ and 89- grades. The utilization of the category and sub-category of CWW prove to be an enhancement with respect to previous research.
Understanding 21st century bordeaux wines from wine reviews using naïve bayes classifier (2020) [16]	Wine Spectator	Predict if a wine rates 90+ or 89- category via Naïve Bayes and SVM and spot the words most predictive of wine quality.	Methods: SVM, Naïve Bayes. Evaluation: Accuracy, Recall, Precision, and F-score	Both models performed closely to each other, and it was possible to identify preferable characteristics for 90+ wines.

Continues in the next page

Table 2.1 – Continuation from previous page

Document	Dataset	Research goal	Method and Evaluation	Results
Discovering the language of wine reviews: A text mining account (2019) [23]	Wine Enthusiast	Predict wine characteristics (grape, color, e.g.) from reviews and investigate the relationship between price and ratings.	Method: SVM. Evaluation: F-Score	Wine experts share a common vocabulary which allows to predict color, grape, and origin country. It was also found the length of the review and average word length show a significant relationship with price and rating.
Wineinformatics: Regression on the grade and price of wines through their sensory attributes (2018) [17]	Wine Spectator	Predict the quality and price of wine	Method: SVM. Evaluation: ME, MAE, MSE, RMSE, Normalized Mean Absolute Error (NMAE), Normalized Root Mean Squared Error (NRMSE), Accuracy	It was possible to predict the grade with little error, but the price was not successfully predictable, resulting in a difference of $13.

Continues in the next page

Table 2.1 – Continuation from previous page

Document	Dataset	Research goal	Method and Evaluation	Results
Wineinformatics: A quantitative analysis of wine reviewers (2018) [20]	Wine Spectator	Test if expert wine-reviewers from Wine Spectator are reliable and consistent with their reviews. Using these reviews, it was assessed which textual features made a wine be given a 90 or higher ranking.	Methods: Naïve Bayes and SVM. Evaluation: Accuracy, precision, recall, specificity	The SVM model scored a 87% accuracy which proves the consistency and prestigious standing of Wine Spectator's reviewers in the wine industry.
Exploring Emotions on Wine Websites: Finding Joy (2018) [43]	NA	Identify which of the 5 emotions are present in blogs and from these which are the ones with more weight in the overall sentiment of the website	Method: IBM Watson. Evaluation: R2, R2 Adj, RMSE	Joy is the most mentioned sentiment.
A picture is worth a thousand words: Translating product reviews into a product positioning map (2017) [44]	Wine Enthusiast	Combining ontology-learning-based text mining and psychometric techniques to migrate online reviews into a product positioning map.	Evaluation: Mean Absolute Percentage Error (MAPE)	It was possible to generate a taxonomy for wine speak and a critics' general propensity to use topics in their wine reviews.
				Continues in the next page

Table 2.1 – Continuation from previous page

Document	Dataset	Research goal	Method and Evaluation	Results
A Wineinformatics Study for White-Box Classification Algorithm to Understand And Evaluate Wine Judges (2016) [21]	Wine Spectator	Evaluate the consistency of wine judges - the more consistent the wine judge, the higher performance classification models have	Methods: Decision tree, association rules, KNN, Naïve Bayes, SVM. Evaluation: Accuracy	Naïve Bayes and SVM models are the best at this predicting exercise.
Wineinformatics: Uncork Napa's cabernet Sauvignon by association rule based classification (2016) [18]	Wine Spectator	Predict the quality of the wines through attributes extracted from wine reviews, with an emphasis on showing the correlation between the attributes of region-specific wines and their quality	Method: Association rules. Evaluation: Accuracy and Coverage	Satisfactory accuracy results with high coverage scores, reveal that region-specific wines will yield greater predictability given their inherent similarities. Also, the reviews revealed common rules of association among wines, which proposes insights as to common ground shared by wines in terms of qualitative attributes.

Continues in the next page

Table 2.1 – Continuation from previous page

Document	Dataset	Research goal	Method and Evaluation	Results
Sentiment Analysis of Wine Aroma (2016) [45]	NA	Classify sentiments expressed on wine reviews, more specifically on its aroma.	Methods: SVM, Vectorisation of documents, Feature selection. Evaluation: Precision, Recall, F-measure, Accuracy	Aroma was categorized into 3 categories, and the model classification matched well with manual classification. The performance was improved with feature selection.
Understanding the wine judges and evaluating the consistency through white-box classification algorithms (2016) [22]	Wine Spectator	Understand wine judges and evaluate their consistency while scoring a wine 90+ or 90-.	Methods: Decision trees, Naive Bayes, KNN. Evaluation: Accuracy	Naïve Bayes scored the best results in assessing judges' consistency.
The Computational Wine Wheel 2.0 and the TriMax Triclustering in Wineinformatics (2016) [46]	Wine Spectator	Propose the CWW 2.0 and show it is possible to group similar wines based solely on their reviews.	Methods: Clustering	TriMax clustering - designed specifically for Wineinformatics - produced promising and cohesive results and these can be used to, for example, define plate grouping or wine searching.
Continues in the next page				

Table 2.1 – Continuation from previous page

Document	Dataset	Research goal	Method and Evaluation	Results
Very Quaffable and Great Fun: Applying NLP To Wine Reviews (2016) [47]	Wine Enthusiast	Predict wine properties based on writings from experts.	Method: SVM. Evaluation: F-Score, Precision, Recall	Using NLP techniques, it was possible to produce classifiers that could predict color, grape variety, price, and country of origin with satisfactory F-Scores.
Predicting And Visualizing Wine Characteristics Through Analysis Of Tasting Notes From Viewpoints (2015) [48]	Wine Enthusiast	Visualization of characteristics of wines based on sentiment analysis	Method: SVM. Evaluation: F-measure	Radar charts which tell us about the most prominent sentiments on wine reviews, per region.
Wineinformatics: Applying data mining on wine sensory reviews processed by the computational wine wheel (2014) [38]	Wine Spectator	Introduce a new data science area: Wineinformatics, performed clustering and predicted whether a wine is scored 90+ or 89- based on savory reviews.	Method: Clustering, association rules.	When using review attributes, it is possible to make clusters based on similarities of region, price, vintage, type, and variety. Also, it is possible to predict the wine grade range based on reviews.

In addition to the above results, the query returned another entry that does not fit the summary structure proposed above, since it is a book: *Wineinformatics: A New Data Science Application*. This is a compilation of techniques used and discoveries from previous research by Bernard Chen and is a useful guide for new research projects, but also a valuable handbook to pave the subject's way [14].

Regarding the theme approached in the retrieved papers, there is an eminence of very particular matters which reflects their origin and authorship. Hence, the metrics and conclusions might be similar across papers, reducing the diversity scope of the present review, given the limited research in this specific area.

As is evident from the above summary table, there was no relevant literature focused on Portuguese wines, which makes room for an application as such, since other geographic-specific datasets have already been used for these types of studies, this is the case for Napa Valley wines in California, studied by Chen *et al.* [18]. Additionally, other specific datasets were also utilised, an example of which is the 21st century Bordeaux wines [16] which focusses not only on a type of wine but also on a specific timeline. Moreover, from the known sources, all the above rely on datasets with expert reviews, opening space for new research settings.

From the above table, there was one that claimed the use of customer-generated content. Although the online stores in question are not disclosed, user feedback is considered a good source for sentiment analysis and their reviews are considered a good proxy for the quality of the wine since these customers are more sensitive to the taste of wine given their recurrent purchases of particular wine labels, making this data very valuable to manufacturers [41].

Another hint we might get from this literature is the still very plausible discussion between the reliability of expert reviews. As previously mentioned, there are a lot of articles from the same group of academics, which may accentuate a given conclusion to be more acceptable, when contradictory research has been written by very few others. Apropos of limitations, it might be worth noticing the CWW was brought up from reviews written by experts and we can question if there are some vocabulary differences between experts and consumers, if one wishes to use UGC as a subject.

As can be seen in the above table, the authors often resort to Wine Spectator and Wine Enthusiast, and there are some references to Robert Parker's reviews. Regarding Wine Spectator, the magazine's wine sensory data is widely used since it is considered to be a pioneer in the wine culture due to the extensive online wine reviews which are seen as consistent and reliable [46]; Wine Enthusiast follows the same reputation, as it hosts a substantial structured catalog with information about each wine such as "the producer, region and country, grape variety, color, alcohol percentage, price, and where to buy it" [47]; lastly, Robert Parker is described as remarkable, influential, and contemporaneous, with great impact in the wine industry and for that, his reviews are considered equally relevant for these studies [40].

2.4. Summary

Wineinformatics is a new field of data science that focusses on knowledge about wine, adding wine reviews to the established used of physiochemical laboratory data, making the use of them the differentiator factor of the subject [42].

This emerging study subject has registered mainly classification problems regarding grade/ rating category prediction, addressed by the common usage of models such as SVM and Naïve Bayes. However, new studies reveal other successful model implementations and there has been some discussion and encouragement around the utilization of algorithms such as Neural Networks – which may become more popular in the future of Wineinformatics [29, 39].

Despite all the findings, there are still some concerns about consistency and true value of expert reviews, which raises doubts and opens room for exploring UGC in similar settings and predictive scenarios. This is one of the main lacking subjects in the literature revised: the vast majority uses only expert reviewers to retrieve information about the wines analysed.

Following the same reasoning, literature suggests the importance of online reviews, recognising that consumers usually read these, demonstrating trust in previous buyers' feedback. In fact, UGC is considered to provide relevant detailed information which is helpful for future consumers comprehend the product quality [49]. Considering this, one can question if consumer reviews are interesting targets for sentiment analysis and related tasks, given their relevance combined with their accessibility, as they are often displayed below the product purchasing panel.

Taking all this into account, we see as a promising future research agenda the focus in UGC in public websites, such as Vivino. Using the various methods found or other innovations, this type of study may reveal some new valuable insights to the wine industry or, at least, confirm known trends or preferences. For that, it might be interesting as well to apply to particular regions and retrieve better and more conclusive insights for the Wine Industry stakeholders.

To conclude, we can finally answer the proposed Research Question, in the introduction of this chapter: "Are there studies on Text Mining techniques applied to Wine Reviews?". We can affirmatively respond to it, given the available literature on Text Mining techniques applied to Wine Reviews. However, when we reduce the scope to Portuguese Wine Reviews, there is no existing literature, to the best of our knowledge.

CHAPTER 3

Data

Given the purpose of this study, public reviews from Vivino were extracted to serve as its base. The Vivino API was used to obtain relevant data: this includes the textual reviews and other enriching fields. Below, Table 3.1 depicts the information extracted for each review.

Table 3.1: Collected data for each record

Field	**Description**
Winery	The name of the winery of which wine is being reviewed
Region	The name of the region of which wine is being reviewed
Country	The name of the country of which wine is being reviewed
Year	The year of which wine was made
Wine ID	The wine unique identifier
Wine Name	The wine name
Grapes	The grapes of which the wine was made
Average Rating	The average of the ratings that wine received
Review Count	The number of reviews that wine received
Price Amount	The price of the wine
Currency Code	The currency in which the price amount is expressed
Is Natural	If the wine is natural or not
Acidity	The acidity level inferred from existing reviews
Fizziness	The fizziness level inferred from existing reviews
Intensity	The intensity level inferred from existing reviews
Sweetness	The sweetness level inferred from existing reviews
Tannin	The tannin level inferred from existing reviews
Review ID	The review unique identifier
Rating	The rating given by the user with the written review
Note	The textual review
Created At	The date in which the review was published
Language	The language in which the review was written

The extraction took place during the month of October, 2023. It resulted in 517,957 entries, dating from 2011 to 2023.

3.1. Data Analysis

Before any further exercises, a preliminary data understanding was performed. For instance, the existence of duplicate records and missing values was studied. This resulted in the removal of all duplicates as well as those variables with a significant amount of missing values, to it: Grapes, with 100% of missing values; Fizziness, with 99% of missing values; and Tannin, with 35% of missing values.

In the first place, we looked at the distribution of these reviews throughout the years, as Figure 3.1 mirrors. We are able to spot a clear growth until 2019, which may depict Vivino as a platform and Portuguese wines popularity crescendo. While during COVID-19 the numbers have maintained, we can spot the reduction of reviews in the last couple of years.

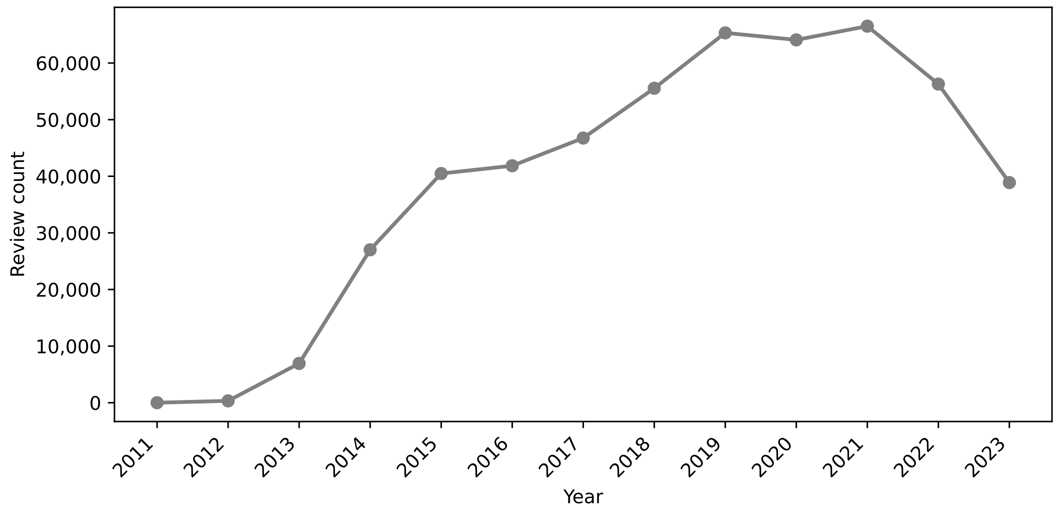

FIGURE 3.1. Review count over the years

Turning our attention to the ratings which accompanied these written reviews, depicted in Figure 3.2, one can clearly understand the tendency towards greater ratings rather than lower ones.

As seen by the previously presented literature, price is one common point of analysis regarding wines. Regardless of the existence of extremely pricey wines, the majority are below the 50 EUR mark, as Figure 3.3 exhibits.

At this point, with 510,018 non-null records, we took interest in the most common written languages. As Figure 3.4 shows, those were Portuguese and English. From the sample above, the English language remained the one of interest. This was due to technical limitations. Although some of the used tools are not able to process multiple languages, all of them, even the ones capable of doing it, present more robust and better trained models for English.

Afterwards, we investigated what was assigned as an unknown language. We calculated the character length and word count of the reviews. Then, we found that the unknown records were more prevailing on the left-hand side of these distributions, as

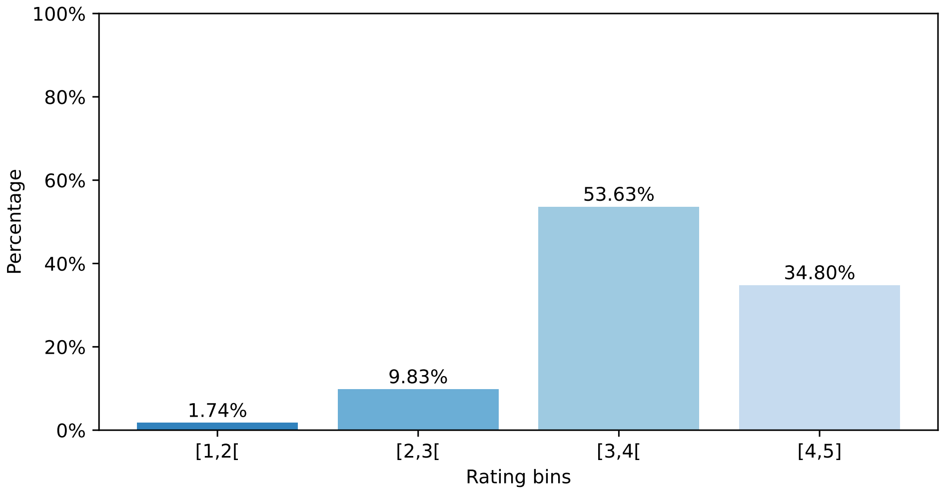

FIGURE 3.2. Review distribution per Rating bin, in percentage ($n = 510,018$)

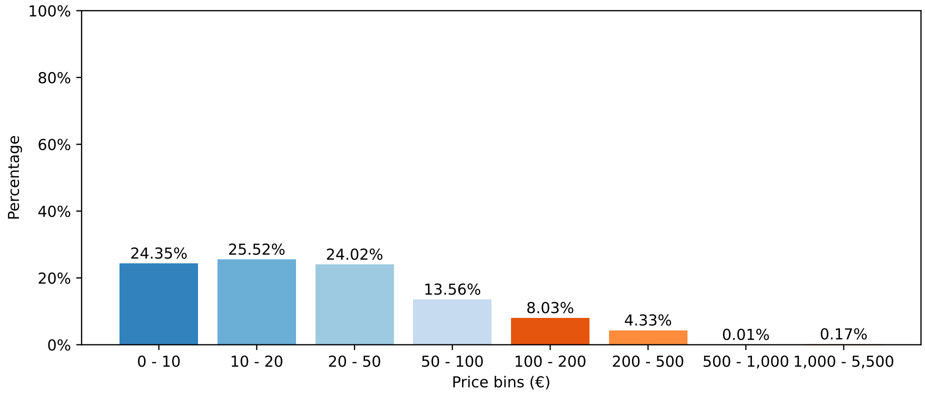

FIGURE 3.3. Review distribution per Price bin, in percentage ($n = 510,018$)

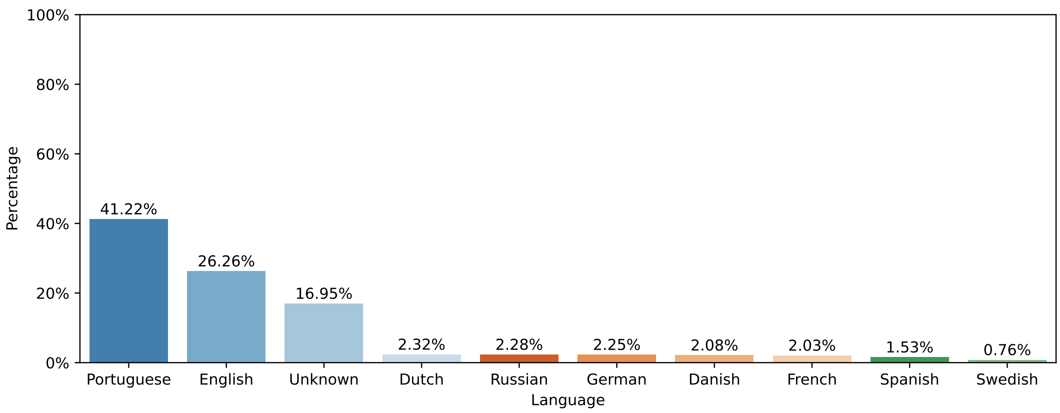

FIGURE 3.4. Top 10 languages with most reviews, in percentage ($n = 510,018$)

shown by Figures 3.5 and 3.6. Given the study of their contents, these records were all discarded since no language could be identified due to their extremely short nature, use of acronyms and emojis, typos, or special characters.

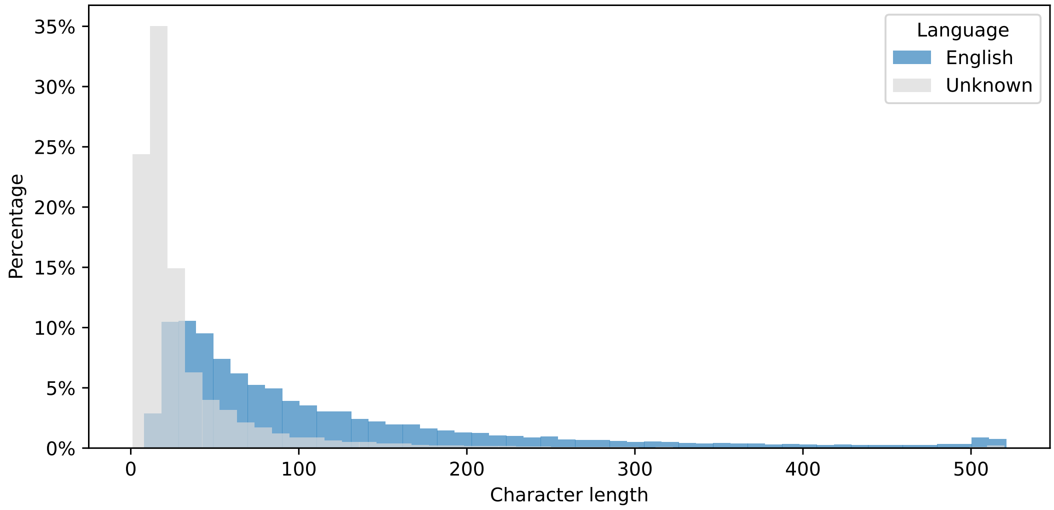

FIGURE 3.5. Review character length distribution per language, in percentage ($n_{Eng} = 133,924, n_{Unk} = 86,440$)

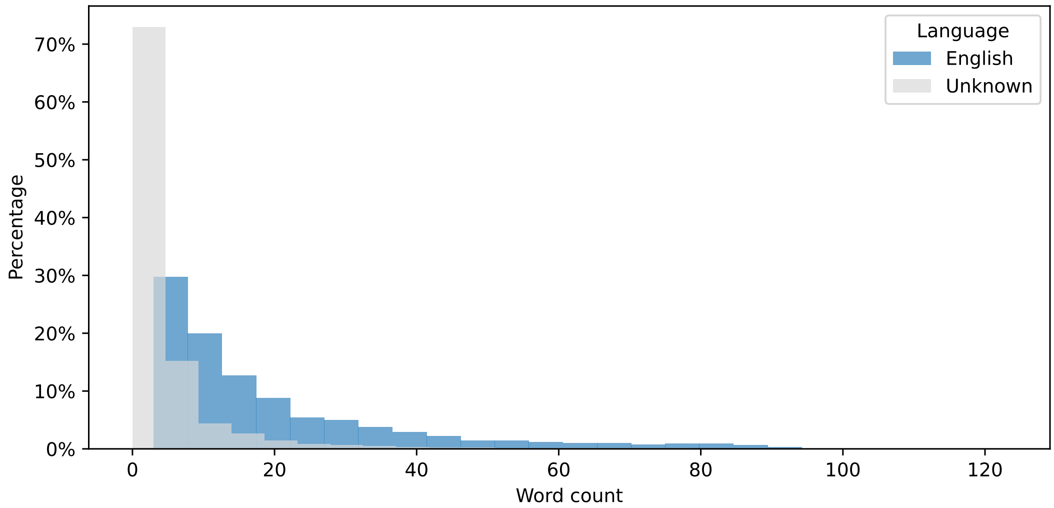

FIGURE 3.6. Word count distribution per language, in percentage ($n_{Eng} = 133,924, n_{Unk} = 86,440$)

Regarding reviews in English, these were filtered to include only the ones with at least three words. Then, in preparation for the upcoming study, an English dataset was created with 129,673 records.

3.2. Data Preparation

Preprocessing is an essential step in improving data quality. This entangles the removal of any inconsistencies in the dataset [50]. In that pursuit, the below tasks were performed. In order to enrich the analysis, two preprocesses were constructed in parallel. They are as follows, in the next sections.

3.2.1. Basic Preprocessing

- Removal of emojis
  Since one of the libraries further discussed does not process emojis and the majority of reviews did not contain them, these were removed to reduce noise in the inconsistent use cases of said characters and to simplify the comparison between utilized libraries.
- Removal of newline character
  This character, represented as "\n", just shows the user performed a line break and it carries no meaning to the exercises about to be performed. In fact, its presence may inadvertently disrupt data. Hence, it was removed.
- Removal of punctuation
  This function will remove any character that is not alphanumeric, a white-space, or an underscore. Unless these special characters are removed, they may cause words to return unavailable in the dictionary [50].
- Removal of stop words
  Words such as articles, prepositions, conjunctions and others of the same sort are to be discarded as they do not account for any valuable meaning [50].
- Text normalisation
  This process assures consistency in treatment between capitalized and non-capitalized words and reduces dimensionality and variability.
- Reduce repetitions
  Repeated letters, like in 'gooood', are used to express intensity [50]. Given that they are not matching dictionary words, the repetition of two letters is the maximum allowed.

For these tasks, several Python libraries were employed. Firstly, *re*[1] library was used for regular expressions, which facilitated the removal of punctuation and excessive character repetitions. Additionally, *nltk*[2] was leveraged to eliminate English stopwords, ensuring that common, uninformative words were excluded. Lastly, for emoji removal, the *demoji*[3] library was applied, further cleaning the text.

3.2.2. Lemmatisation

On top of the previous processing, reviews were lemmatised. This corresponds to the process of reducing the inflected parts of a word to its lemma, which is the form of a word registered in a dictionary [51]. Using spaCy's lemmatiser[4], the root word is provided, conserving the conveyed meaning in a single form, as the following examples show:

- words → word

[1] https://docs.python.org/3/library/re.html
[2] https://www.nltk.org
[3] https://pypi.org/project/demoji/
[4] https://spacy.io/api/lemmatizer

- corpora → corpus
- better → good
- best → good

These two parallel preprocessed corpora were object of the next described tasks.

CHAPTER 4

Extracting Knowledge from User Reviews

This chapter presents the study findings. First, we introduce the methods utilized in each further section. Then, we present Sentiment Analysis results, uncovering reviews' polarity. Afterwards, the Topic Modelling task exposes the topics discussed in the reviews, followed by a section which intertwines the latter tasks in order to analyse overall sentiment regarding each of the retrieved topics. Later, we take a deeper dive on the first task, assessing which of the five basic emotions is the most present in the studied reviews. Finally, we embark on one of the most common tasks among the related literature: review classification, categorising our reviews through their text into different rating classes.

4.1. Methods Employed

Figure 4.1 illustrates the process adopted. The framework consists of four parallel tasks: Sentiment Analysis, Topic Modelling, Emotion Detection, and Review Classification. In a second layer, we crossed the first two tasks by inferring topics by sentiment polarity.

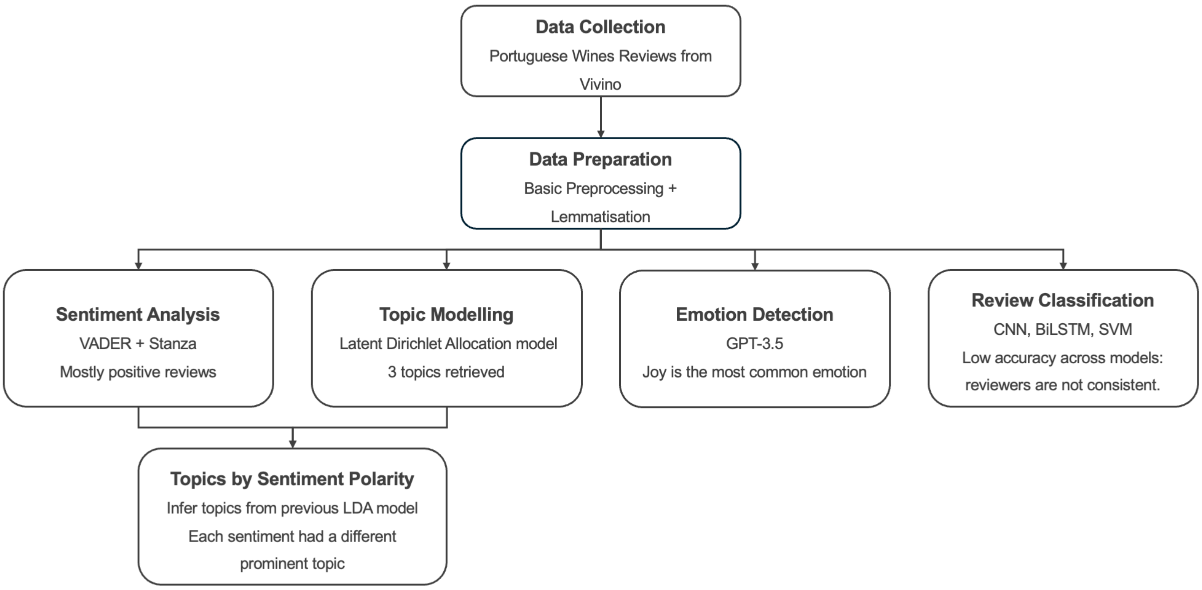

FIGURE 4.1. Illustration of the research process adopted

In the next sub-sections, we present each method employed in the study.

4.1.1. Sentiment Analysis

Sentiment Analysis is a well-established NLP approach to commonly determine the polarity (negative, neutral or positive) of opinions expressed by users through their reviews on social networks [52].

When employing sentiment analysis, there are two main approaches to extract the sentiment and classify it: machine learning or lexicon-based. As proposed by Bonta, Kumaresh, and Janardhan [53], the first type is used to build algorithms and, consequently, models by feature selection over labelled datasets, which implies a previous manual labelling preprocessing and heavy training of the domain-specific data. On its turn, lexicon-based approaches are not domain-dependent and do not require labelled datasets since they recur to word lexicons where words or sequences of them are associated with a respective sentiment score [54].

In the present study, we used Stanza and VADER. Stanza is a deep learning-based library designed to be both flexible and comprehensive. Stanza leverages pre-trained NN models that capture complex language features, enabling the detection of sentiment in texts of varying lengths and complexities. For sentiment analysis, Stanza classifies texts into polarity categories through a pipeline of tokenisation, part-of-speech tagging, lemmatisation, and dependency parsing, leveraging word embeddings to capture semantic nuances [55].

VADER [56] was specially tailored to the context of social media. It is a rule-based tool relying on lexicons with valence scores reflecting the sentiment intensity. Words with stronger positive connotations have higher positive scores, while those with stronger negative connotations have lower negative scores. To gauge sentiment polarity, VADER uses five key rules to adjust sentiment intensity:

- Punctuation, particularly the exclamation point (!), which increases the intensity of the sentiment without altering its underlying meaning.
- Capitalisation, especially when using ALL-CAPS to highlight a sentiment-relevant word among other noncapitalised words, amplifies the intensity of the sentiment without changing its underlying meaning.
- Degree modifiers, also known as intensifiers, booster words, or degree adverbs, influence sentiment intensity by either amplifying or diminishing it (e.g., fairly, rather, quite).
- The contrasting conjunction "but" indicates a change in sentiment polarity, with the sentiment expressed after the conjunction being the more dominant.
- Negation, by analysing the trigram that comes before a sentiment-bearing word, reversing the text's polarity.

4.1.2. Topic Modelling

Topic Modelling is a popular NLP task based on statistical machine learning to discover latent topics in a document collection. Algorithms such as Latent Dirichlet Allocation [57] are usually employed in this pursuit. The idea behind LDA models is that texts contain multiple topics in different proportions, which can be characterised by the use of specific vocabulary [10].

In the present work, we have used LDA to extract the main topics discussed in the wine reviews. The algorithm automatically identifies different lists of words in the text collection and annotates individual documents with probabilistic topic labels [10]. Statistically, each document is represented by a distribution over a number of topics, and each topic is defined by a distribution over a specific vocabulary of words. So, by aggregating topic probabilities across thousands of documents based on metadata, one can, for instance, understand which topics are prevalent within that context [10].

While creating the model, one should decide how many topics are to be discovered. Here, we have iterated through some options and selected three because it returned the highest record of semantic coherence and had the clearest interpretation from a human point of view. This measure, literature suggests, evaluates the interpretability of topics learned [54].

The present LDA model was implemented with the help of the *gensim* [58] module in Python.

4.1.3. Emotion Detection

Emotion detection is the process of identifying a person's emotions like joy, sadness, or surprise. Although some physical aspects such as heart rate, hand shivering, sweating, and voice pitch have been proven to convey an emotional state, detecting emotions in text is still a rather difficult task [59].

Beyond sentiment analysis, emotion detection is a natural additional task since emotions are finer-grained information extracted from opinions and as valuable as sentiment polarity [60]. Several techniques have been used in this pursuit, as discussed by Nandwani and Verma [59], Boitel, Mohasseb, and Haig [61] or Carneros-Prado *et al.* [62], for example. However, in this study, we have used GPT-3.5, which has shown strong performance in various contexts, particularly in tasks related to sentiments and emotions [63].

For it, we utilised OpenAI's API to query ChatGPT. We prompted the 'gpt-3.5-turbo' model to analyse each wine review and identify the predominant emotion, given the options: sadness, joy, anger, fear, or disgust. In case no emotion was detected, it would return none.

4.1.4. Review Classification

For the last experiment, we have focused on the information retained in the wine reviews to compare the performance in the classification tasks of two NN models plus an SVM.

Such an endeavour was very present in the Literature Review, as discussed in Chapter 2. This task was commonly carried out in an attempt to evaluate the consistency of wine judges [16, 19–22, 29, 39, 40, 42]. Following these, our objective was to evaluate the consistency of our UGC, using some of the techniques used in the reviewed studies: CNN, BiLSTM, and SVM.

All three models addressed two-class and four-class text classification tasks, as seen in Katumullage *et al.* [29]. Adding to this, we applied these models to three variations of our dataset, given the present class unbalance already exposed (see Figure 3.2). These will be the unbalanced (original), the undersampled, and the oversampled versions.

Class unbalance entails skewness in the training data, leading to increased difficulty in classification [64]. As such, several techniques are available to overcome this, but common approaches to solve it are undersampling, reducing the sample size, or oversampling, enlarging the sample size. Both approaches have their drawbacks: undersampling will probably disregard useful data, whereas oversampling can increase the likelihood of overfitting [64]. In an attempt to study the effects and results of both solutions, we applied both strategies and compared their performance with the original unbalanced dataset. For undersampling, we utilised RandomUnderSampler [65], and for oversampling, we applied Synthetic Minority Oversampling Technique (SMOTE) [65], both from *imblearn*[1] Python library.

Regarding CNN, this is a type of deep neural network originally developed by LeCun *et al.* [66] for image analysis. The architecture of our CNN model is designed to capture essential features from text sequences through convolution layers, followed by a fully connected network for classification. We first tokenise and pad the sequences. Each text sequence is padded to a fixed length of 100 tokens to maintain consistency across samples.

The model architecture consists of the following components:

- Embedding Layer: The input sequences are mapped to dense vectors of fixed size. This layer learns word embeddings during the training process, capturing semantic relationships between words.
- Convolutional Layer: A 1D convolutional layer is applied to extract local features from the embedded sequences. The ReLU activation function is used.
- Pooling Layer: A max-pooling layer with a pool size of 4 is used to reduce the dimensionality of the feature maps, highlighting the most prominent features and reducing computational costs.
- Flatten and Fully Connected Layers: Output of the previous layer is flattened and passed through a dense layer, followed by a dropout layer (with a 0.5 rate) to prevent overfitting.
- Output Layer: The final dense layer uses a softmax activation function, producing probability distributions over the target classes.

The model was trained for five epochs and, to boost robustness and follow the example set by Katumullage *et al.* [29], we employed 10-fold cross-validation.

As with the CNN model, with BiLSTM the input text data undergoes tokenisation and padding. Ensuring uniformity across samples, as previously done, we encode the class labels using an encoder to transform categorical classes into numerical values which are subsequently fed into the model for classification.

[1] https://pypi.org/project/imbalanced-learn/

The model architecture consists of the following components:

- Embedding Layer: This layer transforms the tokenised input sequences into dense vectors of fixed dimensionality, similarly to the CNN, and learns these embeddings during training, capturing the semantic relationships between words.
- BiLSTM Layer: The core component of the model is this layer. By processing the input sequences in both forward and backward directions, this layer is able to capture both past and future dependencies in the text.
- Fully Connected Layers: After the LSTM layer, the model includes a dense layer with the ReLU activation function, followed by a dropout layer (with a rate of 0.5) to mitigate overfitting.
- Output Layer: The final layer uses a softmax activation function to output probabilities for the target classes.

The model was trained with a 10-fold-cross validation for five epochs, and for each fold the BiLSTM was retrained, avoiding retained bias from previous folds.

SVM is considered the most famous and successful black-box classification algorithm across various application areas [20]. Unlike the NN models, SVMs require explicit feature extraction from text. For that, we utilised TF-IDF vectorisation to convert text data into numerical features. TF-IDF is a commonly used technique in text mining, which translates the importance of a word within a document relative to the entire corpus [67]. Once the TF-IDF vectoriser is implemented, this representation is then used as the input of our SVM models.

The model architecture consists of the following components:

- Kernel Choice: We use a linear kernel, suitable for text classification tasks, which helps to create a decision boundary that maximises the margin between different classes.
- Regularisation Parameter: The regularisation parameter $C = 1$ controls the trade-off between maximising the margin and minimising the classification errors.
- Probability Estimates: Although SVM are not inherently probabilistic classifiers, we enabled probability estimation to allow the model to output class probabilities.

As with the other models, we employed 10-fold-cross validation.

4.2. Sentiment Analysis

For this exercise, the previously presented NLP libraries Stanza and VADER were applied to the English-written reviews. These models are commonly used in sentiment analysis [68], and therefore chosen for the present study.

Attribute scores are used to represent sentiment polarity. While Stanza returns only discrete values, VADER, on the other hand, comes in continuous values showing how positive or negative an opinion is. In order to simplify the comparison, the default presentation of both methods was changed and the modified scores shown in Table 4.1 were the ones used. Regarding VADER, the original scores were categorised as follows: scores less than -0.05 were classified as Negative, the ones greater than 0.05 were classified as Positive, and the remaining scores, ranging between -0.05 and 0.05, were classified as Neutral.

TABLE 4.1. Polarity Score Interpretation for Stanza and VADER Libraries

Library	Version	Scores	Interpretation
Stanza	Original	0, 1, 2	Negative, Neutral, Positive
	Modified	-1, 0, 1	Negative, Neutral, Positive
VADER	Original	[-1, 1]	Most Negative to Most Positive
	Modified	-1, 0, 1	Negative, Neutral, Positive

Once the sentiment polarity was assessed for both Basic Preprocessing and Lemmatisation, results from both libraries were compared. They were in accordance about 80% of the times. Given the very similar results and being VADER tailored for social media data, we will proceed exclusively based on the Lemmatised VADER results.

Firstly, it should be noticed that 79% of our reviews were classified as positive, 18% neutral, and only 3% as negative. Although this does not represent the ideal distribution in terms of class representation, it corroborates our polarity-score results since it exhibits similarities with common finding regarding online reviews.

In fact, studies such as the one by Schoenmueller, Netzer, and Stahl [69] have found a mass prevalence of positive ratings on most online platforms. Moreover, other works have even described it as a positively skewed, asymmetric, bimodal, or J-shaped distribution [70, 71], which resembles the one we get, in Figure 4.2.

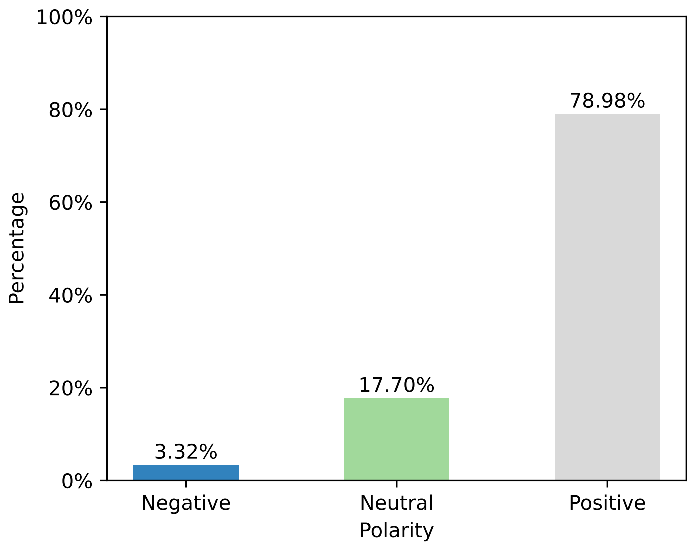

FIGURE 4.2. Review distribution by assessed polarity ($n = 129,673$)

As presented by Schoenmueller, Netzer, and Stahl [69], this phenomena roots to the following possible reasons: (1) compared with offline WOM, writing online reviews takes more effort and consumers may be less likely to report negative experiences; (2) online WOM typically occurs without the sense of strong ties between individuals, and people often hesitate to share negatives with distant social connections; (3) online reviews reach big audiences and when people communicate with larger groups, they are less inclined to share negative information.

Complementing the analysis, we turn our attention to the ratings given along with each review. Figure 4.3 combines the assessed polarity with the rating users can attribute. One can immediately see that positive reviews are associated with higher ratings. In a similar fashion, lower ratings are the most prominent in regard to negative reviews. This, although expected, assures the desired consistency between sentiment expressed on the written reviews and ratings users attributed.

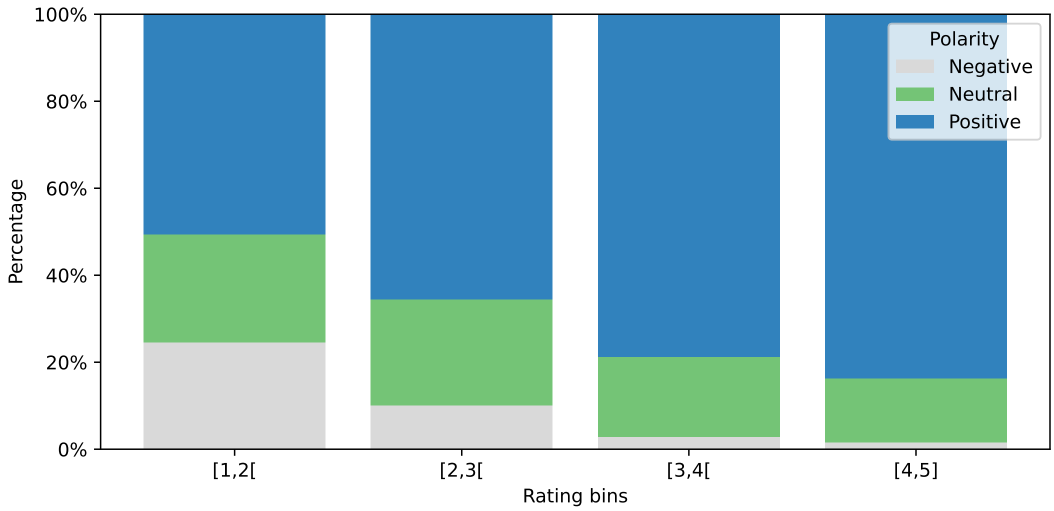

FIGURE 4.3. Sentiment polarity by Rating bins, in percentage ($n_{[1,2[} = 1,934, n_{[2,3[} = 11,651, n_{[3,4[} = 70,317, n_{[4,5]} = 45,771$)

4.3. Topic Modelling

After performing sentiment analysis, we were interested in extracting more meaning from the corpora of reviews. For it, we have developed a topic model solution.

In that sense, we have used the lemmatised text and extended stop words to include the "wine" lemma, since it would not add any meaning to the analysis. In order to represent the topics, the pyLDAvis tool [72] was utilised and it is shown below in Figure 4.4, as an inter-topic distance map.

The present LDA model reached a Perplexity value of -6.88. The Coherence score of 0.58 for the three topics was the maximum of its distribution, shown in Figure 4.5. Lower perplexity values suggest a good probability distribution in predicting the sample. The coherence score assesses quality of the learned topics – the closer to 1 the better.

FIGURE 4.4. Inter-topic distance map

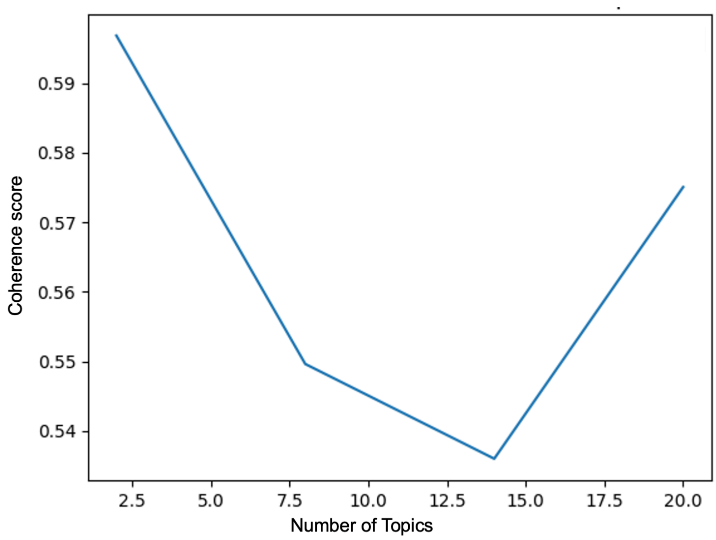

FIGURE 4.5. Coherence scores by the number of topics

Each of the circles returned by pyLDAvis represents a topic, and their area is proportional to the frequency of each topic throughout the corpus [54]. The distance between each group reflects their level of similarity. In that sense, topics that are closer together share a greater number of terms, while those farther apart are more distinct in their composition.

First, Topic 1 shows several terms on the tasting experience and sensory characteristics, as depicted in Figure 4.6. This topic emphasises sensory aspects, particularly related to common tasting notes, aromas, and textures. Terms like "tannin", "dark cherry", "plum", "chocolate", and "vanilla" suggest a discussion around the wine flavour profile and its aroma. At the same time, words such as "smooth", "body", "acidity", and "finish" refer to the texture and overall tasting experience.

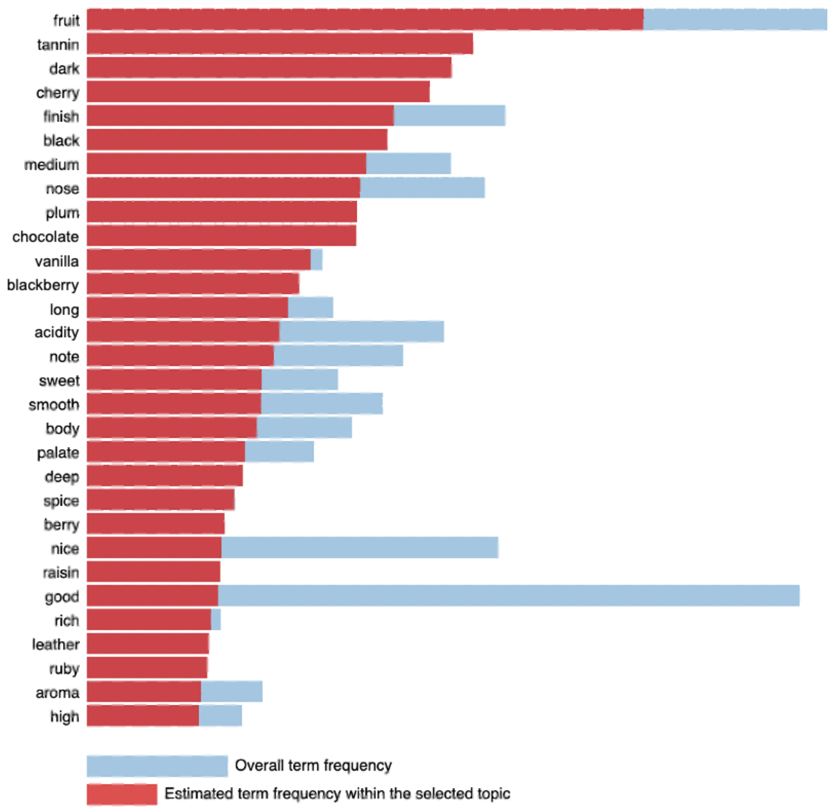

FIGURE 4.6. Topic 1: Wine Tasting Experience and Sensory Characteristics

The terms of Topic 2, shown in Figure 4.7, revolve around the consumer experience and perceived value of the wine. Terms such as "good", "great", "price", "value", and "nice" suggest a discussion of the affordability and quality of the wine. The inclusion of "Portugal", "Douro", "touriga", and "vintage" words indicates the user's acknowledgement of the region or its reputation. In general, this topic points to consumer satisfaction, particularly in relation to the price-to-quality ratio and the origin of the wine.

Lastly, Topic 3 displays terms shown in Figure 4.8 that are related to the wine style, specifically regarding lighter, fresher, and acidic wines. Terms like "light", "fresh", "citrus", "apple", and "mineral" point to the characteristics of certain wine styles, especially white wines or lighter-bodied ones. The focus here turns to the freshness, fruitiness, and crispness of the wine, which coincides with the trends that appeal to consumers lately.

4.4. Topics by Sentiment Polarity

Now that we have sentiment polarity assessed as well as identified the prominent topics discussed in our reviews, it makes sense to combine this information and explore further.

Then, we applied the previous LDA model to infer topic distributions for each polarity in our reviews. We conserved the dominant topic for each review and aggregated

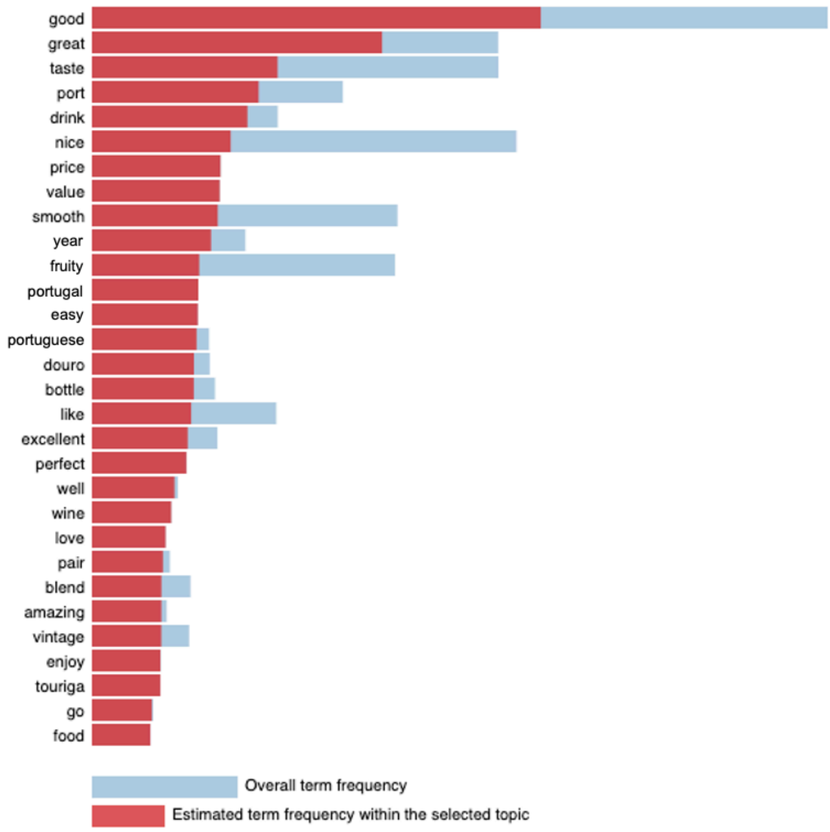

FIGURE 4.7. Topic 2: Wine Value and Consumer Satisfaction Characteristics

that information. Figure 4.9 represents just that: topic distribution for each sentiment assessed.

Immediately it is noticeable that each sentiment has a different topic running the lead. Topic 3 is the most mentioned in positive reviews, and this can be attributed to the pleasure of a refreshing, light, and fruity drink, which is indeed trending among wine drinkers.

However, negative sentiments permeate Topic 2, which was regarding the value of wine and characteristics of consumer satisfaction. This too makes sense as price and other characteristics such as origin could set up expectations on the consumer that if corrupted can motivate the expression of negative feeling towards the drinking experience.

Finally, there is a big high on neutral sentiments around Topic 1, which was revolving the tasting experience and sensory characteristics. This, one could argue, is more of a description and the stating of facts about what are consumers sensing while drinking these wines. As such, no much sentiment surrounds these type of descriptive reviews, with tendency to result in a more neutral tone and, therefore, neither negative or positive sentiments are retrieved.

In order to enrich these results, we went back to the preprocessed text to check originally which bigrams were most common for each bin of ratings. Figure 4.10 shows the results.

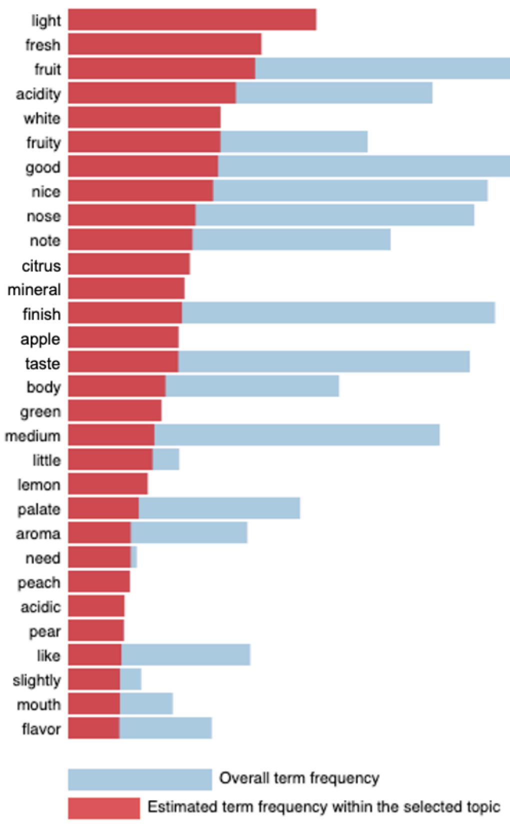

FIGURE 4.8. Topic 3: Wine Style and Freshness

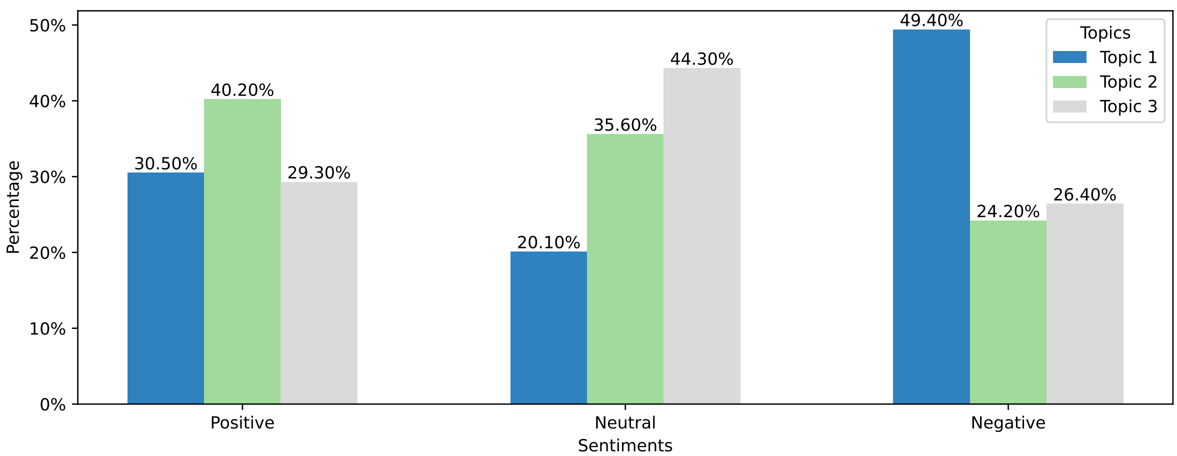

FIGURE 4.9. Topic distribution by sentiment polarity, in percentage ($n_{Pos} = 102,415$, $n_{Neu} = 22,952$, $n_{Neg} = 4,306$)

Regarding the lower quadrants, we can see positive-inclined expressions paired with some tasting notes that probably made consumers enjoy the drink. As we have seen, these better-rated reviews are the most common. However, if we look at the worse ratings, especially focussing in the first quadrant, we can spot something interesting: the common usage of expressions such as "tastes like" or "smells like". This shows that consumers are more recurrent in comparisons when they try to make negative appreciations of a wine they have tasted. As such, this may be signaling that the consumer is not used to a given flavour, rejecting the unpleasant experience which to be described one incurs in comparisons to known flavours, smells, textures, or other ways to convey his experience.

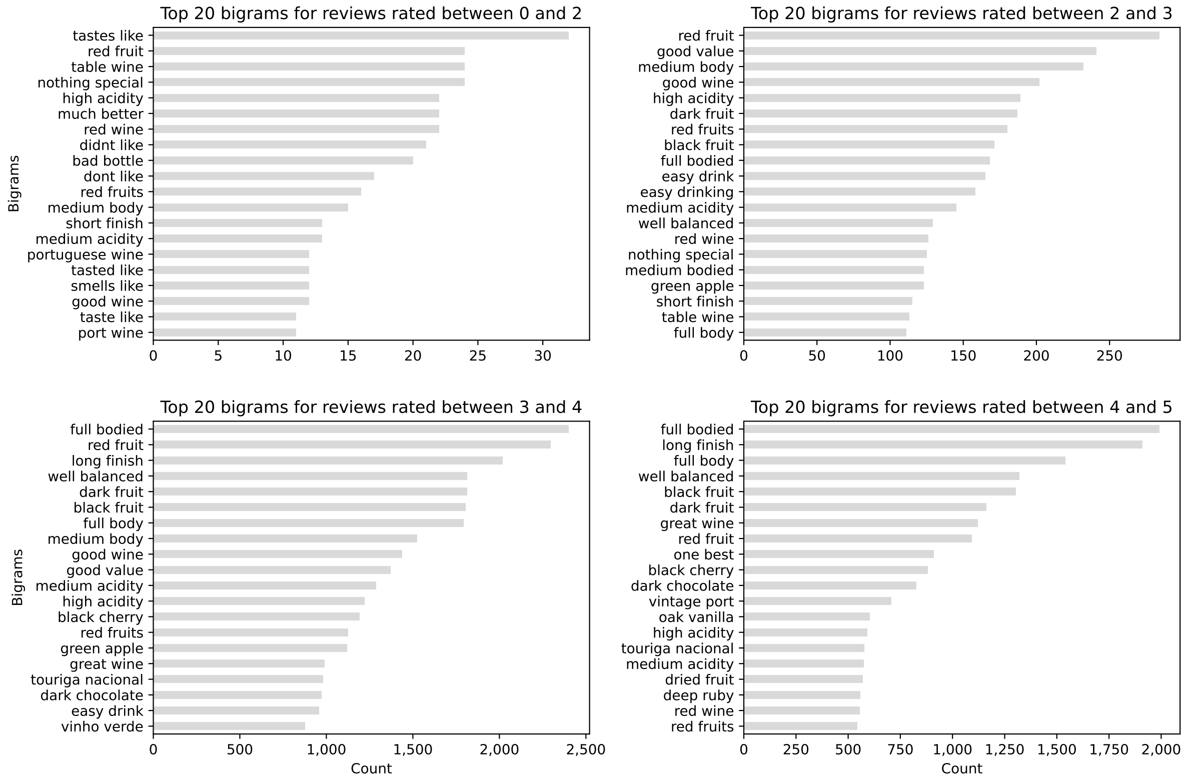

FIGURE 4.10. Top 20 bigrams for each rating bin

4.5. Emotion Detection

Having performed Sentiment Analysis in Section 4.2, the present section aims to further investigate emotions expressed in the reviews, going more in depth than their polarity.

As studied by Treen *et al.* [43] – mentioned among the works analysed in our Literature Review, in Chapter 2 –, here we discover if we are able to retrieve close findings when applying similar Emotion Detection techniques.

While Treen *et al.* [43] recurred to IBM Watson as its platform, using NLP and ML to reveal insights from unstructured data, here we used a generative Large Language Model (LLM). More specifically, we employed GPT-3.5 which, compared to IBM Watson, has achieved competitive performance in emotion classification tasks [62].

As proposed by Treen *et al.* [43], we have focused on five emotions: anger, disgust, fear, joy, and sadness. These are based on Ekman's [73] work, who proposed the previously listed and surprise as the six basic emotions. Although there are no unanimous emotion models, Ekamn's is widely accepted by the emotion recognition community [74]. Hence, we follow using the same group, in order to be able to make a fair comparison to the said article.

As expected, given the results of the Sentiment Analysis exercise, joy was by far the most common emotion assessed in the present task, although approximately 69% of the total reviews did not return an emotion. Still, in respect to negative emotions, disgust was the one with the highest counts, as Table 4.2 below shows.

This goes along with the discoveries of Treen *et al.* [43], who found joy to be the most common emotion detected. However, in their study, sadness was the second scorer and in our case no records were attributed to that emotion. However, a similarity shared with their study is that negative emotions registered a steep low compared to joy, reaffirming wine as a joyous product people desire to share and recommend online.

TABLE 4.2. Emotion Counts

Emotion	**Count**	**Percentage (%)**
joy	40,311	31.086
disgust	83	0.064
fear	13	0.010
anger	2	0.002
unknown	89,264	68.838
Total	**129,673**	**100**

Moreover, even though the present dissertation does not explore the chemo-sensory properties of wines, one can investigate the relationships between these results and the wines' composition through studies such as [75], which compares wine chemical components and sentiments often correlated with them.

According to the study of Tonacci *et al.* [75], which included exclusively red wines, positive and negative emotions were found to be correlated with quantitative and hedonic attributes in their classic sensory analysis. Although wine analysis follows a complex matrix, some compounds were found in the top 25 positively correlated with joy and negatively correlated with disgust, as Figure 4.11 shows.

Albeit more studies must be taken into account before assuming that this correlation means causation, results as such foster the hypothesis that some molecules can be identified as markers capable of eliciting positive or negative emotional reactions [75].

Following Tonacci *et al.* [75], and recurring to Figure 4.11, one can see that some compounds are more positively correlated with disgust and most negatively correlated with joy, usually described by the smells of yeast and overripe fruit. Conversely, some esters were recorded to be positively correlated with joy and negatively correlated with disgust. These are commonly described as sweet, fruity, floral, and rose-like flavours. Furthermore, Tonacci *et al.* [75] state that linalool, limonene, and other compounds characterised by fruity, citrusy, balsamic odor were found to be generally positively correlated with joy and negatively with disgusts.

In some way, these results resemble the bottom quadrants previously presented in Figure 4.10; and now in a similar exercise, we can visualise which unigrams stand out with respect to joy (the only emotion which we have a relevant amount of data), in Figure 4.12. There we see some mentions of fruits, woody odours such as oak, and mentions of

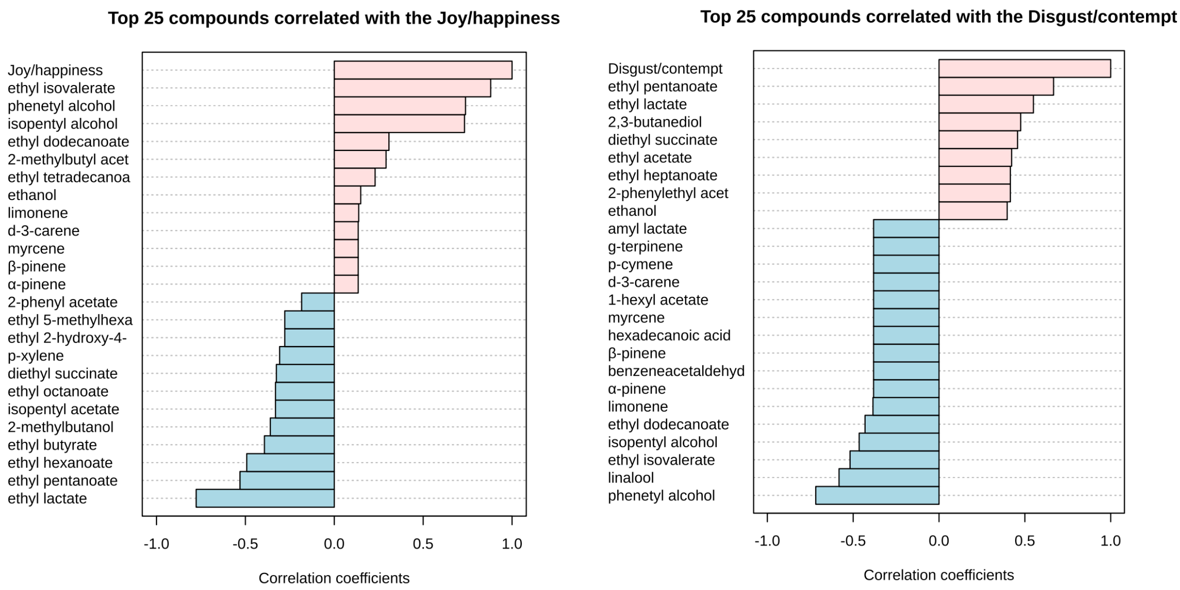

FIGURE 4.11. Top 25 aromatic compounds correlated with joy and disgust, retrieved from Tonacci *et al.* [75]

sweetness, as in sweet and chocolate unigrams. For this, we have extended stop words to include the unigrams "good", "great", "nice", "taste", "well", "wine", "excellent", which were among the most prevalent but accounted for no meaning in the present investigation.

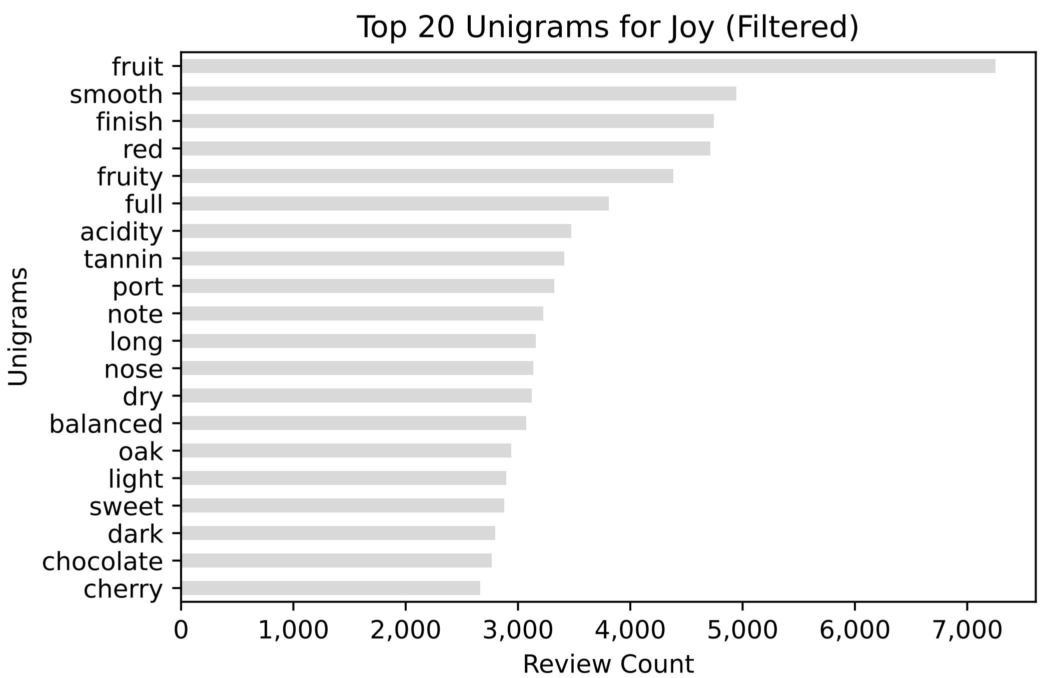

FIGURE 4.12. Top 20 unigrams for the emotion of joy (filtered)

4.6. Review Classification

As stated previously, one of the most common topics discussed in the available literature was the consistency of wine experts when attributing wine ratings.

On this dissertation, we could already prove that, for the data collected, users are most likely to express positive sentiments in their reviews, the topics discussed converge, in summary, into (1) Sensory Characteristics, (2) Wine Value and (3) Wine Style, and that the third topic is the most prominent among positive reviews, while the second is the most common among negative ones. After that, we discovered that joy is the most expressed emotion and that this is suggested to be linked with specific chemosensory attributes.

In order to ultimately understand if the collected textual reviews are consistent in regards to the ratings attributed, we apply two NN models and an SVM.

Moreover, given the differences between rating scales across platforms, expert or not, we adopted the proposal of Szolnoki *et al.* [6] and implemented two versions of rating categories, as Table 4.3 shows.

TABLE 4.3. Proportion of classes for the Two-Class and Four-Class Classifications

Rating Categorisation	Class	Count	Percentage (%)
Two-class	[4,5]	92,594	71.41
	[1,4[	37,079	28.59
Four-class	[4,5]	92,594	71.41
	[3,4[	33,107	25.53
	[2,3[	3,259	2.51
	[1,2[	713	0.55

Regarding metrics to evaluate the models we will present further in this section, we have utilised accuracy, precision, recall and F1-Score. However, as Table 4.3 shows, the count of each class is unbalanced. In that sense, accuracy (the proportion of wines that the classifier has correctly labeled) may not be the most appropriate.

With this in mind, we have produced a classification report[2] for each model version that returned, in addition to accuracy, the other three mentioned metrics by their Macro Average (MA) and Weighted Average (WA). Taking precision as an example, MA Precision calculates precision by class, without weights for the aggregation; meanwhile, WA Precision independently calculates the metric for each class, but when aggregates them applies weights depending on the number of true labels for each category. So, when a model struggles to perform with minority classes, MA will display a bigger penalisation. However, WA favours majority classes.

We have implemented CNN and BiLSTM models with *tensorflow*[3] in Python. For the SVM models, we implemented them using the *scikit-learn*[4] library on Python, as well.

[2] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html
[3] https://tensorflow.org/
[4] https://scikit-learn.org/

Table 4.4 presents the performance metrics for the various models — CNN, BiLSTM, and SVM — across binary and four-class classification tasks. The metrics include Accuracy, Precision, Recall, and F1-Score, with a particular focus on MA for Precision and Recall, aiming at the most accurate evaluation of model performance across classes.

Regarding binary classification task, the MA Precision and Recall metrics reveal more about the models' performance compared to Accuracy, especially when dealing with unbalanced datasets. The SVM model, under unbalanced conditions, shows the highest MA Precision (0.74), while BiLSTM shows MA Precision of 0.70, indicating a more consistent performance in positive classes. This is followed closely by CNN, with MA Precision of 0.68. However, when we turn to Recall metric, the highest values are of the SVM with Oversampling and Undersampling at 0.70.

Indeed, Macro Average metrics are particularly valuable here because they penalise models that do not perform well with the minority classes. Despite the models showing similar or slightly lower Accuracy compared to the highest scorers, their MA Precision and Recall provide a clearer picture of their performance across all classes, highlighting SVM and BiLSTM's robustness in handling both classes better.

In what concerns to four-class results, the classification presented itself as a more challenging task. This is evident by the lower MA metrics across all models, comparatively with two-class results. Nonetheless, SVM under unbalanced conditions achieves again the highest MA Precision, while BiLSTM shows better score at MA Recall.

Overall, these two higher scorers were the ones with the better MA F1-Score results, in both categories. However, the steep drop in performance from binary to four-class classification highlights the increased difficulty of unbalanced multi-class tasks. The struggle with balancing precision and recall across multiple classes is evident from the significant gap between Precision and Recall values, indicating that while the models might correctly identify some classes with high precision, they fail to maintain similar performance across all classes. These lower score results for multi-class classification are in line with the findings of Katumullage *et al.* [29].

Although accuracy is affected by unbalance of the classes, this metric did not improve when undersampling or oversampling were applied to the dataset. This prompt us to some key challenges faced by the models.

This outcome suggests that solely adjusting class distributions may not be sufficient to overcome all challenges posed to these models. Naturally, short reviews often contain limited context, making it challenging for models to discern nuanced sentiments or features that correlate with specific ratings. This can result in a lack of sufficient content for the models to learn from, leading to reduced performance in predicting ratings accurately.

However, this underscores the lack of consistency among Vivino reviewers. As these are informal reviews, no structure is suggested before a user writes their thoughts on the platform nor are there any questions or specific fields influencing the reviews being written. Besides, given the measures we retrieved from the models, it looks like, although

Table 4.4. Accuracy, Precision, Recall, and F-1 Score for each model variation

Model	Accuracy	Precision		Recall		F1-Score	
		MA	WA	MA	WA	MA	WA
2 Classes							
CNN Unbalanced	0.74	0.68	0.73	0.65	0.74	0.66	0.73
CNN with Oversampling	0.72	0.65	0.71	0.64	0.72	0.64	0.71
CNN with Undersampling	0.66	0.64	0.72	0.67	0.66	0.64	0.68
BiLSTM Unbalanced	0.76	0.70	0.74	0.66	0.76	0.67	0.74
BiLSTM with Oversampling	0.74	0.67	0.72	0.65	0.74	0.66	0.73
BiLSTM with Undersampling	0.68	0.66	0.74	0.69	0.68	0.66	0.70
SVM Unbalanced	0.76	0.74	0.75	0.62	0.76	0.63	0.73
SVM with Oversampling	0.71	0.67	0.74	0.70	0.71	0.67	0.72
SVM with Undersampling	0.69	0.67	0.75	0.70	0.69	0.67	0.71
4 Classes							
CNN Unbalanced	0.73	0.46	0.70	0.35	0.73	0.37	0.70
CNN with Oversampling	0.65	0.33	0.67	0.37	0.65	0.32	0.66
CNN with Undersampling	0.45	0.31	0.67	0.45	0.45	0.28	0.52
BiLSTM Unbalanced	0.74	0.49	0.71	0.36	0.74	0.38	0.71
BiLSTM with Oversampling	0.67	0.35	0.70	0.39	0.67	0.34	0.68
BiLSTM with Undersampling	0.49	0.31	0.68	0.46	0.49	0.29	0.55
SVM Unbalanced	0.74	0.64	0.71	0.30	0.74	0.31	0.68
SVM with Oversampling	0.62	0.35	0.72	0.48	0.62	0.36	0.66
SVM with Undersampling	0.49	0.32	0.69	0.49	0.49	0.30	0.56

the overall shortness of the reviews, users are not consistently using the same terms or evoking the same flavours, senses or odors for the same ratings they attribute. This causes models to have a depreciated performance in comparison to the ones trained on expert texts.

This fact, faced to the similar studies available regarding expert consistency, suggests much less congruence among reviewers in Vivino than other websites. Wine Spectator, for its turn, has received an accuracy evaluation as high as 87.21%, with an SVM model, [20], and the magazine returned even higher scores when applying CNN and BiLSTM models: 88.02% and 88.69%, respectively [29].

CHAPTER 5

Conclusions and Recommendations

This study analysed online reviews of Portuguese wines from Vivino using sentiment analysis, topic modelling, emotion detection, and classification.

In addressing the research questions posed in Section 1.3, several key insights were uncovered. Firstly, the most common sentiments expressed towards Portuguese wines in the reviews were predominantly positive, with joy emerging as the most frequently detected emotion.

Secondly, three main themes were identified as drivers of consumer preferences: sensory characteristics, value perceptions, and style preferences. Sensory aspects, such as taste and aroma, were consistently mentioned as influential factors in shaping positive reviews, whereas concerns about price and quality often led to negative feedback.

Lastly, the Machine Learning (ML) models employed demonstrated modest effectiveness in predicting wine ratings based on the textual content of reviews. While they were able to capture certain patterns within the review that correlated with specific ratings, the accuracy remained moderate.

As introduced by the above results, our findings reveal interesting insights that can benefit the wine industry, particularly in the areas of product development and marketing strategy.

5.1. Implications for Wine Industry

One of our primary discoveries regarding consumer preferences is that they focus on the sensory characteristics of wines, such as flavour, aroma, or texture, when leaving a positive review. However, the study shows that consumers frequently comment on the value and price-quality ratio of wines, mainly when giving a negative review.

As such, wine producers can capitalise on this by refining product lines that emphasise sensory qualities that have been well reviewed and trending (e.g., light, fresh, citric wines), thereby aligning with consumer-expressed preferences. At the same time, when dealing with negative sentiments, producers and marketers should pay attention not only to promoting the quality, origin, or composition of the wine, but also to how it is perceived in value terms.

As pricing goes, strategies may be fine-tuned by analysing consumer feedback for each segment, allowing a more effective positioning in the market. However, as previously stated, depending on the amount of information available in reviews and the price of the product, producers can adjust their advertising strategy differently [27].

For the cases that then apply, for example, mid-tier wines with a predominant positive sentiment for their value, can be promoted as "excellent value for money", tapping into consumers who trust simple heuristics such as price, since they are less involved with the product [76].

In addition, these insights can be used to refine campaigns. As introduced and replicated by the results of this study, there is a shifting trend towards lighter and fresher wines [28], plus, evoking our Emotion Detection results, this is a joyful product after all. Portuguese wineries can profit from this by developing marketing campaigns that emphasise these characteristics, promoting not only quality and heritage, but aligning with established trends, painting a colourful picture of the experience that is drinking their wine. As these reviews also come from online outlets, these sorts of campaigns may thrive especially where User-Generated Content (UGC) takes its part in influencing consumer decisions.

Concerning the results in our classification problem, it was shown that our models had moderate capacity for predicting consumer ratings based on textual reviews. This type of method can be a way for wine producers and marketers to predict how consumers will likely rate their products, crossing information with other characteristics of wines already available in stores and the ones to be released. This way, proactive adjustments in both product development and marketing strategies can be made in order of a successful launch or readjusting of established strategies. However, as this study suggests, consumers may not be as consistent as one could wish, so more experimentation should be made, in order for this to be possible.

In that sense, if an effective predicting model is achieved and the corpora of a specific wine segment is constantly showing higher predicted ratings, winemakers can focus on the promotion of those inherent characteristics, most valued by consumers and reviewers. However, if certain terms are commonly associated with lower predicted ratings, this can serve as a signal, warning wine producers to adjust their product or its marketing positioning.

All in all, the dataset, extracted from Vivino, consisting of over 129,673 reviews, provided the foundation for these analyses. The findings, directly influenced by the data extracted, underscore the potential for wineries to leverage UGC as a powerful tool and text mining as a helpful way to ultimately improving competitiveness in the wine industry.

5.2. Implications for Wineinformatics

The present research contributes to the Wineinformatics field by expanding its focus from expert wine reviews to UGC, more precisely reviews from Vivino platform users.

To this day, Wineinformatics has relied on data extracted mostly from renowned wine magazines, such as Wine Spectator, which offer wine tasting notes from connoisseurs, commonly proven to be consistent with the ratings attributed, when putting down their percepted quality and characteristics of certain wines.

Questioning their relatability to the common buyer, this study sets the first stone applying text mining techniques, such as sentiment analysis, topic modeling, and classification problems, to a large dataset of online consumer reviews.

By leveraging UGC, this work broadens the scope of the Wineinformatics field, reflecting the preferences, sentiments and opinions of everyday consumers rather than only trained, highly involved experts. The study showed that consumer reviews can be a rich information source, complementing the insights retrieved from expert review analysis, offering a more comprehensive notion of market trends and consumer perception.

The application of these techniques to Vivino reviews of Portuguese wines introduces a new realm of possibility for Wineinformatics research. Approaches such as this one allow for large-scale analysis of consumer opinion, identifying key trends and emerging preferences across a broad online demographic.

In sum, our findings demonstrated the potential for integrating UGC into sentiment analysis, topic modelling, emotion detection and predictive models for the wine sector, offering a baseline of such applications. These can then be leveraged by wineries in the improvement of the value and pertinence of Wineinformatics to the wine industry, delivering actionable business insights and more profitability.

5.3. Limitations

Although this study demonstrates the value and encourages the use of UGC in Wineinformatics, several limitations should be acknowledged.

Firstly, the data collection was restricted to a single platform, Vivino, which may limit the generalisability of the findings, since it may not represent the full consumer opinions spectrum. Reviews from other wine-related websites or platforms that feature UGC could potentially offer different perspectives and a more comprehensive analysis.

Additionally, all reviews analysed were in English and this may skew the results toward the English-speaking population, since wine is globally consumed and cultural differences may influence published reviews.

Furthermore, while text mining techniques such as the ones employed offer valuable insights, they still have their limitations, mostly when dealing with misspelling, abbreviations, slang, irony, and other unconventional writing forms, which could not be accurately processed.

Finally, this study focusses on reviews of Portuguese wines, hence should not be taken as a capture of the global wine consumer sentiments, opinions or preferences.

5.4. Future Work

Following the present research proposal, it is recommended as future work that similar studies be conducted on wine reviews from other platforms besides Vivino, such as Cellartracker[1]. Expanding the dataset to include reviews from multiple sources would allow

[1] www.cellartracker.com

for a more in depth analysis of consumer preferences and behaviors across different review environments.

Additionally, incorporating reviews written in other languages could provide insights into regional and cultural differences in wine consumption, further enriching the foundational data and putting to test multi-language model processing.

Further research agendas could also focus on more advanced and computational expensive Natural Language Processing (NLP) techniques. Transformers, such as Bidirectional Encoder Representations from Transformers (BERT) (already utilised once in the literature by Katumullage *et al.* [29] on expert reviews), could enhance accuracy, especially treating nuanced reviews.

Finally, the inclusion of more metadata into the equation, such as grape variety, price or region, can provide novel insights from the UGC point of view, building on the expert-based analyses already available. This approach would offer a more holistic understanding of consumer preferences through the lenses of each of those drivers.

References

[1] SAJN Nikolina, "The EU wine sector," European Parliamentary Research Service, Tech. Rep., Jul. 2023. [Online]. Available: https://policycommons.net/artifacts/4430699/the-eu-wine-sector/.

[2] Statista Market Insights, *Wine - Worldwide: Statista Market Forecast*. [Online]. Available: https://www.statista.com/outlook/cmo/alcoholic-drinks/wine/worldwide#revenue (visited on 06/02/2024).

[3] International Wine and Spirits Research, *Seven key trends that will shape the global wine industry in 2024*. [Online]. Available: https://www.theiwsr.com/seven-key-trends-that-will-shape-the-global-wine-industry-in-2024/ (visited on 06/02/2024).

[4] International Organisation of Vine and Wine, "State of the World Vine and Wine Sector. International Organisation of Vine and Wine," Tech. Rep., 2024. [Online]. Available: https://www.oiv.int/sites/default/files/2024-04/OIV_STATE_OF_THE_WORLD_VINE_AND_WINE_SECTOR_IN_2023.pdf (visited on 06/02/2024).

[5] J. Martínez-Navarro and R. Sellers-Rubio, "Three decades of research on wine marketing," *Heliyon*, vol. 10, no. 10, e30938, 2024, ISSN: 2405-8440. DOI: https://doi.org/10.1016/j.heliyon.2024.e30938. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S240584402406969X.

[6] G. Szolnoki, R. Dolan, S. Forbes, L. Thach, and S. Goodman, "Using social media for consumer interaction: An international comparison of winery adoption and activity," *Wine Economics and Policy*, vol. 7, no. 2, pp. 109–119, Dec. 2018, ISSN: 2212-9774. DOI: 10.1016/j.wep.2018.07.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212977418300486.

[7] D. Wilson and S. Quinton, "Let's talk about wine: Does Twitter have value?" *International Journal of Wine Business Research*, vol. 24, no. 4, pp. 271–286, Nov. 2012, ISSN: 1751-1062. DOI: 10.1108/17511061211280329. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/17511061211280329/full/html (visited on 06/09/2024).

[8] S. K. Dixit, *Routledge Handbook of Wine Tourism*, 1st ed. London: Routledge, Oct. 2022, ISBN: 978-1-00-314362-8. DOI: 10.4324/9781003143628. [Online]. Available: https://www.taylorfrancis.com/books/9781003143628 (visited on 06/09/2024).

[9] M. Sigala, E. Christou, and U. Gretzel, Eds., *Social media in travel, tourism and hospitality: theory, practice and cases* (New directions in tourism analysis). London

New York: Routledge, Taylor & Francis Group, 2016, ISBN: 978-1-138-24759-8 978-1-4094-2091-0.

[10] O. Müller, "Structuring Unstructured Data—Or: How Machine Learning Can Make You a Wine Sommelier," in *The Art of Structuring*, K. Bergener, M. Räckers, and A. Stein, Eds., Cham: Springer International Publishing, 2019, pp. 307–316, ISBN: 978-3-030-06233-0 978-3-030-06234-7. DOI: 10.1007/978-3-030-06234-7_29. [Online]. Available: http://link.springer.com/10.1007/978-3-030-06234-7_29 (visited on 06/09/2024).

[11] W. Fu, E.-K. Choi, and H.-S. Kim, "Text Mining with Network Analysis of Online Reviews and Consumers' Satisfaction: A Case Study in Busan Wine Bars," *Information*, vol. 13, no. 3, p. 127, Mar. 2022, ISSN: 2078-2489. DOI: 10.3390/info13030127. [Online]. Available: https://www.mdpi.com/2078-2489/13/3/127 (visited on 06/09/2024).

[12] E. Barbierato, I. Bernetti, and I. Capecchi, "Analyzing TripAdvisor reviews of wine tours: An approach based on text mining and sentiment analysis," *International Journal of Wine Business Research*, vol. 34, no. 2, pp. 212–236, May 2022, ISSN: 1751-1062, 1751-1062. DOI: 10.1108/IJWBR-04-2021-0025. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/IJWBR-04-2021-0025/full/html (visited on 06/09/2024).

[13] M. Santos, X. Antón Rodríguez, and A. Marta-Costa, "Productive efficiency of wine grape producers in the North of Portugal," *Wine Economics and Policy*, Sep. 2021, ISSN: 2212-9774, 2213-3968. DOI: 10.36253/wep-8977. [Online]. Available: https://oaj.fupress.net/index.php/wep/article/view/8977 (visited on 06/09/2024).

[14] B. Chen, *Wineinformatics: A New Data Science Application* (SpringerBriefs in Computer Science). Singapore: Springer Nature Singapore, 2023, ISBN: 978-981-19736-8-0 978-981-19736-9-7. DOI: 10.1007/978-981-19-7369-7. [Online]. Available: https://link.springer.com/10.1007/978-981-19-7369-7 (visited on 06/09/2024).

[15] W. Kwabla, F. Coulibaly, Y. Zhenis, and B. Chen, "Wineinformatics: Can Wine Reviews in Bordeaux Reveal Wine Aging Capability?" *Fermentation*, vol. 7, no. 4, p. 236, Oct. 2021, ISSN: 2311-5637. DOI: 10.3390/fermentation7040236. [Online]. Available: https://www.mdpi.com/2311-5637/7/4/236 (visited on 12/17/2023).

[16] Z. Dong, X. Guo, S. Rajana, and B. Chen, "Understanding 21st Century Bordeaux Wines from Wine Reviews Using Naïve Bayes Classifier," *Beverages*, vol. 6, no. 1, p. 5, Jan. 2020, ISSN: 2306-5710. DOI: 10.3390/beverages6010005. [Online]. Available: https://www.mdpi.com/2306-5710/6/1/5 (visited on 12/17/2023).

[17] J. Palmer and B. Chen, "Wineinformatics: Regression on the Grade and Price of Wines through Their Sensory Attributes," *Fermentation*, vol. 4, no. 4, p. 84, Sep.

2018, ISSN: 2311-5637. DOI: 10.3390/fermentation4040084. [Online]. Available: http://www.mdpi.com/2311-5637/4/4/84 (visited on 12/17/2023).

[18] B. Chen, V. Velchev, B. Nicholson, J. Garrison, M. Iwamura, and R. Battisto, "Wineinformatics: Uncork Napa's Cabernet Sauvignon by Association Rule Based Classification," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA: IEEE, Dec. 2015, pp. 565–569, ISBN: 978-1-5090-0287-0. DOI: 10.1109/ICMLA.2015.44. [Online]. Available: http://ieeexplore.ieee.org/document/7424376/ (visited on 12/17/2023).

[19] G. I. Penagos-Londoño, F. Ruiz-Moreno, R. Sellers-Rubio, S. Del Barrio-García, and A. B. Casado-Díaz, "Consistency of expert product reviews: An application to wine guides," *Wine Economics and Policy*, vol. 11, no. 2, pp. 51–60, Nov. 2022, ISSN: 2212-9774, 2213-3968. DOI: 10.36253/wep-12400. [Online]. Available: https://oaj.fupress.net/index.php/wep/article/view/12400 (visited on 12/17/2023).

[20] B. Chen, V. Velchev, J. Palmer, and T. Atkison, "Wineinformatics: A Quantitative Analysis of Wine Reviewers," *Fermentation*, vol. 4, no. 4, p. 82, Sep. 2018, ISSN: 2311-5637. DOI: 10.3390/fermentation4040082. [Online]. Available: http://www.mdpi.com/2311-5637/4/4/82 (visited on 12/17/2023).

[21] B. Chen, H. Le, T. Atkison, and D. Che, "A Wineinformatics Study for White-box Classification Algorithms to Understand and Evaluate Wine Judges," *Trans. Mach. Learn. Data Min.*, vol. 10, no. 1, pp. 3–24, 2017. [Online]. Available: http://www.ibai-publishing.org/journal/issue%5C_mldm/2017%5C_july/mldm%5C_10%5C_1%5C_3%5C_24.php.

[22] B. Chen, H. Le, C. Rhodes, and D. Che, "Understanding the Wine Judges and Evaluating the Consistency Through White-Box Classification Algorithms," in *Advances in Data Mining. Applications and Theoretical Aspects*, P. Perner, Ed., vol. 9728, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2016, pp. 239–252, ISBN: 978-3-319-41560-4 978-3-319-41561-1. DOI: 10.1007/978-3-319-41561-1_18. [Online]. Available: http://link.springer.com/10.1007/978-3-319-41561-1_18 (visited on 12/17/2023).

[23] E. Lefever, I. Hendrickx, I. Croijmans, A. van den Bosch, and A. Majid, "Discovering the Language of Wine Reviews: A Text Mining Account," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan: European Language Resources Association (ELRA), May 2018. [Online]. Available: https://aclanthology.org/L18-1521.

[24] R. Friberg and E. Grönqvist, "Do Expert Reviews Affect the Demand for Wine?" *American Economic Journal: Applied Economics*, vol. 4, no. 1, pp. 193–211, Jan. 2012, ISSN: 1945-7782, 1945-7790. DOI: 10.1257/app.4.1.193. [Online]. Available: https://pubs.aeaweb.org/doi/10.1257/app.4.1.193 (visited on 10/30/2023).

[25] Z. Lei, D. Yin, S. Mitra, and H. Zhang, "Swayed by the reviews: Disentangling the effects of average ratings and individual reviews in online word-of-mouth," *Production and Operations Management*, vol. 31, no. 6, pp. 2393–2411, Jun. 2022, ISSN: 1059-1478, 1937-5956. DOI: 10.1111/poms.13695. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/poms.13695 (visited on 10/30/2023).

[26] X. Liu, D. Lee, and K. Srinivasan, "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *Journal of Marketing Research*, vol. 56, no. 6, pp. 918–943, Dec. 2019, ISSN: 0022-2437, 1547-7193. DOI: 10.1177/0022243719866690. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0022243719866690 (visited on 10/30/2023).

[27] Y. Chen and J. Xie, "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science*, vol. 54, no. 3, pp. 477–491, Mar. 2008, ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.1070.0810. [Online]. Available: https://pubsonline.informs.org/doi/10.1287/mnsc.1070.0810 (visited on 10/30/2023).

[28] R. Del Rey and S. Loose, "State of the International Wine Markets in 2022: New market trends for wines require new strategies," *Wine Economics and Policy*, Jun. 2023, ISSN: 2212-9774, 2213-3968. DOI: 10.36253/wep-14758. [Online]. Available: https://oaj.fupress.net/index.php/wep/article/view/14758 (visited on 10/30/2023).

[29] D. Katumullage, C. Yang, J. Barth, and J. Cao, "Using Neural Network Models for Wine Review Classification," *Journal of Wine Economics*, vol. 17, no. 1, pp. 27–41, Feb. 2022, ISSN: 1931-4361, 1931-437X. DOI: 10.1017/jwe.2022.2. [Online]. Available: https://www.cambridge.org/core/product/identifier/S1931436122000025/type/journal_article (visited on 10/30/2023).

[30] R. Fragoso and A. A. Vieira, "Efficiency analysis of the Portuguese wine industry using accounting and operational metrics," *Results in Engineering*, vol. 14, p. 100 389, 2022, ISSN: 2590-1230. DOI: https://doi.org/10.1016/j.rineng.2022.100389. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590123022000597.

[31] C. C. Aggarwal, *Machine Learning for Text*. Cham: Springer International Publishing, 2018, ISBN: 978-3-319-73530-6 978-3-319-73531-3. DOI: 10.1007/978-3-319-73531-3. [Online]. Available: http://link.springer.com/10.1007/978-3-319-73531-3 (visited on 11/01/2023).

[32] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Mining and Knowledge Discovery*, vol. 24, no. 3, pp. 478–514, May 2012, ISSN: 1384-5810, 1573-756X. DOI: 10.1007/s10618-011-0238-6. [Online]. Available: http://link.springer.com/10.1007/s10618-011-0238-6 (visited on 11/01/2023).

[33] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7,

pp. 5731–5780, Oct. 2022, ISSN: 0269-2821, 1573-7462. DOI: 10.1007/s10462-022-10144-1. [Online]. Available: https://link.springer.com/10.1007/s10462-022-10144-1 (visited on 01/04/2024).
[34] D. M. Blei, "Probabilistic topic models," en, *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012, ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2133806.2133826. [Online]. Available: https://dl.acm.org/doi/10.1145/2133806.2133826 (visited on 11/29/2024).
[35] A. Shmilovici, "Support Vector Machines," en, in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds., New York: Springer-Verlag, 2005, pp. 257–276, ISBN: 978-0-387-24435-8. DOI: 10.1007/0-387-25465-X_12. [Online]. Available: http://link.springer.com/10.1007/0-387-25465-X_12 (visited on 08/29/2024).
[36] Y. Kim, "Convolutional Neural Networks for Sentence Classification," en, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1746–1751. DOI: 10.3115/v1/D14-1181. [Online]. Available: http://aclweb.org/anthology/D14-1181 (visited on 08/29/2024).
[37] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, ISSN: 1053587X. DOI: 10.1109/78.650093. [Online]. Available: http://ieeexplore.ieee.org/document/650093/ (visited on 08/29/2024).
[38] B. Chen, C. Rhodes, A. Crawford, and L. Hambuchen, "Wineinformatics: Applying Data Mining on Wine Sensory Reviews Processed by the Computational Wine Wheel," in *2014 IEEE International Conference on Data Mining Workshop*, Shenzhen, China: IEEE, Dec. 2014, pp. 142–149, ISBN: 978-1-4799-4274-9. DOI: 10.1109/ICDMW.2014.149. [Online]. Available: https://ieeexplore.ieee.org/document/7022591/ (visited on 11/05/2023).
[39] L. Le, P. N. Hurtado, I. Lawrence, Q. Tian, and B. Chen, "Applying Neural Networks in Wineinformatics with the New Computational Wine Wheel," *Fermentation*, vol. 9, no. 7, p. 629, Jul. 2023, ISSN: 2311-5637. DOI: 10.3390/fermentation9070629. [Online]. Available: https://www.mdpi.com/2311-5637/9/7/629 (visited on 11/09/2023).
[40] Q. Tian, B. Whiting, and B. Chen, "Wineinformatics: Comparing and Combining SVM Models Built by Wine Reviews from Robert Parker and Wine Spectator for 95 + Point Wine Prediction," *Fermentation*, vol. 8, no. 4, p. 164, Apr. 2022, ISSN: 2311-5637. DOI: 10.3390/fermentation8040164. [Online]. Available: https://www.mdpi.com/2311-5637/8/4/164 (visited on 12/26/2023).
[41] S. Zhang, "Sentiment analysis based on food e-commerce reviews," *IOP Conference Series: Earth and Environmental Science*, vol. 792, no. 1, p. 012 023, Jun. 2021, ISSN: 1755-1307, 1755-1315. DOI: 10.1088/1755-1315/792/1/012023. [Online].

Available: https://iopscience.iop.org/article/10.1088/1755-1315/792/1/012023 (visited on 12/26/2023).

[42] Z. Dong, T. Atkison, and B. Chen, “Wineinformatics: Using the Full Power of the Computational Wine Wheel to Understand 21st Century Bordeaux Wines from the Reviews,” *Beverages*, vol. 7, no. 1, p. 3, Jan. 2021, ISSN: 2306-5710. DOI: 10.3390/beverages7010003. [Online]. Available: https://www.mdpi.com/2306-5710/7/1/3 (visited on 12/17/2023).

[43] E. Treen, S. Lord Ferguson, C. Pitt, and J. Vella, “Exploring emotions on wine websites: Finding joy,” *Journal of Wine Research*, vol. 29, no. 1, pp. 64–70, Jan. 2018, ISSN: 0957-1264, 1469-9672. DOI: 10.1080/09571264.2018.1433139. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09571264.2018.1433139 (visited on 08/17/2024).

[44] S. Moon and W. A. Kamakura, “A picture is worth a thousand words: Translating product reviews into a product positioning map,” *International Journal of Research in Marketing*, vol. 34, no. 1, pp. 265–285, Mar. 2017, ISSN: 01678116. DOI: 10.1016/j.ijresmar.2016.05.007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S016781161630074X (visited on 12/18/2023).

[45] N. Wariishi, B. Flanagan, T. Suzuki, and S. Hirokawa, “Sentiment Analysis of Wine Aroma,” in *2015 IIAI 4th International Congress on Advanced Applied Informatics*, Okayama, Japan: IEEE, Jul. 2015, pp. 207–212, ISBN: 978-1-4799-9957-6 978-1-4799-9958-3. DOI: 10.1109/IIAI-AAI.2015.253. [Online]. Available: http://ieeexplore.ieee.org/document/7373902/ (visited on 12/18/2023).

[46] B. Chen, C. Rhodes, A. Yu, and V. Velchev, “The Computational Wine Wheel 2.0 and the TriMax Triclustering in Wineinformatics,” in *Advances in Data Mining. Applications and Theoretical Aspects*, P. Perner, Ed., vol. 9728, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2016, pp. 223–238, ISBN: 978-3-319-41560-4 978-3-319-41561-1. DOI: 10.1007/978-3-319-41561-1_17. [Online]. Available: http://link.springer.com/10.1007/978-3-319-41561-1_17 (visited on 12/26/2023).

[47] I. Hendrickx, E. Lefever, I. Croijmans, A. Majid, and A. Van Den Bosch, “Very quaffable and great fun: Applying NLP to wine reviews,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, Germany: Association for Computational Linguistics, 2016, pp. 306–312. DOI: 10.18653/v1/P16-2050. [Online]. Available: http://aclweb.org/anthology/P16-2050 (visited on 12/26/2023).

[48] B. Flanagan, N. Wariishi, T. Suzuki, and S. Hirokawa, “Predicting and Visualizing Wine Characteristics Through Analysis of Tasting Notes from Viewpoints,” in *HCI International 2015 - Posters’ Extended Abstracts*, C. Stephanidis, Ed., vol. 528, Series Title: Communications in Computer and Information Science, Cham: Springer

International Publishing, 2015, pp. 613–619, ISBN: 978-3-319-21379-8 978-3-319-21380-4. DOI: 10.1007/978-3-319-21380-4_104. [Online]. Available: http://link.springer.com/10.1007/978-3-319-21380-4_104 (visited on 12/26/2024).

[49] Z. Jin, W. Zhangwen, and N. Naichen, "Helping consumers to overcome information overload with a diversified online review subset," *Frontiers of Business Research in China*, vol. 13, no. 1, p. 15, Dec. 2019, ISSN: 1673-7326, 1673-7431. DOI: 10.1186/s11782-019-0062-1. [Online]. Available: https://fbr.springeropen.com/articles/10.1186/s11782-019-0062-1 (visited on 12/26/2023).

[50] J. K. Rout, K.-K. R. Choo, A. K. Dash, S. Bakshi, S. K. Jena, and K. L. Williams, "A model for sentiment and emotion analysis of unstructured social media text," *Electronic Commerce Research*, vol. 18, no. 1, pp. 181–199, Mar. 2018, ISSN: 1389-5753, 1572-9362. DOI: 10.1007/s10660-017-9257-8. [Online]. Available: http://link.springer.com/10.1007/s10660-017-9257-8 (visited on 07/06/2024).

[51] D. Khyani and S. B S, "An Interpretation of Lemmatization and Stemming in Natural Language Processing," *Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology*, vol. 22, pp. 350–357, Jan. 2021.

[52] M. S. Akhtar, D. Gupta, A. Ekbal, and P. Bhattacharyya, "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis," *Knowledge-Based Systems*, vol. 125, pp. 116–135, Jun. 2017, ISSN: 09507051. DOI: 10.1016/j.knosys.2017.03.020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S095070511730148X (visited on 08/17/2024).

[53] V. Bonta, N. Kumaresh, and N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," *Asian Journal of Computer Science and Technology*, vol. 8, no. S2, pp. 1–6, Mar. 2019, ISSN: 2249-0701. DOI: 10.51983/ajcst-2019.8.S2.2037. [Online]. Available: https://ajcst.co/index.php/ajcst/article/view/2037 (visited on 08/17/2024).

[54] M. Cavique, R. Ribeiro, F. Batista, and A. Correia, "Examining Airbnb guest satisfaction tendencies: A text mining approach," *Current Issues in Tourism*, vol. 25, no. 22, pp. 3607–3622, Nov. 2022, ISSN: 1368-3500, 1747-7603. DOI: 10.1080/13683500.2022.2115877. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/13683500.2022.2115877 (visited on 08/09/2024).

[55] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, *Stanza: A Python Natural Language Processing Toolkit for Many Human Languages*, Version Number: 2, 2020. DOI: 10.48550/ARXIV.2003.07082. [Online]. Available: https://arxiv.org/abs/2003.07082 (visited on 07/07/2024).

[56] C. Hutto and E. 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, pp. 216–225, May 2014, ISSN: 2334-0770, 2162-3449. DOI: 10.1609/icwsm.v8i1.14550. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550 (visited on 07/07/2024).

[57] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation," in *The Journal of Machine Learning Research*, vol. 3, 2003, pp. 993–1022.

[58] R. Rehurek and P. Sojka, "Gensim–python framework for vector space modelling," *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, vol. 3, no. 2, 2011.

[59] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, Dec. 2021, ISSN: 1869-5450, 1869-5469. DOI: 10.1007/s13278-021-00776-6. [Online]. Available: https://link.springer.com/10.1007/s13278-021-00776-6 (visited on 08/18/2024).

[60] L. P. Hung and S. Alias, "Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 27, no. 1, pp. 84–95, Jan. 2023, ISSN: 1883-8014, 1343-0130. DOI: 10.20965/jaciii.2023.p0084. [Online]. Available: https://www.fujipress.jp/jaciii/jc/jacii002700010084 (visited on 08/18/2024).

[61] E. Boitel, A. Mohasseb, and E. Haig, "A Comparative Analysis of GPT-3 and BERT Models for Text-based Emotion Recognition: Performance, Efficiency, and Robustness," in *Advances in Computational Intelligence Systems*, N. Naik, P. Jenkins, P. Grace, L. Yang, and S. Prajapat, Eds., vol. 1453, Series Title: Advances in Intelligent Systems and Computing, Cham: Springer Nature Switzerland, 2024, pp. 567–579, ISBN: 978-3-031-47507-8 978-3-031-47508-5. DOI: 10.1007/978-3-031-47508-5_44. [Online]. Available: https://link.springer.com/10.1007/978-3-031-47508-5_44 (visited on 08/18/2024).

[62] D. Carneros-Prado, L. Villa, E. Johnson, C. C. Dobrescu, A. Barragán, and B. García-Martínez, "Comparative Study of Large Language Models as Emotion and Sentiment Analysis Systems: A Case-Specific Analysis of GPT vs. IBM Watson," in *Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023)*, J. Bravo and G. Urzáiz, Eds., vol. 842, Series Title: Lecture Notes in Networks and Systems, Cham: Springer Nature Switzerland, 2023, pp. 229–239, ISBN: 978-3-031-48641-8 978-3-031-48642-5. DOI: 10.1007/978-3-031-48642-5_22. [Online]. Available: https://link.springer.com/10.1007/978-3-031-48642-5_22 (visited on 08/17/2024).

[63] M. M. Amin, R. Mao, E. Cambria, and B. W. Schuller, "A Wide Evaluation of ChatGPT on Affective Computing Tasks," *IEEE Transactions on Affective Computing*, pp. 1–9, 2024, ISSN: 1949-3045, 2371-9850. DOI: 10.1109/TAFFC.2024.3419593. [Online]. Available: https://ieeexplore.ieee.org/document/10572294/ (visited on 08/18/2024).

[64] A. Y. Taha, S. Tiun, A. H. Abd Rahman, and A. Sabah, "MULTILABEL OVER-SAMPLING AND UNDER-SAMPLING WITH CLASS ALIGNMENT FOR IMBALANCED MULTILABEL TEXT CLASSIFICATION," *Journal of Information and Communication Technology*, vol. 20, 2021, ISSN: 1675-414X. DOI: 10.32890/jict2021.20.3.6. [Online]. Available: http://e-journal.uum.edu.my/index.php/jict/article/view/14386 (visited on 08/30/2024).

[65] G. Lemaitre, F. Nogueira, and C. K. Aridas, *Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*, Version Number: 1, 2016. DOI: 10.48550/ARXIV.1609.06570. [Online]. Available: https://arxiv.org/abs/1609.06570 (visited on 08/30/2024).

[66] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, Dec. 1989, ISSN: 0899-7667, 1530-888X. DOI: 10.1162/neco.1989.1.4.541. [Online]. Available: https://direct.mit.edu/neco/article/1/4/541-551/5515 (visited on 08/25/2024).

[67] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *International Journal of Computer Applications*, vol. 181, no. 1, pp. 25–29, Jul. 2018, ISSN: 09758887. DOI: 10.5120/ijca2018917395. [Online]. Available: http://www.ijcaonline.org/archives/volume181/number1/qaiser-2018-ijca-917395.pdf (visited on 08/30/2024).

[68] M. al Shakoush, V. Andrikopoulos, and D. Feitosa, "Sentiment Analysis for Cost Management Actions Related To Changes In IaC Artifacts," in *21st SC@RUG 2024 proceedings 2023-2024*, University of Groningen Press, 2024, p. 99. (visited on 07/07/2024).

[69] V. Schoenmueller, O. Netzer, and F. Stahl, "The Polarity of Online Reviews: Prevalence, Drivers and Implications," *Journal of Marketing Research*, vol. 57, no. 5, pp. 853–877, 2020, ISSN: 0022-2437, 1547-7193. DOI: 10.1177/0022243720941832.

[70] W. W. Moe, O. Netzer, and D. A. Schweidel, "Social Media Analytics," in *Handbook of Marketing Decision Models*, B. Wierenga and R. Van Der Lans, Eds., vol. 254, Series Title: International Series in Operations Research & Management Science, Cham: Springer International Publishing, 2017, pp. 483–504, ISBN: 978-3-319-56939-0 978-3-319-56941-3. DOI: 10.1007/978-3-319-56941-3_16. [Online]. Available: http://link.springer.com/10.1007/978-3-319-56941-3_16 (visited on 07/13/2024).

[71] N. Hu, P. A. Pavlou, Temple University, J. Zhang, and University of Texas, Arlington, "On Self-Selection Biases in Online Product Reviews," *MIS Quarterly*, vol. 41, no. 2, pp. 449–471, Feb. 2017, ISSN: 02767783, 21629730. DOI: 10.25300/MISQ/2017/41.2.06. [Online]. Available: http://misq.org/on-self-selection-biases-in-online-product-reviews.html?SID=ke47j8g6v47ngppd0eisqn12h7 (visited on 07/13/2024).

[72] C. Sievert and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, Maryland, USA: Association for Computational Linguistics, 2014, pp. 63–70. DOI: 10.3115/v1/W14-3110. [Online]. Available: http://aclweb.org/anthology/W14-3110 (visited on 08/09/2024).
[73] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, May 1992. DOI: 10.1080/02699939208411068. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/02699939208411068 (visited on 08/17/2024).
[74] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, and W. Zhang, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Information Fusion*, vol. 83-84, pp. 19–52, Jul. 2022, ISSN: 15662535. DOI: 10.1016/j.inffus.2022.03.009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1566253522000367 (visited on 08/17/2024).
[75] A. Tonacci, G. Scalzini, P. Díaz-Guerrero, C. Sanmartin, I. Taglieri, G. Ferroni, G. Flamini, L. Odello, L. Billeci, and F. Venturi, "Chemosensory analysis of emotional wines: Merging of explicit and implicit methods to measure emotions aroused by red wines," *Food Research International*, vol. 190, p. 114 611, Aug. 2024, ISSN: 09639969. DOI: 10.1016/j.foodres.2024.114611. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0963996924006811 (visited on 08/17/2024).
[76] D. De Toni, R. Pompermayer, F. Lazzari, and G. S. Milan, "The symbolic value of wine, moderating and mediating factors and their relationship to consumer purchase intention," *International Journal of Wine Business Research*, vol. 34, no. 2, pp. 190–211, May 2022, ISSN: 1751-1062, 1751-1062. DOI: 10.1108/IJWBR-01-2021-0006. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/IJWBR-01-2021-0006/full/html (visited on 09/09/2024).