WINEINFORMATICS: COMPARISON AND COMBINATION OF CLASSIFICATION MODELS BUILD WITH WINE REVIEWS FROM DIFFERENT SOURCES FOR CLASS PREDICTION

By

Qiuyun Tian

A thesis presented to the Department of Computer Science and Engineering
and the Graduate School of the University of Central Arkansas
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science

Conway, Arkansas
August 2022

ABSTRACT

Wineinformatics, as one of the new fields of Data Science, uses wine as application domain. Wine reviews in human language format are processed by a Natural Language Processing tool called the Computational Wine Wheel (CWW). In previous research, the source for wine reviews was Wine Spectator. For this project, two wine sources are collected and used for this research: wine reviews from Robert Parker and wine reviews from Wine Spectator. This project focused on three main goals: 1. Enhancing the Computational Wine Wheel by analyzing and processing Robert Parker's review; 2. Evaluating and comparing the consistency of Wine Spectator and Robert Parker's reviews using the new Computational Wine Wheel 3.0; 3. Proposing new methods to combine different sources of reviews using the new Computational Wine Wheel 3.0 for achieving better results. Five datasets were used to predict whether the wine belonged to Classic Wines (95+ scores) or not (94- scores) by implementing two classification methods: the black-box model Support Vector Machines and the white-box model Naive Bayes. After enhancing the Computational Wine Wheel, 75% of datasets had better performance after applying the new CWW 3.0 than applying the CWW 2.0. The proposed combination method "RPcomWS" that combines two reviews from Robert Parker and Wine Spectator into one review achieved accuracy as high as 84.89%, with an F-score of 75.49%, a precision of 84.64%, and a recall of 68.13%. The result shows that the review combination of Robert Parker and Wine Spectator can extract the most wine attributes from reviews to achieve better performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Data Science

Acquiring information and transforming it into a store of knowledge is one of the essential skills for human development [1]. From ancient times, when the oldest and most experienced man was the leader, to today, when students sit in classrooms and learn from books, the principle remains the same: that man thrives by turning information into knowledge patterns and putting it to use. However, with the development of science and technology, information itself and its generation and processing have developed in a different way. The computer was one of the greatest inventions of the last century. First, an infinite amount of information, which is called data, was produced, and stored through computers. Secondly, the processes of large amounts of data that cannot easily be done by humans, then are handled by computers.

As computers become more and more powerful, immeasurable amounts of data are produced in various ways, such as social data [2] from all kinds of social media platforms, machine data [3] from industrial equipment that are installed in machinery, transactional data [4] from daily transactions, and so on. The large amount of data brings a large amount of information as well as challenges. Most of the data is unstructured. In order to process and analyze and make good use of this data, data science has been extended. Data science is a new science field that has been rapidly developed in the past decades, which is dealing with data in a scientific way. As an interdisciplinary field, data science relates to a very broad series of fields, such as data mining, statistics, artificial intelligence, business strategy and much more.

In this thesis, the study of data mining techniques was explored. Data mining is a process of extracting useful patterns and discovering knowledge from large amounts of data sets, and it contains mining techniques such as classification, clustering, regression, association rules and so on. The learning algorithms are normally categorized into four types: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning [5] uses a training set to teach models to predict the desired output, with the typical algorithms of classification and regression. Unsupervised learning [6] uses algorithms to analyze and cluster unlabeled datasets, such as algorithms of clustering and association rules. Semi-supervised learning [7] is between supervised and unsupervised learning that deals with both labeled datasets and unlabeled datasets. Reinforcement learning [8] is using an agent to decide the performance and learn from its experience.

With all the powerful techniques, as shown in Figure 1, data science has been used across different industries and achieved important results. For instance, in hospitals, data science predictive modeling can help the outcomes of disease through the historical data of patients [9]. The fraud detection of data science has been used in banks for helping manage resources and make better decisions [10]. Business markets use data clustering to find interesting patterns for better sales [11]. More common applications of data science, such as auto-driving cars [12], advertisement injecting, and so on, are affecting our daily lives. The diversity of data sources brings a variety of domains to the research of data science. In this paper, the domain that will be focused on is wine.

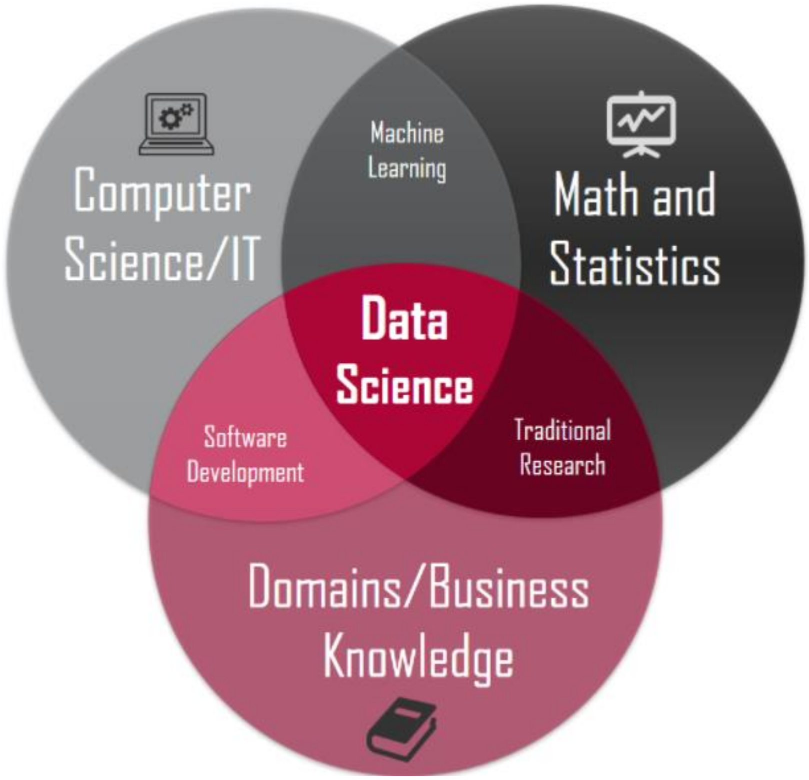

Figure 1. Data Science Venn Diagram

1.2 Wineinformatics

Wine as a popular beverage has a history of more than thousands of years. Back to the time of the fourth millennium B.C. when viticulture originated in the region south of Black and Caspian seas, Europeans kept controlling the occupation until the end of the nineteenth century. Wine production in the New World started to develop rapidly in the late nineteenth century because of the development of the global economy with the massive immigration and the improved transportation and so on [13]. According to the global wine production statistics maintained by the International Organization of Vine and Wine (OIV), more than 260 million hectoliters of wine were produced in 2020 worldwide [14]. The high yield of wine products is due to the large population and land usage, high consumer demand and the improvement of wine production technology. There are four basic steps involved in the winemaking process: picking the grapes, crushing the grapes, fermenting the juice, and aging the wine. Subtle differences in each step affect the taste, smell and all the other aspects of wine. For example, experienced wine makers know the importance of the timing of crushing the grapes by the rapidity with which the grapes are gathered and pressed, so that the whole contents of each vat may be exactly in the same state, and a simultaneous and equal fermentation be secured throughout [15]. Wine, as a culture with a long history, also is a popular economic commodity that has many wine lovers. Fine wine products are not only an enjoyment for tasting, but also a great investment choice due to the aging requirements of wine.

Wineinformatics is a new data science method that uses wine as domain knowledge that incorporates data science techniques and wine-related datasets, which includes physicochemical laboratory data and wine reviews. The physicochemical laboratory data comes from the physicochemical tests that are laboratory-based and focusing on factors like PH level, acidity, the presence of sugar and other chemical properties. These datasets could make interesting discoveries on relationships between the chemical components of wine and quality of human tastings to improve the wine making process. However, the cost of the laboratory test is fairly high, and the physicochemical attributes in this dataset are from the laboratory analysis, which does not directly reflect the sensory feelings that wine brings to human taste. The wine review datasets come from reviews that contain the human sensory perception of wine tasting by wine experts. In the comparison of the physicochemical laboratory datasets and wine review datasets, both datasets provide information to wine makers and markets, while wine review datasets also benefit wine consumers by providing easy understanding and professional information.

1.3 Wine Reviews

Wine reviews, as an essential part of the industry of wine, set trends and guide customers' preferences. Wine reviews describe the taste of wine and sometimes also wine-related information, such as vineyard, soil, etc. There are a lot of forms of wine reviews, such as wine magazines, personal blogs, or social media accounts. Famous wine reviews could effectively change the rank of a wine, as well as the price in the market. Some top publications for wine reviews include Wine Spectator, Wine Advocate by Robert Parker, and Wine Enthusiast Magazine and so on. Figures 2 and 3 are wine review examples from Wine Spectator and Robert Parker. Wine Spectator has a significant influence on the culture of wine with its annual reviews of over 15,000 wines from around the world [16]. Based on the magazine's policy, the wine experts are required to do blind tastings to avoid bias. Therefore, Wine Spectator's wine reviews provide a trustworthy and effective reputation and can be used for science projects. Robert Parker is a world-renowned wine critic. His wine reviews cause significant influence to the reputation and price of the wine.

Chateau Latour   Year : 2003   Price : $1999.99   Score : 98

Intense aromas of blackberry, licorice, currant and mineral. Full-bodied, with very well-integrated tannins and a long, long finish. Very refined and beautiful. Goes on for minutes. This reminds me of the fabulous 1996. But even better. Best after 2012.

Figure 2. The example of Wine Spectator wine reviews on Wine.com

Figure 2 is an example of a review of Wine Spectator of 2003 Chateau Latour. The review describes the aromas of the wine as "intense," "blackberry," "licorice," "currant," and "mineral." It is focusing on the taste of wine with attributes such as "Full-bodied," "well-integrated tannins," "long finish," and so on.

Chateau Latour Year : 2003 Price : $1999.99 Score : 98

2003 was one of the hottest, earliest Bordeaux vintages ever. Some vines suffered from lack of moisture, but old vines and clay subsoil at Enclos saw this vineyard through. The Merlot harvest occurred between September 8 and 13, and the Cabernet Sauvignon was picked between September 22 and 30. The 2003 Latour is a blend of 81% Cabernet Sauvignon, 18% Merlot and 1% Petit Verdot. Six percent of the press wine was added to the final blend. It has a medium to deep garnet-purple color, then wow—it explodes from the glass with bombastic black and blue fruits, followed up by meat, wood smoke, sandalwood and Indian spice accents with underlying floral wafts. The palate is full, rich, velvety, seductive and very long on the finish.

Figure 3. The example of Robert Parker wine reviews on Wine.com

Figure 3 is the wine review from Robert Parker of 2003 Chateau Latour, the same wine as Figure 2. In this review, the taste of the wine was described as "black and blue fruits," "meat," "wood smoke," "very long on the finish." As we can see, the attributes of these two reviews are very different from each other.

Not only are the wine reviews various, each wine publication rates wine slightly differently, but most rating systems are based on Robert Parker's 100-point rating system. Robert Parker is a world-renowned wine critic, who has a high reputation and affection in the wine world. He grades wines based on the taste, smell, and all the other attributes on a scale of 50 to 100.

Exploring Figure 4, the comparison of the rating system between Robert Parker and Wine Spectator is not dramatically different. The top tier wine range for Wine Spectator is 95-100, while Robert Parker is 96-100. They also differ in the range of 80-89. While Wine Spectator separates this range into two ranges, Robert Parker treats 80-89 as one range.

Robert Parker

96-100:	An extraordinary wine
90-95:	An outstanding wine
80-89:	A barely above average to very good wine
70-79:	An average wine
60-69:	A below average wine
50-59:	An unacceptable wine

Wine Spectator

95-100:	Classic : a great wine
90-94:	Outstanding : a wine of superior character and style
85-89:	Very good : a wine with special qualities
80-84:	Good : a solid, well-made wine
75-79:	Mediocre : a drinkable wine that may have minor flaws
50-74:	Not recommended

Figure 4. Rating System of Comparison of Robert Parker and Wine Spectator

With the very descriptive attributes from wine reviews and the precise rating system, applying these wine review datasets on data science techniques such as classification, clustering, and association rules is a practical project. For instance, with a classification algorithm, the rating scores of wines could be the class of the dataset while the descriptive words from wine reviews could be the attributes of the dataset. Clustering algorithms may be used to find out interesting attribute patterns that exist in different wines with the same rating score. However, wine reviews are stored in a human-readable format. To put wine reviews into actual data analysis usage, in this project we use the Computational Wine Wheel as a natural language processing tool to extract the attributes from wine reviews so that they can be processed by computers. In our previous research, the datasets generated by the Computational Wine Wheel have been used on a variety of different topics. For example, evaluating the Wine Spectator and all of its major reviews through both white-box and black-box classification algorithms [17], testing if wine reviews can be used to predict whether a bottle of wine can be held six years or more before it hits the optimal conditions for drinking [18], and so on [19][20].

1.4 Goal of the Research

The new methodology used for extracting key attributes from wine reviews, which is called Computational Wine Wheel, was developed based on only the wine reviews of Wine Spectator. In this research, the first goal was to enhance the Computational Wine Wheel. For expanding the vocabulary of the CWW in this project, a new review source, reviews from Robert Parker, was used. After updating the CWW, the second goal was to evaluate and compare the consistency of Wine Spectator's reviews and Robert Parker's reviews using the new Computational Wine Wheel 3.0. The last goal was to propose new methods to combine Wine Spectator and Robert Parker’s reviews for achieving better performance and acquire more precise information.

CHAPTER 2 DATA

Data, the cornerstone of data science, is a collection of facts, including texts, numbers, audio, observations, and so on. The vast amount of data in our world provides an enormous source for data scientists to use. While most data are unstructured, the step of data preprocessing is very essential for the whole progress of data analysis. This chapter provides an overview of data preprocessing, focusing on problems of real-world data. Under the domain of Wineinformatics, this thesis focused on the data type of wine reviews that are produced by professional wine experts in human language format.

In previous works, the focus of the study has been on the reviews of Wine Spectator. In this work, we explored the utility of the computational wine wheel on new resources, such as reviews by Robert Parker, as well as compared the difference between reviews from Robert Parker and Wine Spectator.

2.1 Data Collection

2.1.1 Wine.com

In this project, the source data was from the website wine.com, which is an e-commerce website. Wine.com is the nation's leading online wine retailer that provides its customers access to the world's largest wine store, and it achieves the sales volume of over two million bottles of wine per year. For offering descriptive and variety guidance to the customers, wine.com provides wine reviews from different professional critics, as shown in Figure 5, including wine magazines like Wine Enthusiast, Decanter, Wine Spectator, and wine experts like James Suckling, Robert Parker, etc.

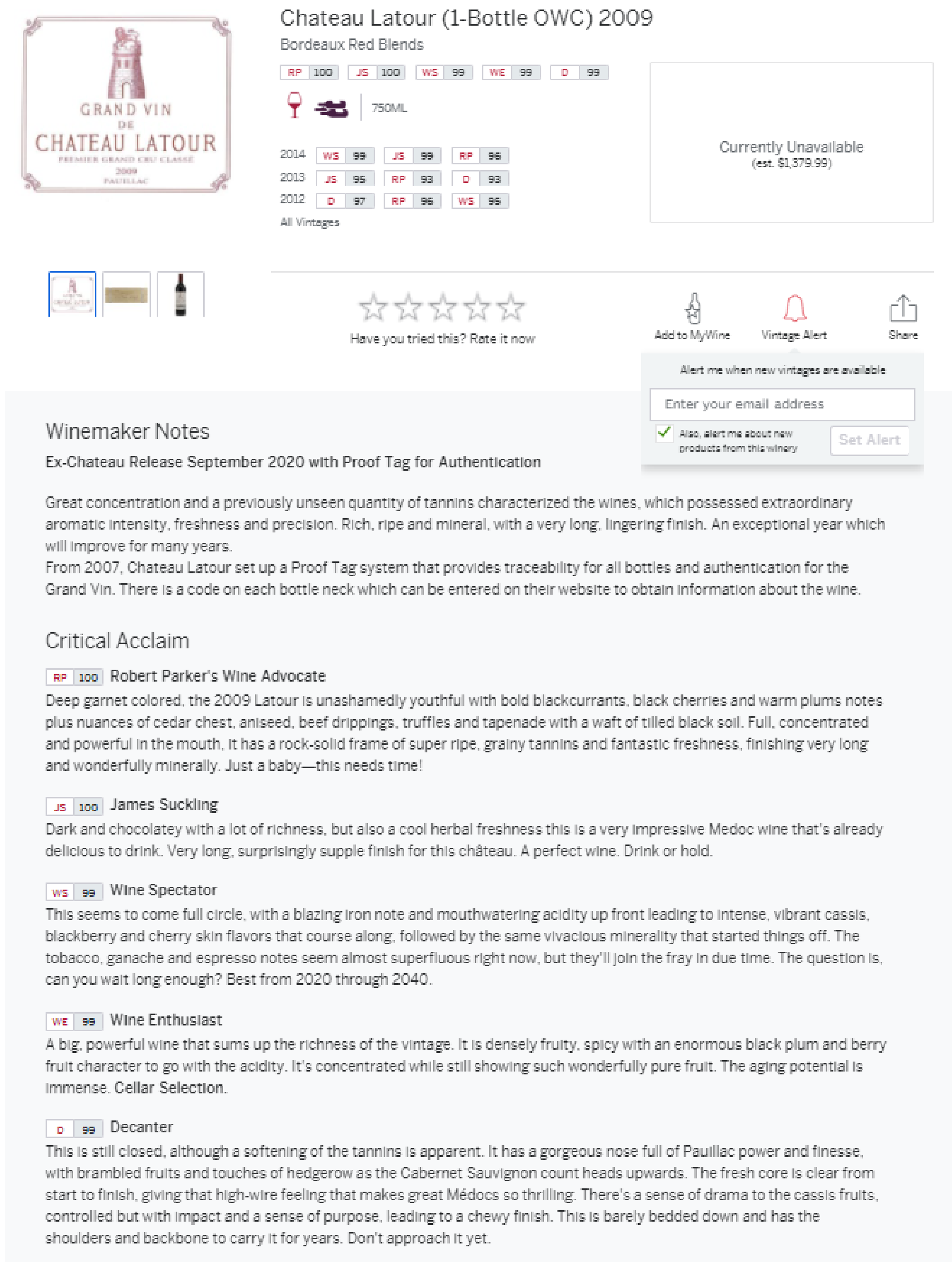

Figure 5. Wine.com Example

Wine.com provided a trustworthy and convenient way to collect wine reviews for our research. Especially for the purpose of comparing reviews from Robert Parker and Wine Spectator, collecting wines that contain both reviews were more straightforward on this website because it displayed all reviews on one page.

2.1.2 Dataset 1: 1855 Elite Bordeaux Wines

According to the global wine production statistics maintained by OIV, France occupied the second of the world wine production places with 46.6 million of hectoliters in 2020 [14]. French wine originated in the 6th century BC, and it became a part of their civilization, as well as winemaking. In the Exposition Universelle de Paris in 1855, Emperor Napoleon III requested a classification system for France's best Bordeaux wines that would be on display for visitors from around the world. Under the classification of 1855, wines were ranked in importance from first to fifth growths, which was directly related to the quality of wine. The list of Bordeaux Wine Official Classification of 1855 is still influential today. Therefore, the wine review collected in this project is based on this list.

In wine.com, we searched all wines in the 1855 Bordeaux Wine Official Classification list made in the 21st century (2000-2020) and included the wine into the dataset 1 if the wine had both Robert Parker and Wine Spectator reviews. Dataset 1 contained 513 wines with a total of 1026 wine reviews. The name, vintage, score and the wine reviews of each wine were collected.

2.1.3 Dataset 2: Bordeaux Wines

One of the most well-known French wine making regions in the world is Bordeaux, which is located on the west coast of central France. This location has the ideal climate and soil for high quality viticulture, so that it produces the famous Bordeaux wine that is considered in a typical old-world style. Annual weather differences create significant vintage variations, making Bordeaux an existing wine region to follow. In wine.com, all wines produced in Bordeaux were searched, and collected if the wine had both Robert

Parker and Wine Spectator reviews. These wine reviews from wines produced in Bordeaux formed the dataset 2. Dataset 2 included 2341 wines with a total of 4682 wine reviews, half from Robert Parker, another half from Wine Spectator. The name, vintage, score, and the wine reviews of each wine were collected.

2.1.4 Dataset 3: Italy Wines

Italian wine is the wine produced in every region of Italy, home of one of the oldest wine-producing regions in the world. According to the global wine production statistics maintained by the International Organization of Vine and Wine (OIV), Italy was the first largest wine producer in the world in 2021, as well as in the past-five-years [14]. Naturally, most Italian wine regions enjoy a Mediterranean climate and a notable coastline. Considering the variable terrain and conditions, most high-quality viticulture in Italy takes place on picturesque hillsides. In wine.com, all wines from Italy were searched, and included in dataset 3 if the wine had both Robert Parker and Wine Spectator reviews. Dataset 3 contained 3198 wines with a total of 6396 wine reviews. The name, vintage, score, and the wine reviews of each wine were collected.

2.1.5 Dataset 4: California Wines

The California wine industry is growing and changing amidst a global revolution in grape growing, wine production, wine marketing and consumer tastes [45]. Unlike the Bordeaux wine industry and Italy wine industry, California wine industry has younger age and newer wine culture. The per capita wine consumption and the quality of wine consumed consistently rise in the United State, and the California wine industry is one of the most important roles of wine production in the US. In wine.com, all wines produced in

California were searched and included in dataset 4 if the wine had both Robert Parker and Wine Spectator reviews. Dataset 4 contained 4180 wines with a total of 8360 wine reviews. The name, vintage, score, and the wine reviews of each wine were collected.

2.1.6 Dataset 5: All Datasets Combination

Dataset 5 was the last dataset, which combined all four datasets above into one big dataset. This dataset contained 10,232 wines with a total of 20,464 wine reviews. The name, vintage, score, and the wine reviews of each wine were collected. Dataset 5 was the biggest dataset that included all instances into one dataset; therefore, it provided a summarized review of all wine reviews.

2.1.7 Datasets Summary

An example of the raw data collection is given in Figure 6. The wine name and vintage were combined into one column, the score and review take two columns.

Name	score	Review
Chateau Mouton Rothschild 1959	100	I am always blown away by the 1959 Mouton, one of the greatest Moutons made in the last thi
Chateau Petrus 1971	95	This wine has been seemingly fully mature since the mid- to late seventies. It is a seductive, opu
Chateau Pichon Longueville Comtesse de Lalan	90	Undoubtedly a top success for the vintage, Pichon-Lalande's 1979 exhibits a dark garnet color w
Chateau Leoville Las Cases 1982	95	I have had perfect bottles of this cuvee, but, perplexingly, the bottles from my cellar tend to be
Chateau Haut-Brion (bin soiled, scuffed labels)	95	I know Jean Delmas, whom I respect as one of the world's greatest wine producers, has always
Chateau Latour 1982	100	1982 was a great vintage—relatively warm and prolific, producing wines of richness and depth.
Chateau La Lagune 1982	92	Unquestionably the greatest La Lagune until the 2005 was conceived, the 1982 exhibits a dense
Chateau Mouton Rothschild 1982	100	Medium brick in color, the nose of the 1982 Mouton Rothschild is a little closed and sluggish to
Chateau Ducru-Beaucaillou (stained label) 198	96	At a charity dinner in Charleston, SC, the 1982 Ducru Beaucaillou from my cellar was the only co
Chateau Cheval Blanc 1982	92	During its first 10-12 years of life, this was a perfect wine, but it now seems to be in a stage whe
Chateau Margaux (very top shoulder) 1982	97	The 1982 Chateau Margaux is a wine that I have tasted many times, constantly contrasted agair
Chateau Lynch-Bages 1982	93	Beautifully mature with sweet, sun-drenched black currant, fig, roasted herb and loamy soil cha
Chateau Cos d'Estournel 1982	95	This 1982 is still displaying a beautiful deep ruby/purple hue as well as a stunning set of aromat
Chateau Certan de May 1982	93	I have been drinking this wine out of half bottle, and this was the first regular bottle I have had

Figure 6. Example of Raw Data Collection

In sum, five datasets were generated from wine.com as shown in Table 1. Dataset 1 contained 513 wines from the list of 1855 Bordeaux Wine Official Classification. Dataset 2 contained 2341 wines that were produced in Bordeaux. Dataset 3 contained 3198 wines that were produced in Italy, and Dataset 4 contained 4180 wines that were produced in California. Dataset 5 contained 10232 wines, which was the total of all four datasets above.

Table 1. Dataset Summary

	Data Source	Wine Count	Wine Review Count
Dataset 1	1855 Elite Bordeaux	513	1026
Dataset 2	Bordeaux	2341	4682
Dataset 3	Italy	3198	6396
Dataset 4	California	4180	8360
Dataset 5	All Above	10232	20464

Most of the previous Wineinformatics research targeted the classification problem on predicting whether a wine can receive 90 points or above; thus, if the wine received a score equal or above 90 points out of 100, the label of the wine was marked as a positive (+) class. Otherwise, the label was marked as a negative (-) class. However, the wines collected in this research were the wines marketed on the high-class e-commerce website, 99% of them received higher than 90 points. Therefore, the targeted classification problem in this project was whether the wine received 95 points or above; thus, if the wine received a score equal or above 95 points out of 100, the label of the wine was marked as a positive (+) class. Otherwise, the label was marked as a negative (-) class. However, even though the class label threshold was 95, some datasets were still imbalanced datasets.

2.2 Data Preprocessing

For making the knowledge discovery more convenient and easier, removing the irrelevant and redundant information or noisy and unreliable data is very necessary. Data preprocessing is a fundamental stage to transform the unstructured data into cleaned forms which can be used for high-profit purposes. Data preprocessing includes data cleaning, normalization, transformation, attributes extraction and so on [22]. Exploring Figure 6, the raw data contained the name of the wine, the year, the scores, and the reviews from both Robert Parker and Wine Spectator. The focus of the data preprocessing was on the wine reviews.

Figure 7 is the example of the Wine Spectator review that was discussed in the last chapter. We wanted to look deeper into it from the perspective of data preprocessing. In this review in Figure 7, the highlighted words/word groups “blackberry,” “licorice,” “currant,” and “mineral” are the aromas description of this wine; “full-bodied,” “well-integrated tannins,” and “a long, long finish” are the body and structure description of this wine. “Very refined” and “beautiful” are an overall description of the wine. These are the important key words of this wine that we wanted to explore. However, wine reviews are stored in a human readable format that cannot be processed directly by computers. In this type of data, the contents and information are too complex and lengthy, including various numbers, punctuation marks, and different types of connective words that are not related to the wine itself. Therefore, instead of removing or deleting the irrelevant and redundant information, we extracted the useful information out of the reviews.

Chateau Latour	Year : 2003	Price : $1999.99	Score : 98

Intense aromas of blackberry, licorice, currant and mineral. Full-bodied, with very well-integrated tannins and a long, long finish. Very refined and beautiful. Goes on for minutes. This reminds me of the fabulous 1996. But even better. Best after 2012.

Figure 7. The example of Wine Spectator wine reviews on Wine.com

2.3 The Computational Wine Wheel

In order to extract the essential attributes from the wine reviews, a new natural language processing application, the Computational Wine Wheel, was used. The Computational Wine Wheel (CWW) works as a dictionary that contains all kinds of wine related attributes, which include fruit flavors (berry, apple, etc.), the body of wine (tannin, acidity, etc.), descriptive adjectives (balance, beautiful, etc.), and so on. By using CWW in the programming, we were able to extract the keywords from wine reviews and generate the datasets for the computer to process.

The Computational Wine Wheel has multiple levels and branches to separate the very broad categories of wine attributes into more specific and detailed subcategories. In the datasets of this chapter, the Computational Wine Wheel 2.0 was used. Table 2 is a partial demonstration of CWW2.0, which contains four parts of attributes. In the complete CWW2.0, there are 14 “CATEGORY_NAME” attributes, 34 “SUBCATEGORY_NAME" attributes, 1932 “SPECIFIC_NAME” attributes, and 986 “NORMALIZED_NAME” attributes. The “SPECIFIC_NAME” attributes were used to match with the key words in the wine reviews, then the corresponding “NORMALIZED_NAME” attributes, “SUBCATEGORY_NAME” attributes and “CATEGORY_NAME” attributes were encoded as vectors in the datasets.

Table 2. Partial Demonstration of the Computational Wine Wheel 2.0

	A	B	C	D	E
1	CATEGORY_NAME	SUBCATEGORY_NAME	SPECIFIC_NAME	NORMALIZED_NAME	CATEGORY
2	FRUITY	CITRUS	BLOOD ORANGE	BLOOD ORANGE	3
3	FRUITY	CITRUS	CITRUS	CITRUS	3
4	FRUITY	CITRUS	CITRUS GLAZED	CITRUS	3
5	FRUITY	CITRUS	CITRUS PEEL	CITRUS PEEL	3
6	FRUITY	CITRUS	CITRUS PEEL FLAVORS	CITRUS PEEL	3
7	FRUITY	CITRUS	CITRUS ZEST	CITRUS ZEST	3
8	FRUITY	CITRUS	CITRUS ZEST NOTES	CITRUS ZEST	3
9	FRUITY	CITRUS	CITRUS-LACED	CITRUS	3
10	FRUITY	CITRUS	CLEMENTINE	CLEMENTINE	3
11	FRUITY	CITRUS	FRESHLY SQUEEZED LIME	LIME	3
12	FRUITY	CITRUS	GRAPEFRUIT	GRAPEFRUIT	3
13	FRUITY	CITRUS	GRAPEFRUIT PEEL	GRAPEFRUIT PEEL	3
14	FRUITY	CITRUS	HINTS OF CITRUS	CITRUS	3
15	FRUITY	CITRUS	HINTS OF ORANGE	ORANGE	3
16	FRUITY	CITRUS	KEY LIME	KEY LIME	3
17	FRUITY	CITRUS	LEMON	LEMON	3
18	FRUITY	CITRUS	LEMON AROMAS	LEMON	3
19	FRUITY	CITRUS	LEMON MOUSSE	LEMON MOUSSE	3
20	FRUITY	CITRUS	LEMON PEEL	LEMON PEEL	3

For example, Table 3 presents two reviews of one wine: the left one is Robert Parker's review, and the right one is the Wine Spectator's review. The program started with finding attributes from the wine reviews that matched the list of 1,932 "SPECIFIC_NAME" attributes. In this process, each review was treated as a list of words, so that each word took turns to be compared with the attributes in the "SPECIFIC_NAME" list. The list of 1932 "SPECIFIC_NAME" attributes was separated into two lists which were the list of single words and the list of word groups, then sorted from longest to shortest length. The program looped through the list of word groups and deleted the matched attributes from the review first, so that it avoided the repeated single word matching situation. For example, if the review contained "blueberry cream", the program looped through the list of word groups first and deleted the word group "blueberry cream," therefore when the program looped through the list of single words it did not extract the word "blueberry" to add the redundant attributes. When the program found the attribute that matches the "SPECIFIC_NAME" attribute, as shown in Figure 8, the corresponding

“NORMALIZED_NAME” attribute was assigned 1, otherwise, it was assigned 0. The corresponding “SUBCATEGORY” and “CATEGORY” attributes were incremented continuously for one wine review as shown in Figure 9.

Table 3. Robert Parker and Wine Spectator’s review

Robert Parker’s review	Wine Spectator’s review
2003 was one of the hottest, earliest Bordeaux vintages ever. Some vines suffered from lack of moisture, but old vines and clay subsoil at Enclos saw this vineyard through. The Merlot harvest occurred between September 8 and 13, and the Cabernet Sauvignon was picked between September 22 and 30. The 2003 Latour is a blend of 81% Cabernet Sauvignon, 18% Merlot and 1% Petit Verdot. Six percent of the press wine was added to the final blend. It has a medium to deep garnet-purple color, then wow—it explodes from the glass with bombastic black and blue fruits, followed up by meat, wood smoke, sandalwood and Indian spice accents with underlying floral wafts. The palate is full, rich, velvety, seductive and very long on the finish.	Intense aromas of blackberry, licorice, currant and mineral. Full-bodied, with very well-integrated tannins and a long, long finish. Very refined and beautiful. Goes on for minutes. This reminds me of the fabulous 1996. But even better. Best after 2012.

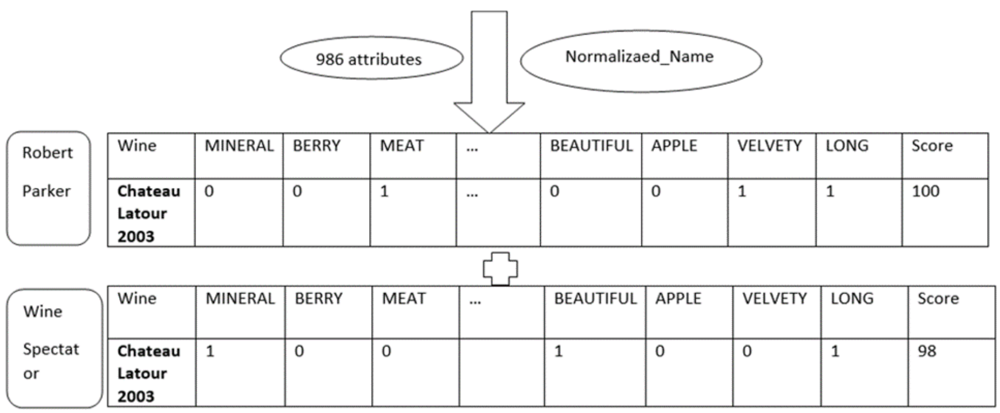

Figure 8. The First Step of Extracting Word with Computational Wine Wheel

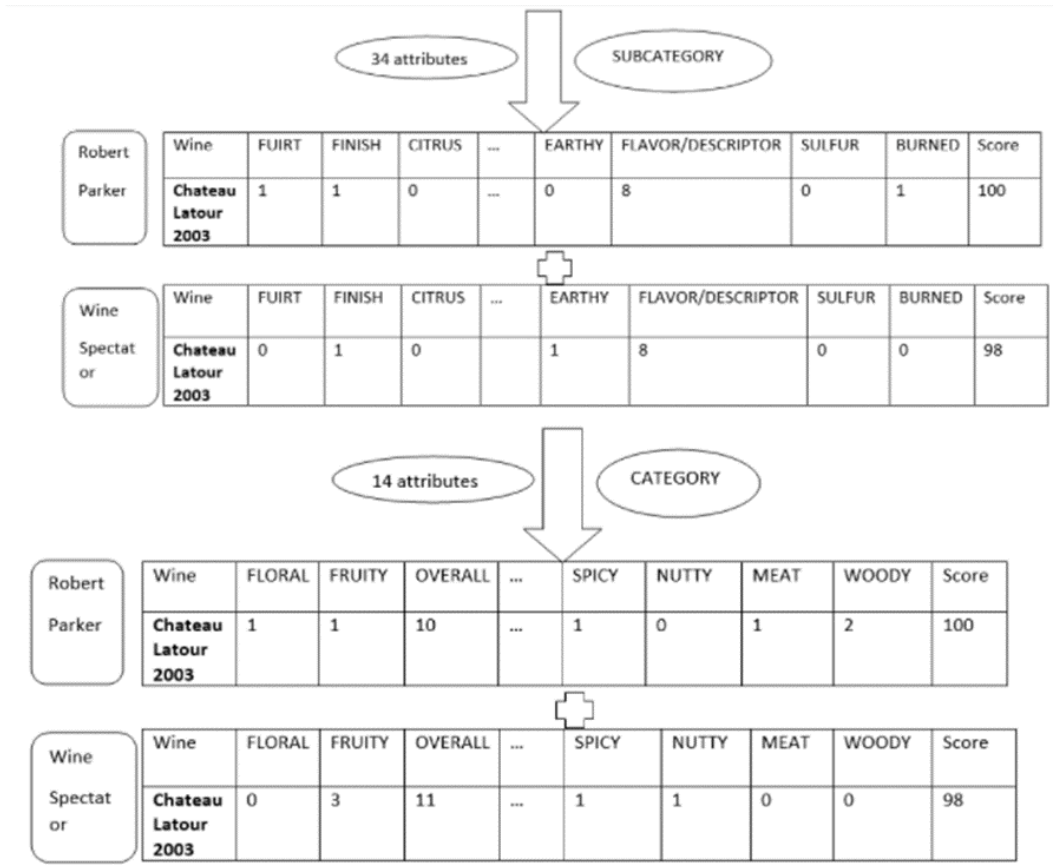

Figure 9. The Second and Third Step of Extracting Word with Computational Wine Wheel

2.4 Normalization on Category and Subcategory

The "NORMALIZED_NAME" attributes were the normalized format of the "SPECIFIC_NAME" attributes, because of the multiple formats of language expression. For instance, "good structure" and "impressive structure" were both normalized as "well-structured." The "SUBCATEGORY_NAME" attributes and "CATEGORY_NAME" were the summarized format of the "NORMALIZED_NAME" attributes, while "SUBCATEGORY" attributes were more detailed and "CATEGORY" attributes were more general. For instance, in Table 4, "lemon aromas" was normalized as "lemon" and "lime" was normalized as "lime" while both corresponded to "citrus" under the SUBCATEGORY_NAME. "Black currant" and "cassis" were both normalized as "black currant” and corresponded to “berry” under the SUBCATEGORY_NAME. All four attributes were categorized as “fruity” under the CATEGORY_NAME.

Table 4. Example of Corresponding Attributes

CATEGORY_NAME	SUBCATEGORY_NAME	NORMALIZED_NAME	SPECIFIC_NAME
fruity	citrus	lemon	Lemon aromas
		lime	lime
	berry	black currant	black currant
			cassis

The “NORMALIZED_NAME” attribute was a binary dataset, but the corresponding attributes under “SUBCATEGORY” and” CATEGORY” were continuous attributes. The continuous datasets needed to be normalized by applying the normalization algorithm to rescale their values to avoid the imbalance of weight of features. Normalization can not only efficiently speed up the computational process, but also make the understanding of the data easier. In this project, Min Max Normalization was used to rescale the values of the continuous datasets from 0-1. Z was the normalized value, x was the original value, min(x) was the minimum value among all x, and max(x) was the maximum value among all x. The equation is as follows.

$$z = \frac{x - \min{(x)}}{\max(x) - \min{(x)}}$$

CHAPTER 3 METHODS

Data mining supports the fundamental concept of extracting patterns and discovering relationships between parameters in huge amounts of data [23]. After the data preprocessing and management in Chapter 2, this chapter will present the methodology applied to the data collected.

As shown in Figure 10, data mining techniques as one of the essential cores in the Data mining field have provided effective analysis for decision makers in solving problems arising in particular areas. There are many various types of data mining techniques, so choosing the suitable technique for the corresponding domain or research goal can have a serious impact. Data mining techniques fall into three categories: classification, clustering, and association rule. Association rules mining is to discover the associations and relations among item sets of large data [24], clustering mining is to divide data into groups of objects that are similar to one another and dissimilar to objects in other groups [25], and classification mining is to predict the categorical class label based on the classification models. For the research purpose, classification is the main focus of this project.

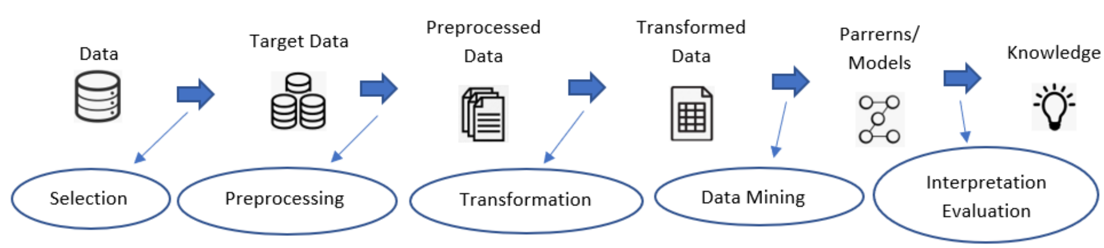

Figure 10. Data Mining Process

3.1 Classification

Classification techniques have become one of the most useful and important techniques for data science. The principal concept of classification is to categorize new data by using the previously categorized data as a basis. Therefore, two sets of data are required for classification; one is the previously categorized dataset that is called the training dataset, another one is the new dataset that is called the testing dataset. There are two steps in classification. As demonstrated in Figure 11, the first step is constructing a model based on the training dataset by applying classification algorithms, and the second step is classifying the testing dataset into a class label by using the model [26].

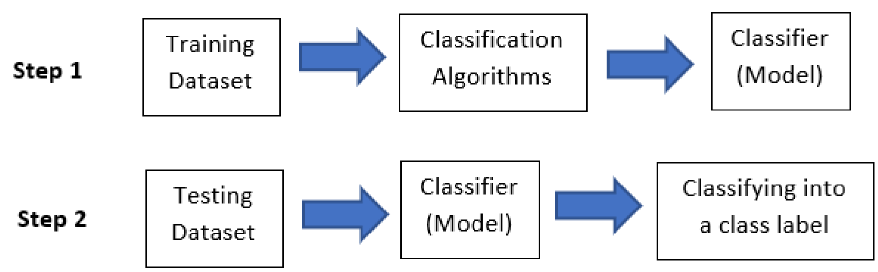

Figure 11. Classification Steps

As elaborated in Chapter 2, the cost of data collection and organization is not cheap; therefore, the collected data is expected to be entirely utilized. To satisfy the classification algorithms, training datasets and testing datasets are both required. In this case, how would one maximize the utility of the limited amount of data? The cross-validation method is one of the best choices to achieve this goal.

3.1.1 Cross Validation

The statistical method, cross-validation, is used to evaluate and compute learning algorithms by dividing data into two parts: one part for training and the other one for testing.

The typical mechanism of cross-validation is that both training and testing sets must cross over in successive rounds so that each data point has chances to be used as a training data point to build the model and a testing data point to be validated, which utilizes the dataset as much as possible. In this project, fivefold cross-validation was used. In the fivefold cross-validation, the data is partitioned into five equally sized folds firstly. Five iterations of training data and testing data are performed so that each iteration contains a different fold of testing data while the rest of the data are used as training data [27]. By using fivefold cross-validation, one data point will not be present in both the training dataset and testing dataset; therefore, the model will not be able to cheat. As shown in Figure 12, the orange highlighted fold is the testing dataset, while the blue highlighted folds are the training datasets. The training dataset in each iteration is much more than the testing dataset because the classification model should be done in a way that has enough instances to train on to avoid giving the poor results when used for testing [28].

Iteration 1	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Iteration 2	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Iteration 3	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Iteration 4	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Iteration 5	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5

Figure 12. Five-Fold Cross Validation

In this project, one more step was added into the fivefold cross-validation process to avoid the possible imbalance class problems in each fold, which was separating the data based on their class labels at first. Then we randomly selected 1/5 portion of data from each separated class dataset and combined them into one-fold. After applying the fivefold cross-validation on the data, five sets of training data and five sets of testing data were ready to be processed by the classification algorithms.

3.1.2 Black Box & White Box

The first step of classification as mentioned before is to build the classification model. Classification models are built with training data by applying the classification algorithms, which can be categorized into two approaches: black-box models and white-box models. Black-box models are considered as the proposals that are very difficult to explain and to be understood in practical applications, which include those proposals containing a complicated mathematical function, distance function, and the representation space [29]. One of the most common black-box-based models is the hyperplane-based models, such as support vector machines [30], which use a subspace to separate the problem’s classes. Another important black-box-based model type is inspired by the biological neural networks, which exploit the possible functional similarities between the human brain and artificial information processing systems [31].

White-box models, also called the explainable models, are those models based on patterns or rules, which can be understood and explained in practical applications [29]. The common model decision trees [32], a good demonstration of a white-box model, provide accurate results and understandable models for analyzing the application domain. Some other popular white-box models are linear regression models [33], Bayesian Networks [34], Naive Bayes [35], and so on.

To evaluate a classification model is not only according to the accuracy of its predictions, but also relates to the explainability and the interpretability of its predictions [36]. The debates between black-box and white-box models have never stopped since the essential difference between these two models is the trade-off of accuracy and the interpretation of predictions. In general, black-box models are regarded as having high performance of accuracy but hard to interpret, while white-box models are regarded as having understandable internal logic and structure with simpler computation and higher cost [37]. In this project, SVM classifier as the representative of black-box model and naive bayes classifier as the representative of white-box model were implemented and applied on datasets, so advantages and disadvantages on both types of models can be deeply studied and compared.

3.2 Support Vector Machines

Support Vector Machine (SVM), one of the most popular machine learning algorithms, was introduced in 1992 by Boser, Guyon, and Vapnik in COLT-92 [38]. Support vector machines are a set of related supervised learning methods for classification and regression. The mapping function is used to predict which maps independent variables to dependent variables. The mapping function is a hyperplane or a set of hyperplanes in a high-dimensional space to separate data points into classes.

For instance, two classes of data points in Figure 13 are represented by the red color and green color. A line was drawn to separate the two classes, which is called the decision boundary in SVM. A hyperplane is the best choice of the decision boundary, which means it separates the data in the best way. The distances from the decision boundary (black line) to the nearest data points from both sides are called margins. While the margins from both sides are equal, the margins are the maximum values, and they form the maximum margin classifier. For real world data, choosing the decision boundary that allows misclassifications is essential; therefore, the soft margin classifier, also called support vector classifier, is used. The name support vector classifier comes from the fact that the data points on the edge and within the soft margin, the data points in the black circle in Figure 14, are called support vectors.

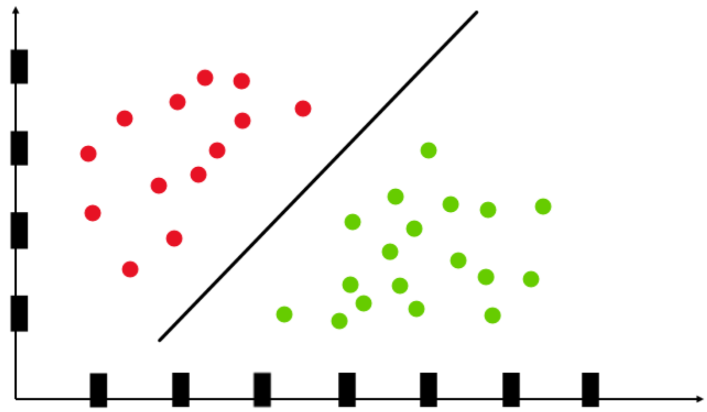

Figure 13. Example 1 of SVM

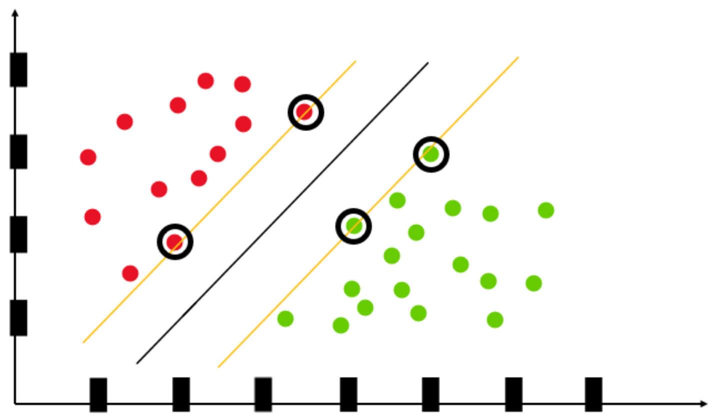

Figure 14. Example 2 of SVM

When the data are one-dimensional, the support vector classifier is a single point; when the data are two-dimensional, the support vector classifier is a line. However, not all datasets are linearly separable, which means they cannot be separated with a straight line like the example in Figure 13. For instance, the dataset in Figure 15 is not separable linearly. To separate this dataset, we start with adding a Y-axis on the graph, as shown in Figure 16. The Y value for each data point equals the square of X value. Now each data point has X and Y coordinates, which makes the data two-dimensional. Then a support vector classifier, the red line in Figure16, can be drawn to separate the two classes. The main idea of SVMs is moving the data into a relatively higher dimension, and then finding a support vector classifier that separates the higher dimensional data into groups. The mechanism is very powerful and can be applied to infinite dimensional data because the data can always be separated by its one-dimensional higher space.

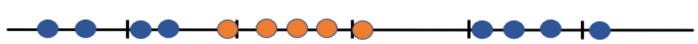

Figure 15. Example 3 of SVM

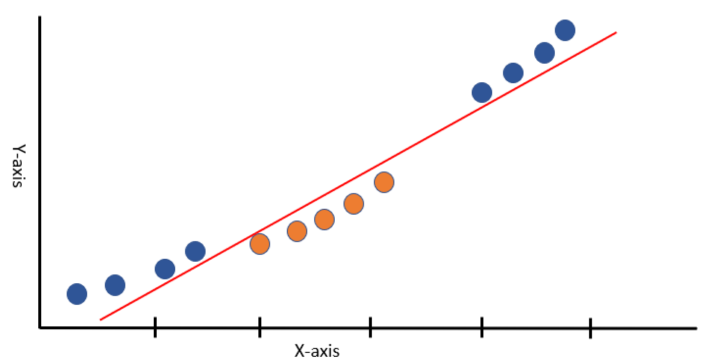

Figure 16. Example 4 of SVM

A kernel function is used to systematically find the support vector classifier in higher dimensions. In the example of Figure 16, the Polynomial Kernel is used. The equation for Polynomial kernel is $K(x, y) = (x^T y+c)^d$. Some other popular kernel functions include gaussian radial basis function [39], linear kernel, multi-layer perceptron, and so on [38].

Finding the support vector classifier is the most critical part in SVM algorithms. To implement SVM classification, two sets of input data are taken. The first input dataset is the training dataset, which finds the support vector classifier; the second input dataset is the testing dataset, which is used to predict the classes by mapping the data points based on the support vector classifier. In this project, SVM light [40] with the linear kernel was used.

3.3 Naive Bayes

Naive Bayes classifiers are a collection of white-box classification algorithms based on Bayes' Theorem, which follow the principle that making the predictions probabilistically, and every pair of features being classified is independent of each other. Bayes' Theorem is named after a eighteenth century mathematician Thomas Bayes, which is a mathematical formula for determining conditional probability. Conditional probability is regarded as the likelihood of an event or outcome occurring, according to the occurrence of the existing events. For example, event A is winning the red box in the raffle, and there is a 60% chance of winning the red box. Event B is having candies in the red box, and there is a 50% chance of having candies in the red box. Therefore, the probability of winning the red box and having candies in the red box is equal to P (candies | red box) • P (red box) = (0.50) • (0.60) = 0.3. Bayes' Theorem provides a way to update the existing prediction by incorporating new or additional evidence. The formula of Bayes' Theorem is as follows.

$$P(A|B) = \frac{\mathrm{P(A) * P(B|A)}}{\mathrm{P(B)}}$$

P(A): The probability of A occurring

P(B): The probability of B occurring

P(A|B): The probability of A given B

P(B|A): The probability of B given A

In practical datasets, it's more likely to have more than one attribute of one data point and more than one class in one dataset. Therefore, the naive bayes classifier is used. While B represents n attributes of data point as a vector B = (b_1, b_2, …, b_x), A has m classes in the dataset (a_1, a_2, … a_y). The formula of the naive bayes is as follows.

$$class\ a_1:\ P(a_1|B)\ =\ \frac{P(a_1)\cdot P(b_1|a_1)(P(b_2|a_1)...P(b_n|a_1)}{P(b_1,b_2,...b_n)}$$

$$class\ a_2: P(a_2|B)\ =\ \frac{P(a_2)\cdot P(b_1|a_2)(P(b_2|a_2)...P(b_n|a_2)}{P(b_1,b_2,...b_n)}$$

If there are y classes in the dataset, one data point will have y probabilities. The naive bayes classification is to derive the maximum posterior probability, which is the maximal P(A|B).

The datasets in this project contained both binary attributes and continuous attributes. In order to implement Naive Bayes on the datasets, the continuous attributes were converted into binary attributes. Two sets of continuous attributes were in the datasets, one set was the "CATEGORY" attributes, another set was the "SUBCATEGORY" attributes. Both sets of attributes were the counts of the frequency. Therefore, converting the continuous attributes into binary attributes was to set the frequency of the "CATEGORY" attributes and "SUBCATEGORY" attributes as binary. For example, if the "fruity" under the "CATEGORY" attribute counted 3 for this wine, it was signed "1" under the binary attribute "fruity3;" if the "fruity" counted 0 times for this wine, it was signed "1" under the binary attribute "fruity0." After converting the continuous attributes into binary attributes, the Bernoulli Naive Bayes classifier for binary attributes was implemented.

3.3.1 Bernoulli Naive Bayes

Bernoulli Naive Bayes is a variant of Naive Bayes, which is often used for discrete data, where attributes are only in binary form. The formula is as follows.

$$p(A_i|B) = \frac{N_{ic}+1}{N_c + C}$$

N_{ic}: The number of reviews containing attribute A_i and in class B

N_c: The number of class B

The probability to calculate is the frequency of the key word occurring in the wine reviews. For example, there are 100 reviews in the dataset with 60 negative classes. In the 60 negative classes, 23 classes contain the attribute "blueberry." Therefore, the probability P("blueberry"|negative) = 23/60=0.38. However, not all attributes occur in the dataset; therefore, Laplace smoothing was used to avoid zero prior probabilities by adding 1 to the dividend and adding the number of values in B to the divisor.

3.4 Evaluation Matrix

To evaluate the result, four different statistical measures were used: True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN). As shown in Table 5, a true positive prediction means that the prediction of the model was correct, and the wine was predicted as positive class; a false positive prediction means that the prediction of the model was incorrect, and it predicted the wine was also positive class. A true negative prediction means that the prediction of the model was correct, and the wine was predicted as a negative class; a false negative prediction means that the prediction of the model was incorrect, and the wine was predicted as a negative class.

Table 5. Evaluation Matrix

Evaluation Matrix	Predicted (positive)	Predicted (negative)
Actual (positive)	TP	FN
Actual (negative)	FP	TN

To better understand the result of the classification, four metrics of measurements were used as well: Accuracy, Precision, Recall, and F-score. Accuracy was the percentage of wines that have been correctly classified across all the wines. Essentially it tells us how many wines were correctly predicted as positive and negative.

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision, also called positive predictive value, was the number of wines that were predicted as positive and were correct. This showed how many wines were positive and being correctly predicted out of all the positive predictions.

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall, also known as sensitivity, was the proportion of the positive wines that were predicted correctly. This metric showed how well the model can correctly predicted wines that were positive.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F-score was a measure that took the mean of the combined precision and recall. This metric looked at precision and recall as a single metric and measured how many of the positive wines were predicted correctly.

$$F - Score = 2 * \frac{\text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})}$$

CHAPTER 4 ENHANCING CWW WITH ROBERT PARKER REVIEWS

4.1 Robert Parker

When wine reviews were written by those who had close relationships with the wine industry before the 1970s, it was more like wine promotion than critical assessment [42]. A United States lawyer, Robert Parker, who had no formal training in the wine tasting, decided to use his reputedly olfactory and gustatory talents for describing and critiquing the qualities of wine with his own way. Robert M. Parker, Jr., one of the world's most influential wine critics, has been regarded as the most successful in the field of wine writing [43]. Robert Parker's wine reviews are described as "remarkably powerful contemporary rhetoric which has had an unprecedented impact in the world of prestigious wine for more than two decades" [31]; therefore, unlike Wine Spectator's reviews, which are shorter and more precise, Robert parker's reviews are more descriptive and colorful.

As shown in Table 6, two wine reviews about the 2017 E. Pira e Figli Barolo Mosconi are from Robert Parker (the left column) and Wine Spectator (the right column). The review from Robert Parker contains much more information than Wine Spectator, not only the taste of the wine quality, but also wine-related information such as it's from the "hot vintage." On the other hand, the review from Wine Spectator is much shorter and pithy with very straight forward descriptions of the taste of the wine. Studying these two different types of reviews is of great significance to the research of this project.

Table 6. Wine Review of E. Pira e Figli Barolo Mosconi 2017 from Robert Parker and Wine Spectator

Robert Parker's Review	Wine Spectator's Review
Made with organically grown clones of Michel and Lampia, the E Pira-Chiara Boschis 2017 Barolo Mosconi softly presents bright berry aromas, raspberry, wild cherry, crushed limestone and delicate floral tones of lilacs and violets. Like other wines from this hot vintage, this expression from the Mosconi cru in Monforte d'Alba has that unique floral signature that is precious and unexpected. The wine shows great depth and balance with a pretty intensity that spreads over the palate. The tannins are dry with some dustiness, but the mouthfeel is spot-on in terms of length and polish.	Cherry and plum fruit flavors are accented by vanilla, toast, hay, white pepper and tar notes in this expressive, solidly built Barolo, which is fluid, with a dense matrix of tannins shoring up the long finish, showing fine complexity, balance and length. Best from 2025 through 2048.

4.2 Computational Wine Wheel Update

As mentioned in Chapter 2, Computational Wine Wheel has been the data preprocessing tool that used on the wineinformatics research in the past. CWW works as a dictionary using one-hot encoding to convert the attributes of wine reviews to vectors. However, the Computational Wine Wheel was built based on the reviews of Wine Spectator and was used on the reviews of Wine Spectator as well. In this project, all the words in the CWW that generated from Wine Spectator's reviews were used to map the attributes of the reviews from Robert Parker, which is a brand-new performance since ever. Based on the study of the reviews from Robert Parker and Wine Spectator from the last section, the difference between the reviews of Wine Spectator and Robert Parker is obvious and distinguished. To improve the efficiency of the application on the CWW, as well as to make it more compatible with the reviews from Robert Parker, the update iteration of CWW was necessary.

An update iteration for Computational Wine Wheel means enlarged the vocabulary, which was done by adding new wine attributes that were extracted from new wine reviews. For expanding the wine attribute dictionary, the Robert Parker wine reviews of 513 wines that are the 1855 Bordeaux Wine Official Classification made in the 21st century (2000-2020) were used for new attributes extraction.

4.3 Hand Extraction of CWW3.0

The extraction process for CWW3.0 was purely manual, and it contained a few steps. For example, Figure 17 shows a wine review of the 2016 Mouton Rothschild from Robert Parker. The first step was hand extracting the attributes that would be important to determine the quality of the wine and listing them as shown in Table 7 under the column "Hand extracted attributes." The second step was to programmatically extract the attributes from the Robert Parker review by using the CWW2.0 to see how many attributes the CWW2.0 already contained. The second column in Table 7 named "Program extracted attributes" shows the attributes extracted by the program. The third step was to compare the attributes that were extracted by hand with the attributes extracted by program of CWW2.0. The third column in Table 7 named "Common Attributes" displays the attributes extracted both by hand and program. To check the efficiency of the attribute extraction of the CWW2.0 was to figure out how many important attributes the program actually extracted. The extraction rate equals the count of attributes both extracted by hand and program (column "Common Attributes" in Table 7) divided by the count of hand extracted attributes (column "Hand extracted attributes" in Table 7). As shown in Table 7, the common attributes count was 20, divided by the hand extracted attributes count 26, so the extraction rate was 20/26 = 77%. By looking at the extraction rate, we know how many percent of the attributes that the CWW2.0 contains after applying it on Robert Parker's reviews. The last step was to add new attributes that were useful to the CWW.

Chateau Mouton Rothschild 2016

Composed of 83% Cabernet Sauvignon, 15% Merlot, 1% Cabernet Franc and 1% Petit Verdot, the 2016 Mouton Rothschild has an opaque garnet-purple color. WOW—the nose explodes from the glass with powerful blackcurrant cordial, black raspberries, blueberry pie and melted chocolate notions, plus suggestions of aniseed, camphor, lifted kirsch and the faintest waft of a subtle floral perfume in the background. Full-bodied, concentrated, bold and totally seductive in the mouth, it has very fine-grained, silt-like tannins, while jam-packed with tightly wound fruit layers, finishing in this wonderful array of mineral sparks. Magic.

Figure 17. Hand Extraction Process Example 1.

Table 7. Example of Extraction Rate Progress

Hand extracted attributes	Program extracted attributes	Common Attributes
powerful, blackcurrant, black raspberries, blueberry, pie, melted chocolate, aniseed, camphor, kirsch, subtle, floral, full-bodied, concentrated, bold, seductive, fine-grained, silt-like tannins, jam-packed, tightly wound, fruit, layers, finishing, wonderful, mineral, sparks, magic,	powerful, black raspberries, blueberry, pie, melted chocolate, kirsch, subtle, floral, full-bodied, concentrated, bold, seductive, jam-packed, tightly wound, fruit, layers, finishing, wonderful, mineral, sparks, purple color, tannins, explodes,	powerful, black raspberries, blueberry, pie, melted chocolate, kirsch, subtle, floral, full-bodied, concentrated, bold, seductive, jam-packed, tightly wound, fruit, layers, finishing, wonderful, mineral, sparks,
Total count: 26	Total count: 23	20

The review in Table 7, the hand extracted attributes that were not in the common attributes were considered to add into the CWW.

From this review, six new attributes were added into the CWW. In the CWW2.0, “BLACKCURRANT” existed as a “BLACK CURRANT;” however, the program was not able to extract “BLACKCURRANT” as “BLACK CURRANT,” adding “BLACKCURRANT” into the CWW was important and the normalized name of it was the “BLACK CURRANT.” “FINE-GRAINED,” “SILT-LIKE TANNINS,” “ANISEED,” “CAMPHOR,” and “MAGIC” as new descriptive attributes were also added into the CWW.

Broadly speaking, three major types of attributes were added. The first type was the new savory/descriptive attributes. Because the styles of reviews from Wine Spectator and Robert Parker are so different, so many descriptive words that Robert Parker uses frequently that rarely show up in Wine Spectator's reviews. For example, "CUMIN SEED," "ROSE HIP TEA," "TILLED SOIL," "VEGETABLE," "BLACK FOREST CAKE," and so on. However, considering the actual frequency of utilization, as well as avoiding redundancy, those descriptive attributes that show up rarely were not added, such as "ACACIA FLOWER," "TART ACID," "SMORGASBORD," and so on.

For the situation of having a long group of words that were not extracted as one attribute, such as "INTENSE AROMATICS," the program extracted "INTENSE" and "AROMATICS," which was not an accurate description and also caused the redundancy. Therefore, the second type of attributes, the multiple-word groups, were added. For example, "INTENSE AROMATICS" was normalized as "INTENSE."

The third type was adding different formats of existing attributes in CWW2.0. This scenario can be very tricky. For example, in Figure 18, the attribute "FRESH FIGS" was extracted by the program as "FRESH" and "FIGS," and they were normalized as "FRESH" and "FIGS." Let's look into Figure 18, the words "FIGS" and "FRESHNESS" were extracted by the program as "FRESHNESS" and "FIGS," as well as normalized as "FRESH" and "FIGS." However, these two reviews had big differences between the "FIGS" while they were normalized as the same attributes. The "FRESH FIGS" in Figure 18 should be extracted as one attribute and then normalized as "FIGS," while the "DRIED FIGS" and "FRESHNESS" in Figure 19 should be extracted as two attributes and then normalized as "DRIED FIGS" and "FRESH." Therefore, two new attributes "FRESH FIGS" (normalized as “FIGS”) and “DRIED FIGS” (normalized as “DRIED FIGS”) were added. Adding this type of attributes was one of the most important steps out of the whole enhanced CWW project, because this step significantly improved the accuracy and reduced the redundancy of CWW.

Château Coutet 2010

Medium lemon-gold colored, the 2010 Coutet is relatively evolved, revealing honey nut, ginger ale and baked peaches scents with touches of straw, fresh figs, quince paste and nutmeg. The palate is considerably more mature than when I last reviewed this wine a couple of years age, so I suspect this bottle may have been slightly oxidized, nonetheless it offers loads of nutty nuances accenting the dried fruits, finishing deliciously spicy.

Hand Total: 12

Program extracted attributes:
['dried fruit', 'baked peach']
['straw', 'finishing', 'spicy', 'nut', 'nutmeg', 'nutty', 'deliciously', 'fresh', 'quince', 'figs', 'ginger', 'honey']

Program Total: 14
Extraction Rate: 6/12=50%
Differences: honey nut,ginger ale,fresh figs,mature,oxidized,finishing deliciously spicy,

Figure 18. Hand Extraction Process Example 2.

Chateau Mouton Rothschild 2001

Deep garnet-brick colored, the 2001 Rauzan-Ségla rolls out of the glass with open, expressive, mature notes of potpourri, star anise, incense and cigar box with a pretty core of kirsch, redcurrant jelly and dried figs. The light to medium-bodied palate delivers mature spice and dried berry layers with a light grip of chewy tannins and bags of freshness, finishing with a peppery kick. The blend this year is 63% Cabernet Sauvignon, 34% Merlot and 3% Cabernet Franc, harvested between September 28 and October 11 at an average yield of 46 hectoliters per hectare. It was aged for 18 months in French oak, 50% new. The alcohol is 12.5%.

Hand Total: 22

Program extracted attributes:
['star anise', 'chewy tannins', 'medium-bodied', 'dried berry']
['pretty', 'oak', 'kirsch', 'finishing', 'freshness', 'core', 'deep', 'cigar', 'alcohol', 'expressive', 'layers', 'light', 'peppery', 'incense', 'spice', 'delivers', 'figs', 'grip']

Program Total: 22
Extraction Rate: 16/22=73%
Differences: mature,potpourri,cigar box,dried figs,redcurrant,

Figure 19. Hand Extraction Process Example 3.

Eventually, 769 new attributes were added into the CWW3.0. The next process was to get rid of unnecessary attributes. The most essential principle of this step was to reduce the redundancy for programming. Similar to adding attributes in the last process, there were three types of attributes for deletion. Firstly, the plural formats were very important for programming extraction. During our programming extraction, the one-word attributes were separated from multi-word attributes. One-word attributes always need to have both single and plural in CWW so that both formats can be extracted. While in the case where the plural is added directly to s or es, multi-word attributes just need the single format in CWW and both single and plural will be extracted. Therefore, the unnecessary multi-word attribute plurals were deleted, such as "RED CURRANTS," "ASIAN SPICES," "SMOKED MEATS," and so on.

Secondly, the needless word groups were removed as well. For example, "HINTS OF ORANGE," "TOUCH OF SMOKE," "NOTES OF BEEF," these attributes were normalized as "ORANGE," "SMOKE," and "BEEF." In the vocabulary of CWW, "ORANGE," "SMOKE," "BEEF" exists, and those attributes will always be extracted and normalized no matter what. Therefore, removing the redundant word groups helped the efficiency of the programming. The meaning of having word groups in CWW was to capture either a different or more accurate description, otherwise it's meaningless.

The last type to delete was the very broad attributes that show up frequently in reviews without clear direction, such as "WEIGHT," "STRUCTURE," and so on. For example, in a review of "the plate has fantastic intensity with a very elephant, modest weight, featuring super-ripe…", the program extracted "WEIGHT" while it did not have a clear expression about how the wine was. Instead of extracting "WEIGHT," "MODEST

WEIGHT" is more meaningful. Ultimately, 111 attributes were deleted from the Computational Wine Wheel.

4.4 Computational Wine Wheel 2.0 vs. Computational Wine Wheel 3.0

After the processes of adding attributes and deleting attributes, as shown in Table 8, the CWW3.0 final version ended up with 14 categories, 34 subcategories, 1191 normalized attributes, and 2589 specific terms. 632 new specific terms and 192 new normalized attributes were added.

Table 8. Comparison of CWW2.0 and CWW3.0

	CWW2.0	CWW3.0
Data Source	Wine Spectator	Wine Spectator + Robert Parker
Categories	14	14
Subcategories	34	34
Specific Terms	1932	2589
Normalized Attributes	986	1191

The updated iteration of Computational Wine Wheel, CWW3.0, was expanded by adding new attributes that were extracted from the wine review of Robert Parker. All of the wine reviews were from wines that were the 1855 Bordeaux Wine Official Classification made in the 21st century (2000-2020). The data source of CWW2.0 was wine reviews from Wine Spectator. CWW3.0 contained 657 more specific-term attributes and 205 more normalized attributes than CWW2.0

To have a better understanding of the differences between CWW2.0 and CWW3.0, Table 9 the CWW2.0 and Table 10 the CWW3.0 with detailed statistics are displayed. As mentioned in Chapter 2, the Computational Wine Wheel has multiple levels and branches to separate the very broad categories of wine attributes into more specific and detailed subcategories, which includes "CATEGORY_NAME," "SUBCATEGORY_NAME," “SPECIFIC_NAME,” and “NORMALIZED_NAME.” Table 9 is a detailed statistical record of CWW2.0, and Table 10 is the record of CWW3.0 with new added attributes counts.

Table 9. CWW 2.0 Statistical Record

CATEGORY_NAME	SUBCATEGORY_NAME	SPECIFIC_NAME	NORMALIZED_NAME
CARAMEL	CARAMEL	71	40
CHEMICAL	PETROLEUM	9	5
	SULFUR	11	10
	PUNGENT	4	3
EARTHY	EARTHY	72	31
	MOLDY	2	2
FLORAL	FLORAL	61	39
FRUITY	BERRY	49	28
	CITRUS	37	23
	DRIED FRUIT	67	60
	FRUIT	22	9
	OTHER	25	18
	TREE FRUIT	39	31
	TROPICAL FRUIT	48	27
FRESH	FRESH	41	29
	DRIED	25	21
	CANNED/COOKED	16	15
MEAT	MEAT	25	13
MICROBIOLOGICAL	YEASTY	5	4
	LACTIC	14	6
NUTTY	NUTTY	25	15
OVERALL	TANNINS	90	4
	BODY	50	23
	STRUCTURE	40	2
	ACIDITY	40	3
	FINISH	184	5
	FLAVOR/DESCRIPTORS	649	432
OXIDIZED	OXIDIZED	1	1
PUNGENT	HOT	3	2
	COLD	1	1
SPICY	SPICE	83	44
WOODY	RESINOUS	24	9
	PHENOLIC	6	4
	BURNED	47	26

Table 10. CWW 3.0 Statistical Record

CATEGORY_NAME	SUBCATEGORY_NAME	SPECIFIC_NAME	SPECIFIC_NAME Added	NORMALIZED_NAME	NORMALIZED_NAME Added
CARAMEL	CARAMEL	97	26	56	16
CHEMICAL	PETROLEUM	11	2	6	1
	SULFUR	11	0	10	0
	PUNGENT	4	0	4	1
EARTHY	EARTHY	128	56	47	16
	MOLDY	2	0	2	0
FLORAL	FLORAL	87	26	45	6
FRUITY	BERRY	84	35	39	11
	CITRUS	56	19	35	2
	DRIED FRUIT	76	9	65	5
	FRUIT	42	20	16	7
	OTHER	22	-3	9	-9
	TREE FRUIT	55	16	40	9
	TROPICAL FRUIT	67	19	36	9
FRESH	FRESH	75	34	44	15
	DRIED	50	25	39	18
	CANNED/COOKED	18	2	17	2
MEAT	MEAT	36	11	21	8
MICROBIOLOGICAL	YEASTY	5	0	4	0
	LACTIC	14	0	6	0
NUTTY	NUTTY	27	2	20	5
OVERALL	TANNINS	124	34	6	2
	BODY	61	11	17	-6
	STRUCTURE	51	11	2	0
	ACIDITY	61	21	4	1
	FINISH	233	49	13	8
	FLAVOR/DESCRIPTORS	889	240	467	35
OXIDIZED	OXIDIZED	2	1	2	1
PUNGENT	HOT	3	0	2	0
	COLD	1	0	1	0
SPICY	SPICE	85	2	53	9
WOODY	RESINOUS	31	7	12	3
	PHENOLIC	6	0	5	1
	BURNED	51	4	28	2

In Table 10, two extra columns are added that are named "SPECIFICE_NAME added" and "NORMALIZED_NAME added," which record the new added attributes amount. The yellow highlighted numbers are negative, which means those are the deleted attributes. The red highlighted numbers represent the largest amount of new added attributes. Newly added "SPECIFIC" and "NORMALIZED" attributes also affect the continuously increasing number of "CATEGORY" and "SUBCATEGORY." Therefore, all four branches in CWW were utilized for the classification.

CHAPTER 5 EVALUATION OF ROBERT PARKER AND WINE SPECTATOR

After updating the Computational Wine Wheel, the second goal of this research was to evaluate and compare the consistency of the wine reviews from Wine Spectator and Robert Parker by using the new Computational Wine Wheel 3.0. In this chapter, the results of implementing SVM classifiers and Naive Bayes classifiers on the datasets after applying the Computational Wine Wheel will be discussed. In order to better study the performance of the new iteration Computational Wine Wheel 3.0, the results of datasets after applying the last version CWW2.0 are demonstrated as the reference.

5.1 Classification Results

After applying the Computational Wine Wheel on the datasets, CWW3.0 can generate 1239 (14 CATEGORY + 34 SUBCATEGORY + 1191 NORMALIZED) attributes and CWW2.0 can generate 1034 (14 CATEGORY + 34 SUBCATEGORY + 986 NORMALIZED) attributes. Since there are three different sets of attributes, the input data were prepared differently for four different experiments based on the methodology used in one of the previous research [17] to maximize the power of the Computational Wine Wheel: The first experiment used wine reviews with only "CATEGORY" attributes, resulting in a continuous dataset with 14 attributes from both CWW2.0 and CWW3.0; The second experiment used reviews with only "NORMALIZED_NAME" attributes, resulting in a binary dataset with 986 attributes from CWW2.0 and 1191 attributes from CWW3.0, which is similar to that used in most previous studies [19, 46, 47]; The third experiment used reviews with both "CATEGORY" and "NORMALZIED_NAME" attributes, resulting in a mixed dataset with 1000 attributes from CWW2.0 and 1205 attributes from CWW3.0. The fourth experiment used reviews with "CATEGORY," "NORMALIZED_NAME," and

"SUBCATEGORY" attributes, resulting in a mixed dataset with 1034 attributes from CWW2.0 and 1239 attributes from CWW3.0, which provides all information that can be extracted from the computational wine wheel. Figure 20 demonstrates the attributes contained in each experiment. The four experiments method will be used in both Chapter 5 and Chapter 6 to demonstrate the result of classification. As introduced in Chapter 2, all five datasets from different data sources were used in this research and the results are provided in the following sections.

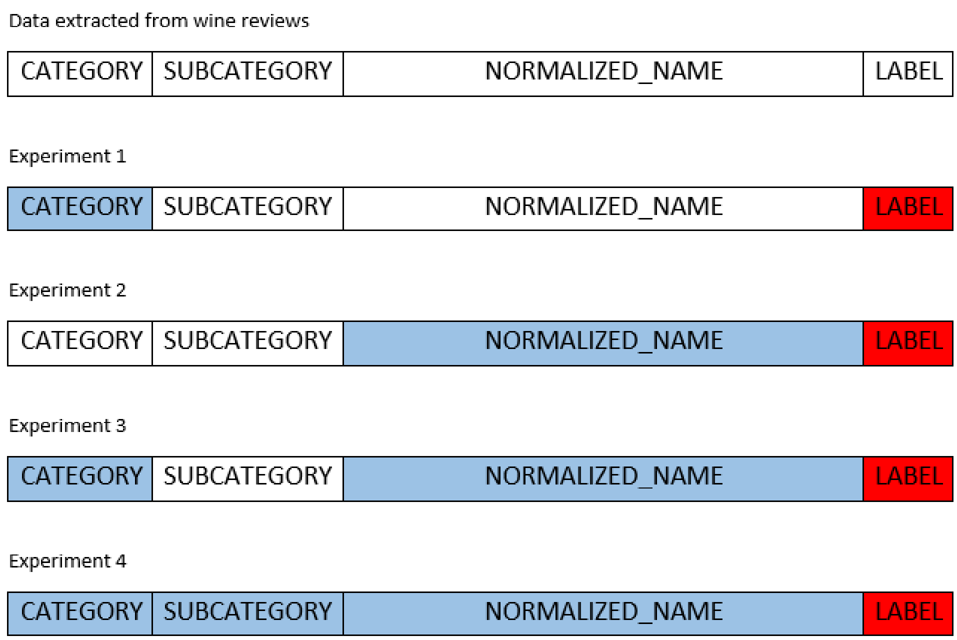

Figure 20. Four Experiments for Data Classification

5.1.1 Dataset 1 : 1855 Elite Bordeaux Wine

Dataset 1 contained 513 wines that were in the 1855 Bordeaux Wine Official Classification list made in the 21st century. Figure 21 compares the SVM accuracy results of all four experiments of Wine Spectator and Robert Parker after applying both Computational Wine Wheel 2.0 and 3.0.

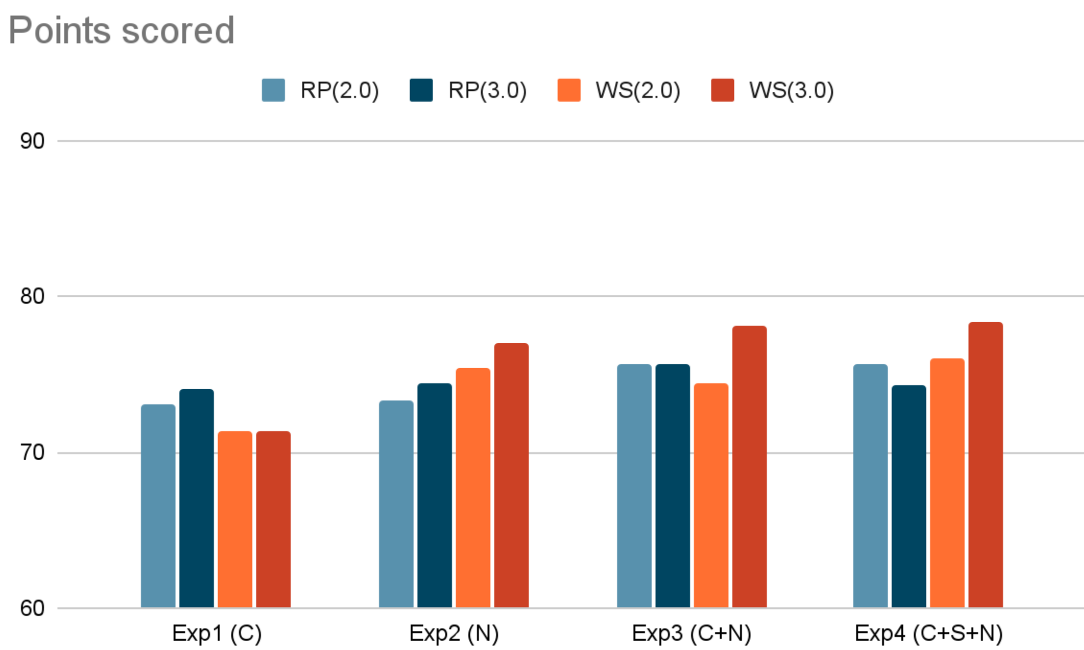

Figure 21. SVM Accuracy Results of Dataset 1

Looking at the two Wine Spectator columns, experiment 1 (CATEGORY attributes only) stayed the same on both, but other three experiments had obvious improvements after applying on the CWW3.0, especially experiment 3 improved 3.71%. However, looking at the two Robert Parker columns, both experiment 1 and 2 improved around 1%, while experiment 3 stayed the same and experiment 4 decreased 1.36%. These results were surprising because what was expected was the obvious improvement on Robert Parker's reviews instead of Wine Spectator since the updating of CWW3.0 was based on Robert Parker's reviews.

The second classification method was Naive Bayes. Figure 22 displays the accuracy results of implementing Naive Bayes Classification.

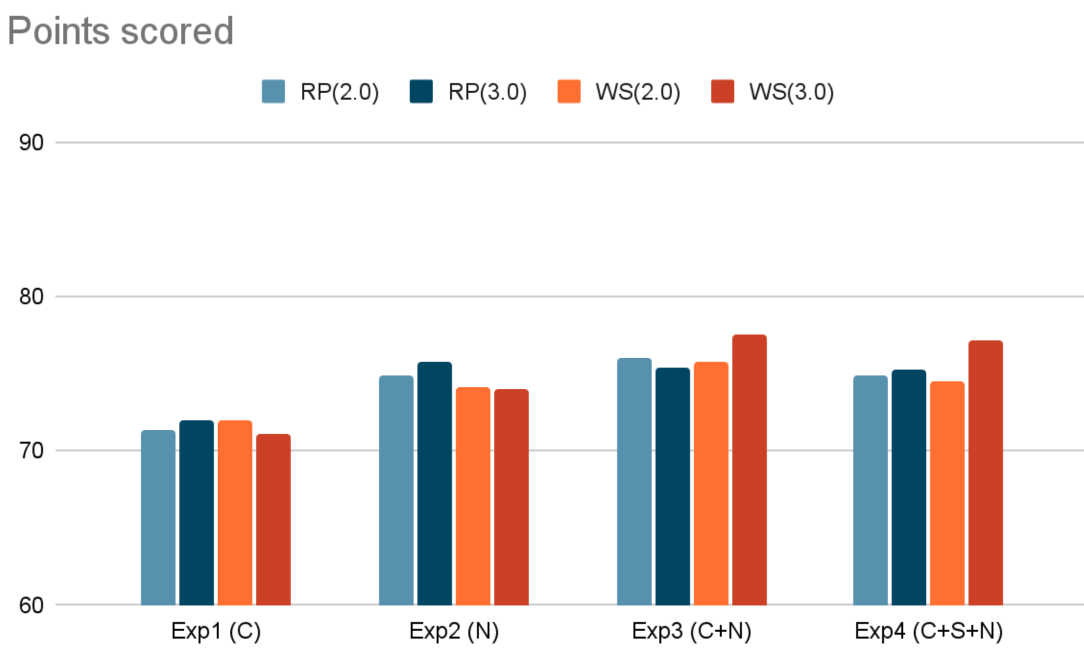

Figure 22. Naive Bayes Accuracy Result of Dataset 1

On experiment 2 (NORMALIZED attribute only), the results of Robert Parker increased 1% from applying CWW2.0 to CWW3.0. On both Experiment 3 and 4, Wine Spectator's datasets of CWW3.0 increased about 2-3% compared to applying CWW2.0, which are obvious improvements.

From Dataset 1's results, SVM classification achieved the highest result of 78.36%, which was the result of Wine Spectator after applying CWW3.0 on Experiment 4 (all three sets of attributes). The highest result from Naive Bayes was 77.52%, which was the result of Wine Spectator after applying CWW3.0 on Experiment 3 (CATEGORY + NORMALIZED attributes).

5.1.2 Dataset 2 : Bordeaux Wine

Dataset 2 included 2341 wines that come from Bordeaux, one of the most famous wine-making locations. Figure 23 displays the accuracy result after implementing SVM on dataset 2.

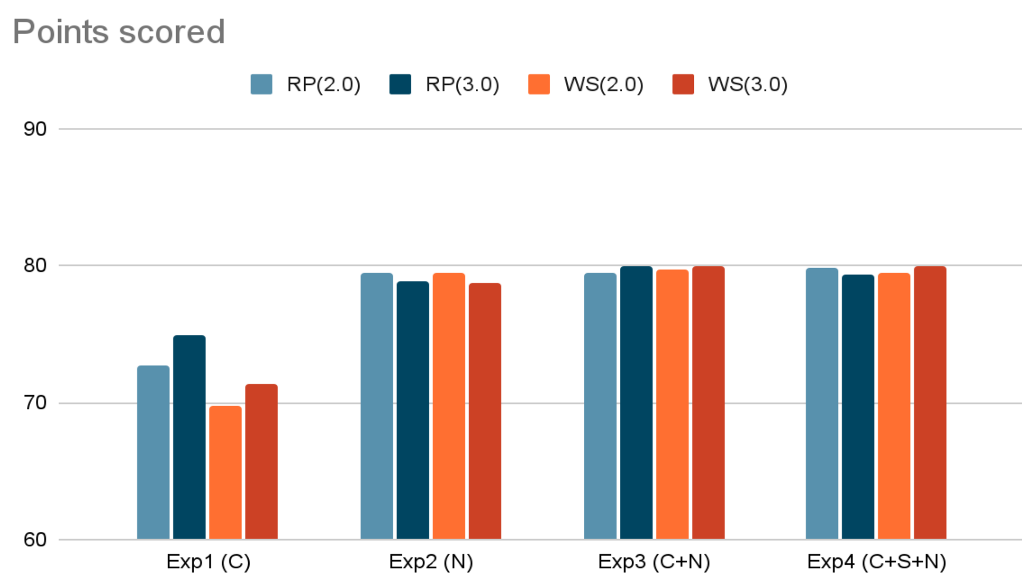

Figure 23. SVM Accuracy Results of Dataset 2

On Experiment 1, compared to results from applying CWW2.0 and CWW3.0, both Wine Spectator and Robert Parker's results increased from 1.5%-2%. However, on Experiment 2, compared to results from applying CWW2.0 and CWW3.0, both Wine Spectator and Robert Parker’s results decreased. The highest result from Dataset 2 was 79.97%, which happened twice, one was the result from Robert Parker after applying CWW3.0 on Experiment 3 (CATEGORY + NORMALIZED attributes), and the other one was the result from Wine Spectator after applying CWW3.0 on Experiment 4 (all three sets attributes).

Figure 24 displays the results after implementing Naive Bayes classification. Based on the pattern of the graph, except Wine Spectator on Experiment 2, all of the datasets after applying CWW3.0 increased compared to applying CWW2.0, which proves that the enhancement of the Computational Wine Wheel actually helped improve the accuracy.

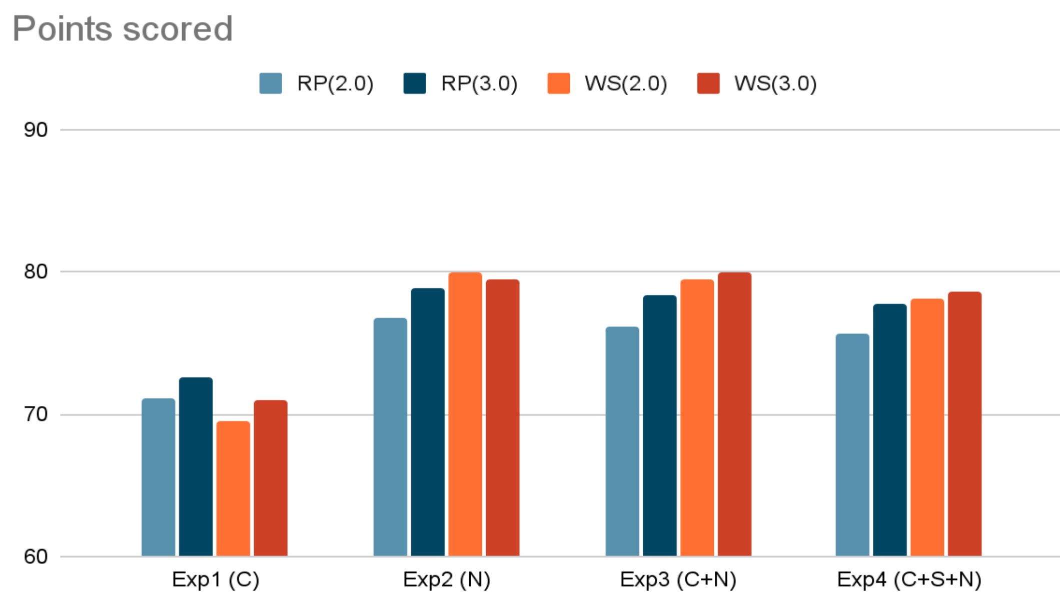

Figure 24. Naive Bayes Accuracy Result of Dataset 2

The highest result after implementing Naive Bayes on Dataset 2 was equal to the highest result of SVM, 79.97%, which was the result of Wine Spectator after applying CWW3.0 on Experiment 3. Compared to Figure 23 and Figure 24, the fluctuation and difference between Wine Spectator and Robert Parker after applying CWW2.0 and CWW3.0 were more obvious on Naive Bayes results than SVM.

5.1.3 Dataset 3 : Italy Wine

Dataset 3 contained 3198 wines that were produced in Italy, which is one of the oldest wine-producing regions in the world. Figure 25 displays the SVM accuracy results of Dataset 3.

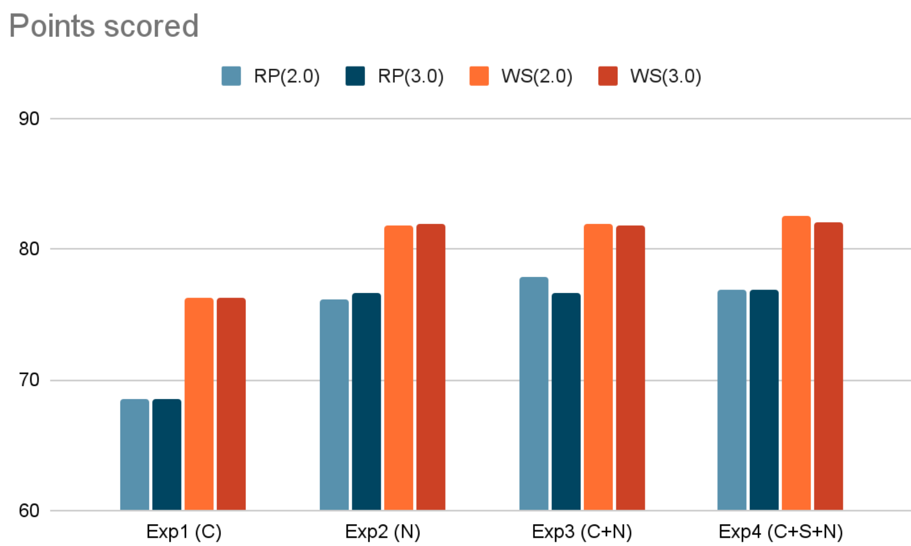

Figure 25. SVM Accuracy Results of Dataset 3

In Figure 25, the results of Wine Spectator after applying CWW3.0 had no obvious increase compared to applying CWW2.0, but a slight decrease on Experiment 3 and Experiment 4. The results of Robert Parker had an average of 6% lower accuracy than results of Wine Spectator in all four experiments. The highest result of SVM on Dataset 3 was 82.49%, which was the result of Wine Spectator after applying CWW2.0 on Experiment 4 (all three set attributes).

Figure 26 displays the results of Dataset 3 after implementing Naive Bayes classification.

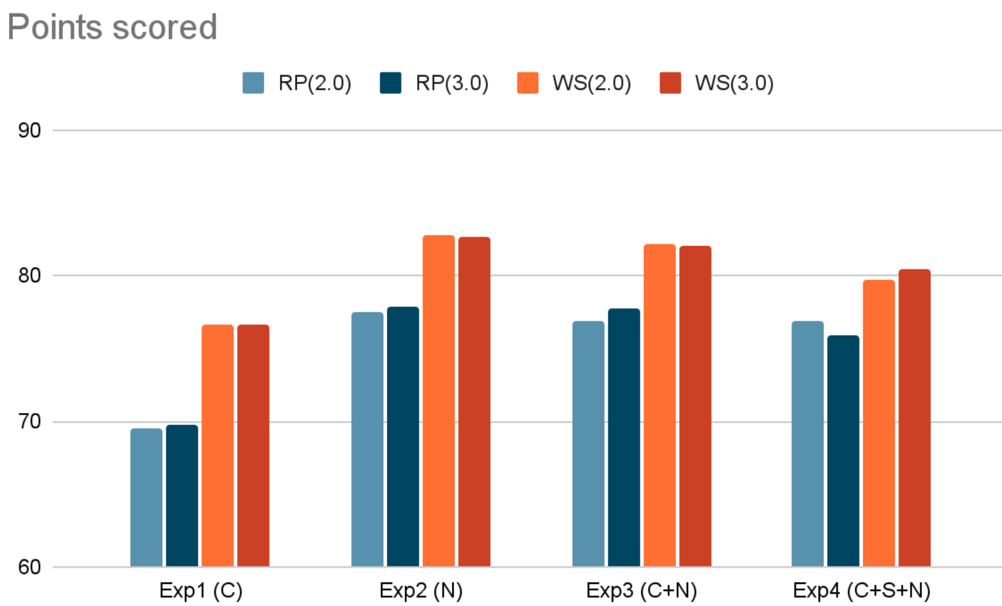

Figure 26. Naive Bayes Accuracy Result of Dataset 3

For Wine Spectator datasets, the results in Figure 26 after applying CWW3.0 decreased on Experiment 1, 2, and 3 compared to after applying CWW2.0, but a slight increase on Experiment 4. For Robert Parker datasets, the results after applying CWW3.0 increased on Experiment 1, 2, and 3 compared to after applying CWW2.0, but a slight decrease on Experiment 4, which was the opposite movement to Wine Spectator. The highest accuracy result from Naive Bayes classification was 82.8%, which was the result of Wine Spectator after applying CWW2.0 on Experiment 2 (NORMALIZED only). On all four experiments, Wine Spectator's results were higher than Robert Parker's, which was a similar movement as shown in SVM's Figure 25.

Compared to results of SVM, Naive Bayes results were all higher on Experiment 1, 2, and 3, except Robert Parker after CWW2.0 on Experiment 3. SVM achieved higher accuracies on Experiment 4. The highest accuracy from Dataset 3 was 82.8, which was the result of implementing Naive Bayes on Wine Spectator reviews after applying CWW2.0 on Experiment 2.

5.1.4 Dataset 4 : California Wine

Dataset 4 contained 4180 wines that were produced in California, which is one of the most important wine productions in the US industry. Figure 27 displays the SVM accuracy results of Dataset 4.

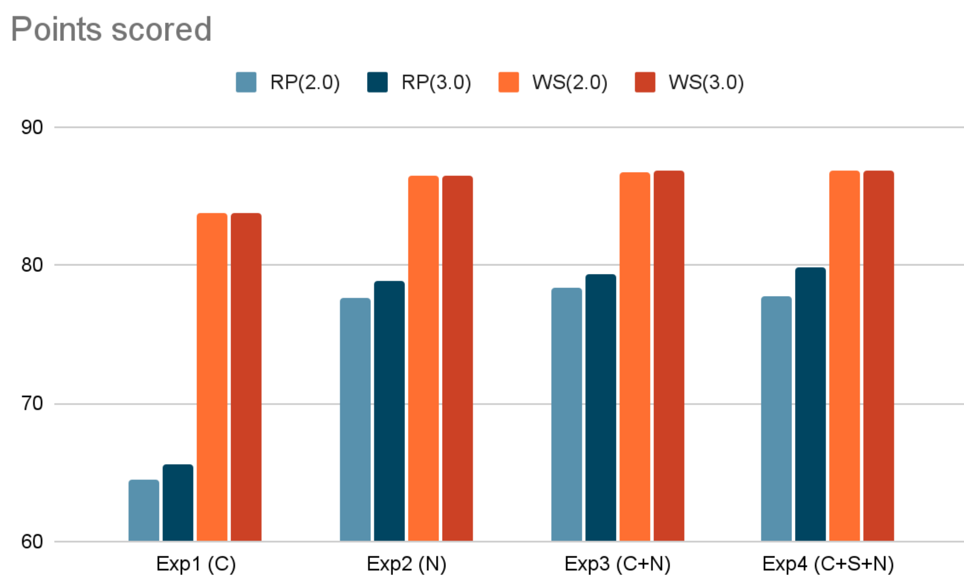

Figure 27. SVM Accuracy Results of Dataset 4

On Experiment 1, Robert Parker's accuracies in Figure 27 were almost 20% lower than Wine Spectator's accuracies. On Experiment 2, Experiment 3 and Experiment 4, Robert Parker's accuracies from after applying CWW3.0 all increased compared to after applying CWW2.0, while Wine Spectator's accuracies stayed almost the same after applying CWW2.0 and CWW3.0. The highest result after applying SVM on Dataset 4 was 86.8%, which was the result of Wine Spectator after applying CWW3.0 on Experiment 4 (all three set attributes).

Figure 28 shows the accuracy results of Dataset 4 after implementing Naive Bayes classification. The accuracies from Robert Parker were much lower than the accuracies from Wine Spectator.

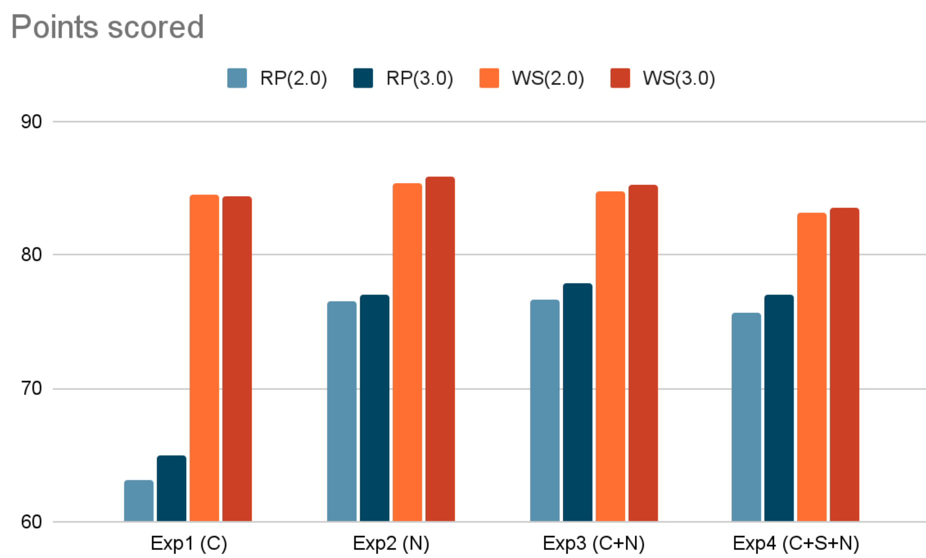

Figure 28. Naive Bayes Accuracy Result of Dataset 4

On all four experiments in Figure 28, the accuracies of Robert Parker after applying CWW3.0 increased compared to applying CWW2.0, while the Wine Spectator's accuracies also increased after applying CWW3.0 compared to CWW2.0 except Experiment 1 (CATEGORY only). The highest accuracy after implementing Naive Bayes on Dataset 4 was 85.84%, which was the Wine Spectator after applying CWW3.0 on Experiment 2 (NORMALIZED only).

The results from both SVM and Naive Bayes showed that the Wine Spectator datasets achieved much higher accuracies than Robert Parker. The results from SVM all achieved higher accuracies than the results from Naive Bayes, except the result of Wine Spectator dataset after applying CWW3.0. The highest accuracy result from Dataset 4 was

86.8%, which was the result of implementing SVM on Wine Spectator dataset after applying CWW3.0 on Experiment 4 (all three sets attributes).

5.1.5 Dataset 5 : All Datasets Combination

Dataset 5 contained 10,232 wines that were formed by the former four datasets, which contained all the wine data from this research. Figure 29 displays the SVM accuracy results of Dataset 5.

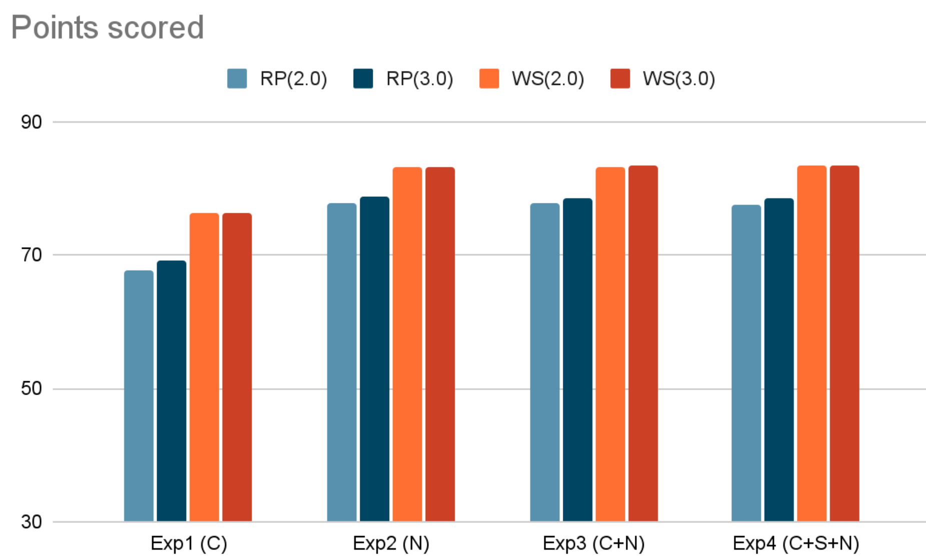

Figure 29. SVM Accuracy Results of Dataset 5

From the flow of the graph in Figure 29, it is obvious to see that the accuracies of Robert Parker after applying CWW3.0 increased compared to applying CWW2.0, while the accuracies of Wine Spectator after applying CWW3.0 stayed almost the same compared to applying CWW2.0. Because this dataset contained all wine data in this project, this flow explained that CWW3.0 increased the SVM accuracy for Robert Parker's wine reviews compared to Wine Spectator's wine reviews, which was understandable for CWW3.0 was expanded based on Robert Parker's reviews and CWW2.0 was based on Wine Spectator's reviews. The highest accuracy of Dataset 5 after implementing SVM classification was

81.95%, which was the Wine Spectator after applying CWW2.0 on Experiment 2 (NORMALIZED only).

Figure 30 displays the accuracy results of Dataset 5 after implementing Naive Bayes classification.

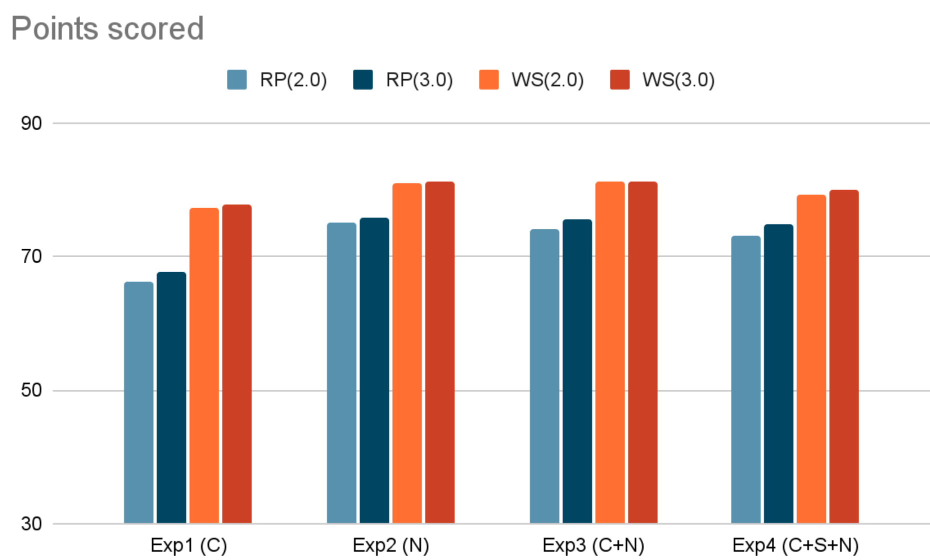

Figure 30. Naive Bayes Accuracy Result of Dataset 5

Figure 30 has the similar flow on the graph as SVM results in Figure 29, which was the increase on Robert Parker applying CWW3.0 compared to applying CWW2.0 and no obvious movements on Wine Spectator applying CWW3.0 compared to applying CWW2.0. The highest accuracy from Dataset 5 after implementing Naive Bayes was 80.32%, which was the result of Wine Spectator after applying CWW2.0 on Experiment 2 (NORMALIZED only).

For both Robert Parker and Wine Spectator's reviews, SVM achieved better accuracies on most experiments than Naive Bayes, except Wine Spectators on Experiment 1. The highest accuracy from Dataset 5 was 83.48%, which was the SVM result of Wine Spectator after applying CWW3.0 on Experiment 4 (all three sets attributes).

5.2 Dataset Results of Accuracy and Balance Statistic

This chapter contained five datasets. Each dataset contained two data sources, one was Robert Parker and the other one was Wine Spectator. Each data source generated two sub-datasets, one was the dataset after applying CWW2.0 and the other one was the dataset after applying CWW3.0. Each sub-dataset contained four experiments, which were Experiment 1 (CATEGORY attributes only), Experiment 2 (NORMALIZED only), Experiment 3 (CATEGORY + NORMALIZED attributes), and Experiment 4 (CATEGORY + SUBCATEGORY + NORMALIZED attributes). This project in this chapter in total contained 80 data subsections for classification. Two classification algorithms were implemented in this project, therefore, 160 sets of results with the evaluation metrics introduced in Chapter 3 were generated. However, it is not applicable to discuss every set of results in this chapter. The focus of the result discussion in this chapter is the accuracy, as demonstrated above. To have a better understanding on the application of the Computational Wine Wheel 2.0 and 3.0, the results of all five datasets were compared.

These five datasets were not all balanced datasets. Table 11 displays the positive and negative cases counts for all five datasets.

Table 11. Positive and Negative cases count for all datasets

	Wine Spectator 95+	Wine Spectator 94-	Robert Parker 95+	Robert Parker 94-
Dataset 1	183 (36%)	330 (64%)	197 (38%)	316 (62%)
Dataset 2	811 (35%)	1530 (65%)	975 (42%)	1366 (58%)
Dataset 3	759 (24%)	2439 (76%)	1008 (32%)	2190 (68%)
Dataset 4	677 (16%)	3503 (84%)	1945 (47%)	2235 (53%)
Dataset 5	2430(24%)	7802(76%)	4125 (40%)	6107(60%)

As shown in Table 11, Robert Parker had more balanced datasets compared to Wine Spectator, especially Dataset 4 contained 47% positive and 53% negative. Dataset 4 from Wine Spectator was the most imbalanced dataset, which contained 16% positive and 84% negative cases. The imbalanced dataset can easily cause high accuracy because the prediction tends to predict the high percentage class (in dataset 4 will be the negative class for Wine Spectator), which explains why the results from Wine Spectator achieved much higher accuracy than Robert Parker on Dataset 4.

5.3 Dataset Results of CWW2.0 and CWW3.0 Comparison

After enhancing the Computational Wine Wheel, the comparison of CWW2.0 and CWW3.0 proves the functionality of the new iteration CWW3.0. As mentioned in previous chapters, CWW2.0 was based on the reviews of Wine Spectator, while CWW3.0 was enhanced by using the reviews of Robert Parker. Therefore, to have a better vision of whether the CWW3.0 improves the results or not, this section will detail the four metrics of measurements introduced in Chapter 3: Accuracy, Precision, Recall, and F-score.

5.4 Evaluation Results

5.4.1 Accuracy Result

Table 12 displays the accuracy movement of each dataset. All five datasets' results were collected. The "RP/WS increased" columns show how many times the accuracy of the dataset after applying CWW3.0 increased compared to applying CWW2.0, while the "RP/WS decreased" columns show how many times the accuracy decreased. Each dataset contained two classification algorithms with four experiments, which ended at a total of 8 times differences. If there was an equal accuracy on both CWW2.0 and CWW3.0, it was not counted in this table.

Table 12. Statistical Record of Accuracy improvement after applying CWW3.0

	RP increased	RP decreased	WS increased	WS decreased
Dataset 1	5	2	5	2
Dataset 2	6	2	6	2
Dataset 3	4	2	2	4
Dataset 4	8	0	6	1
Dataset 5	8	0	7	0
Total	31/37=84%	6/37=16%	26/35=74%	9/35=26%

As shown in Table 12, Robert Parker's datasets had 84% possibility to increase the accuracy after applying CWW3.0 compared to CWW2.0, while Wine Spectator's datasets had 74% chance to increase the accuracy after applying CWW3.0.

5.4.2 Precision Result

Table 13 demonstrates the precision movements of all five datasets. The "RP/WS increased" columns show how many times the precision of the dataset after applying CWW3.0 has increased compared to applying CWW2.0, and the "RP/WS decreased" columns show how many times the precision decreased.

Table 13. Statistical Record of Precision improvement after applying CWW3.0

	RP increased	RP decreased	WS increased	WS decreased
Dataset 1	5	3	6	2
Dataset 2	7	1	7	1
Dataset 3	6	1	5	2
Dataset 4	8	0	5	2
Dataset 5	8	0	6	1
Total	34/39=87%	5/39=13%	29/37=78%	8/37=22%

As the “Total” column shown in Table 13, the precision of Robert Parker’s datasets after applying CWW3.0 had the possibility of 87% to increase, and the Wine Spectator’s dataset’s precision had the chance of 78% to increase.

5.4.3 Recall Result

Table 14 demonstrates the recall movements of all five datasets. The “RP/WS increased” columns count the times of the recall of the dataset after applying CWW3.0 increased compared to applying CWW2.0, and the “RP/WS decreased” columns show how many times the recall decreased.

Table 14. Statistical Record of Recall improvement after applying CWW3.0

	RP increased	RP decreased	WS increased	WS decreased
Dataset 1	4	2	7	1
Dataset 2	6	2	5	3
Dataset 3	4	3	2	5
Dataset 4	7	1	3	4
Dataset 5	8	0	4	3
Total	29/37=78%	8/37=22%	21/37=57%	16/37=43%

Table 14 calculates the final possibility of recall increase based on the records of all five datasets. The Robert Parker’s datasets had 78% possibility to increase the recall after applying CWW3.0, and the Wine Spectator’s datasets had 57% chance to increase the recall.

5.4.4 F-score Result

Table 15 shows the F-score movement of all five datasets. The “RP/WS increased” columns count the times of the F-score results after applying the CWW3.0 that increased compared to applying CWW2.0, and the “RP/WS decreased” columns count the times of the F-score have decreased.

Table 15. Statistical Record of F-score improvement after applying CWW3.0

	RP increased	RP decreased	WS increased	WS decreased
Dataset 1	6	2	7	1
Dataset 2	6	2	5	3
Dataset 3	5	1	3	4
Dataset 4	8	0	5	2
Dataset 5	8	0	5	2
Total	33/38=87%	5/38=13%	25/37=68%	12/37=32%

As shown in the “Total” row in Table 15, Robert Parker’s reviews had 87% possibility to increase the F-score after applying CWW3.0 compared to CWW2.0, and Wine Spectator’s reviews had 68% possibility to increase the accuracy after applying CWW3.0.

In sum, after applying the new iteration Computation Wine Wheel 3.0, both Robert Parker and Wine Spectator’s dataset had at least 57% possibility to increase the results. Also, the Robert Parker’s datasets had much higher possibility to increase the results of Accuracy, Precision, Recall, and F-score compared to Wine Spectator after applying CWW3.0. Each measurement metric has all detailed results in Appendix A.

CHAPTER 6 COMBINATION REVIEW OF ROBERT PARKER AND WINE SPECTATOR

The third goal of this research was to evaluate and compare the consistency of the wine reviews from two customized combination datasets from Robert Parker and Wine Spectator by using the new Computational Wine Wheel 3.0. In this chapter, the results of two types of combination datasets that formed by reviews from Robert Parker and Wine Spectator after implementing SVM classifiers and Naive Bayes classifiers will be discussed and compared.

6.1 Customized Combination Dataset

The first combination dataset type was to combine every instance from Robert Parker and Wine Spectator's reviews into one dataset. For example, the Elite Bordeaux dataset contained a total of 1026 wine reviews: 513 reviews from Robert Parker and 513 reviews from Wine Spectator. The first combination type was the "RPmixWS" dataset that contained all 1026 wine reviews from the Elite Bordeaux dataset with each review and its score. Therefore, the "RPmixWS" dataset provided more instances for the classification models to process.

The second combination dataset type, named as "RPcomWS," was combining two reviews into one review by connecting them with a space distance, and the score for each instance in "RPcomWS" was the average score of Wine Spectator and Robert Parker. As shown in Table 16, the first column contains a review from Robert Parker and its score 91; the second column contains a review from Wine Spectator and its score 95. Both reviews are describing the same wine. After applying the second combination dataset type

“RPcomWS,” the third column shows the review of combined reviews and its score 93, which is the average of scores of Robert Parker and Wine Spectator.

Table 16. Example of Combination Dataset “RPcomWS”

Robert Parker Review	Wine Spectator Review	RPcomWS Review
Deep garnet color. Aromas of warm cassis, dried plum, cloves and pencil shavings. Medium+ acidity and medium to firm, fine tannins. Long finish. Tasted August 2009.	This shows lots of mulled spice, warm tobacco leaf and well-roasted cedar accents, but isn't short on fruit, offering enticing layers of red currant, plum and blackberry confiture. The long finish is riddled with sweet smoke, black tea and iron notes. A gorgeous wine from an overlooked vintage.	Deep garnet color. Aromas of warm cassis, dried plum, cloves and pencil shavings. Medium+ acidity and medium to firm, fine tannins. Long finish. Tasted August 2009. This shows lots of mulled spice, warm tobacco leaf and well-roasted cedar accents, but isn't short on fruit, offering enticing layers of red currant, plum and blackberry confiture. The long finish is riddled with sweet smoke, black tea and iron notes. A gorgeous wine from an overlooked vintage.
RP Score : 91	WS Score : 95	RPcomWS Score : (91+95)/2 = 93

Therefore, the Elite Bordeaux dataset contained 513 reviews after “RPcomWS” combination. For this combination type, each instance contained both Robert Parker and Wine Spectator’s information, which meant more wine attributes were extracted from each instance.

6.2 Datasets and Classification

As introduced in Chapter 2, all five datasets from different data sources were used in this research: Dataset 1 of 1855 Elite Bordeaux Wine, Dataset 2 of Bordeaux Wine, Dataset 3 of Italy Wine, Dataset 4 of California Wine, Dataset 5 of All Wine Combination. Four different experiment methods used in Chapter 5 were also used in Chapter 6. Experiment 1 was using wine reviews with only “CATEGORY” attributes; Experiment 2 was using wine reviews with only “NORMALIZED” attributes; Experiment 3 was using wine reviews with “CATEGORY” + “NORMALIZED” attributes; Experiment 4 was using wine reviews with “CATEGORY” + “NORMALIZED” + “SUBCATEGORY” attributes.

6.2.1 Dataset 1 : 1855 Elite Bordeaux Wine

Dataset 1 contained 513 wines that were in the 1855 Bordeaux Wine Official Classification list made in the 21st century; combination “RPmixWS” contained 1026 instances, and combination “RPcomWS” contained 513 instances. Figure 31 displays the accuracy results of Dataset 1 after implementing the SVM classification. All results decreased after applying CWW3.0 compared to applying CWW2.0, except “RPmixWS” dataset on Experiment 2.

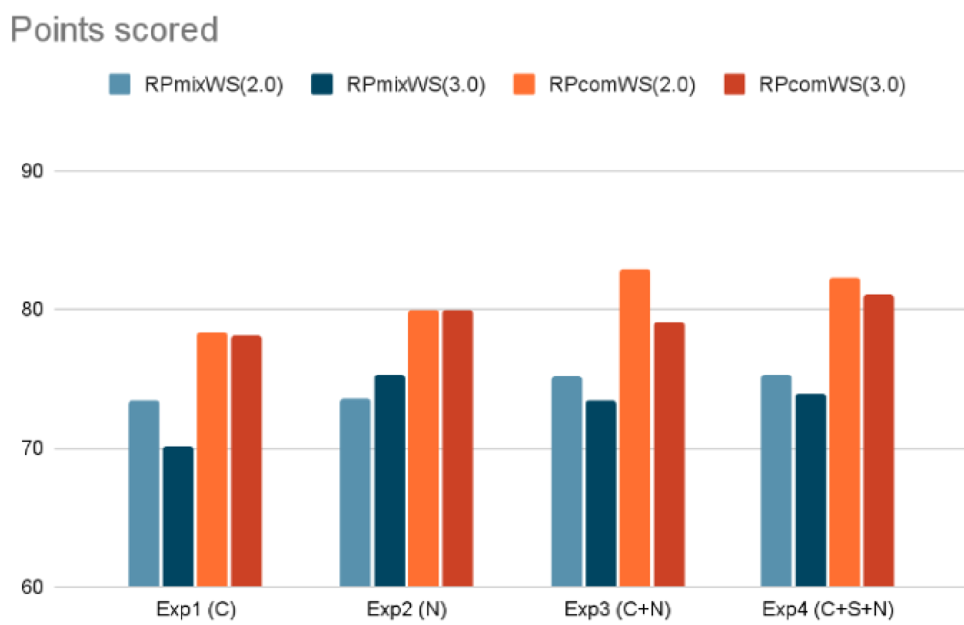

Figure 31. SVM Accuracy Results of Dataset 1

As shown in Figure 31, the datasets of “RPcomWS” achieved much higher accuracies than the datasets of “RPmixWS,” which proved that the second combination style helped gaining more information from the reviews for the classification modeling. The highest result from SVM classification on Dataset 1 was 82.85, which was the result of “RPcomWS” in Experiment 3 after applying CWW2.0.

The second method was Naive Bayes classification. Figure 32 shows the Naive Bayes results of Dataset 1. From the graph flow in Figure 32, the datasets of “RPmixWS” all decreased after applying CWW3.0 compared to applying CWW2.0, while the datasets of “RPcomWS” increased after applying CWW3.0 compared to applying CWW2.0.

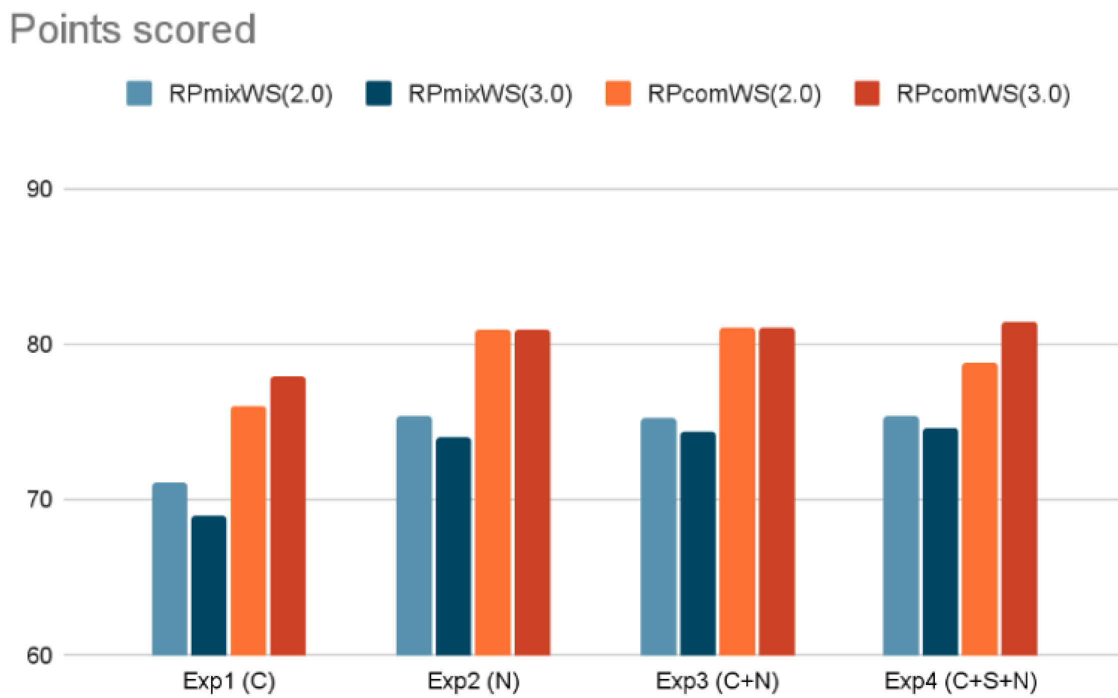

Figure 32. Naive Bayes Accuracy Result of Dataset 1

Each set of accuracy results from “RPcomWS” was much higher than the same experiment “RPmixWS” result in Figure 32. The highest accuracy result of Dataset 1 after implementing Naive Bayes classification was 81.48%, which was the result of “RPcomWS” in Experiment 4 after applying CWW3.0.

Both SVM and Naive Bayes results demonstrated that datasets of “RPcomWS” achieved higher accuracies than datasets of “RPmixWS.” Both SVM and Naive Bayes results also showed that datasets of “RPmixWS” were more likely to decrease the accuracy after applying CWW3.0 compared to applying CWW2.0. The highest accuracy result from Dataset 1 was 82.85, which was the SVM accuracy result of “RPcomWS” in Experiment 3 after applying CWW2.0.

6.2.2 Dataset 2 : Bordeaux Wine

Dataset 2 included 2341 wines that came from Bordeaux; combination “RPmixWS” contained 4682 instances, and combination “RPcomWS” contained 2341 instances. Figure 33 displays the accuracy results of Dataset 2 after implementing SVM. All datasets slightly increased after applying CWW3.0 compared to applying CWW2.0, except dataset “RPmixWS” on Experiment 2.

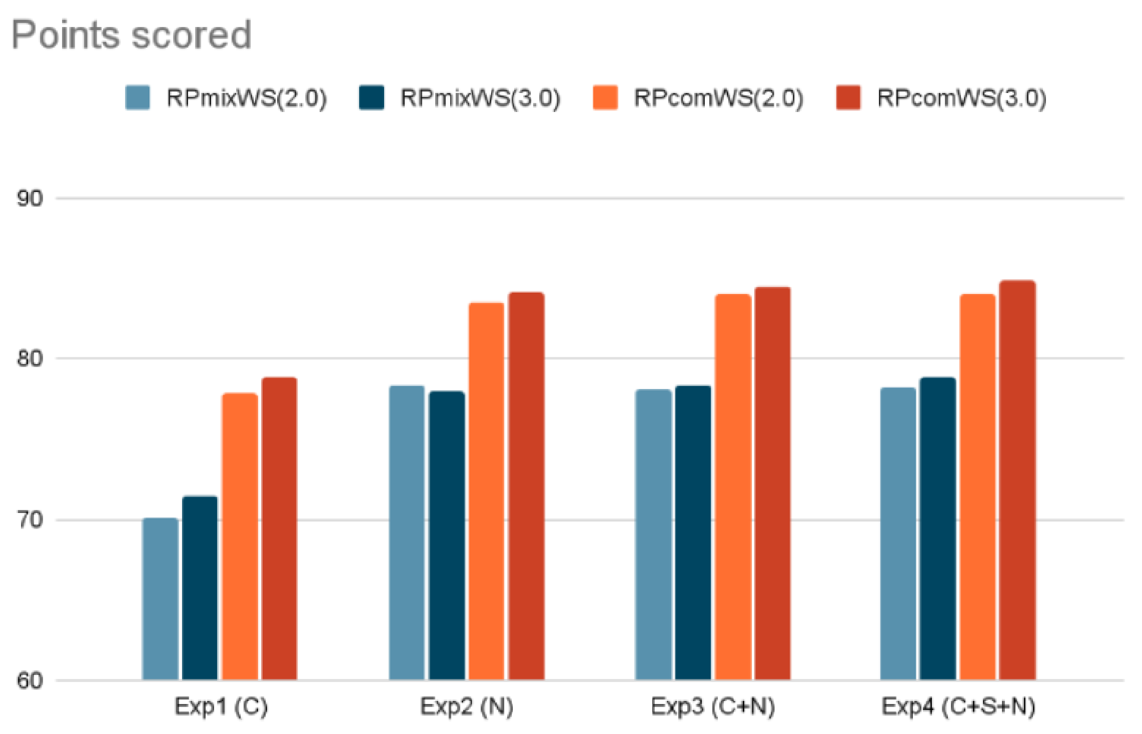

Figure 33. SVM Accuracy Results of Dataset 2

All datasets of “RPcomWS” achieved higher accuracy results than all datasets of “RPmixWS,” with an average of 6% higher accuracy for each experiment. The highest accuracy from SVM result was 84.89, which was the result of “RPcomWS” in Experiment 4 after applying CWW3.0.

Figure 34 shows the accuracy result of Dataset 2 after implementing Naive Bayes classification. The difference between datasets after applying CWW2.0 and CWW3.0 was not very obvious, while the difference between datasets of "RPcomWS" and datasets of "RPmixWS" was obvious that "RPcomWS" was achieving much higher accuracy. The highest accuracy from Naive Bayes results in Dataset 2 was 83.26, which was the result from "RPcomWS" in Experiment 2 after applying CWW2.0 and result from "RPcomWS" in Experiment 3 after applying CWW3.0.

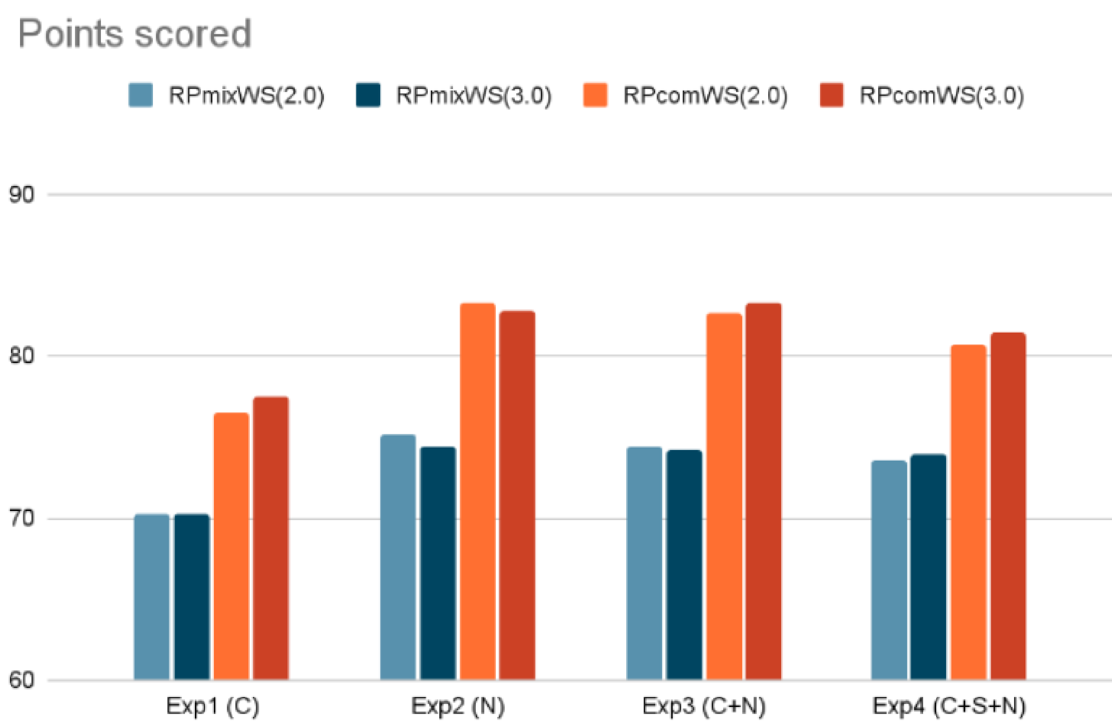

Figure 34. Naive Bayes Accuracy Result of Dataset 2

Comparing Figure 33 and Figure 34, both results show that CWW2.0 and CWW3.0 caused very slight differences in all datasets. The datasets of "RPmixWS" achieved better results on all experiments after implementing SVM classification compared to Naive Bayes. The highest accuracy from Dataset 2 was 84.89, which was the SVM result of "RPcomWS" in Experiment 4 after applying CWW3.0.

6.2.3 Dataset 3 : Italy Wine

Dataset 3 contained 3198 wines that were produced in Italy; combination "RPmixWS" contained 6396 instances, and combination "RPcomWS" contained 3198 instances. Figure 35 shows the accuracy results of Dataset 3 after implementing SVM classification. The accuracy results of "RPcomWS" achieved much higher than results of "RPmixWS." Datasets of "RPcomWS" in Experiment 2, 3 and 4 all received accuracy higher than 83.65%, while all datasets of "RPmixWS" in all four experiments had accuracies lower than 80%.

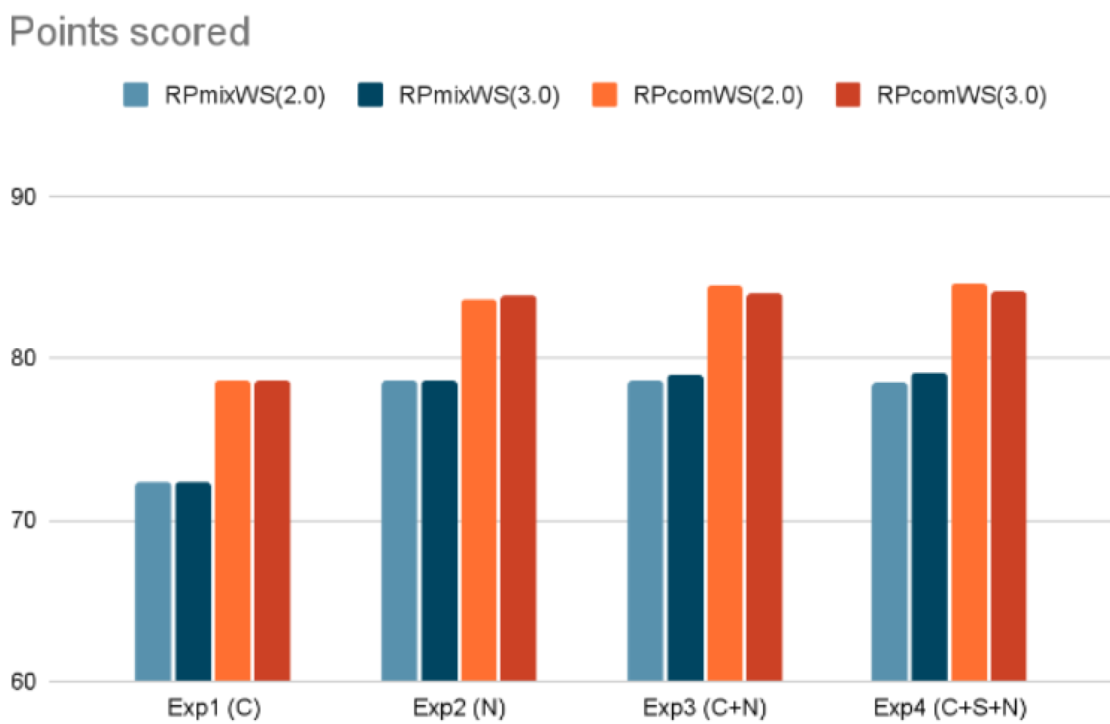

Figure 35. SVM Accuracy Results of Dataset 3

As shown in Figure 35, all datasets after applying CWW3.0 had no obvious affection compared to applying CWW2.0. The datasets of "RPmixWS" in Experiment 3 and 4 slightly increased after applying CWW3.0, while datasets of "RPcomWS" in the same experiments slightly decreased after applying CWW3.0. The highest accuracy result from SVM was 84.68%, which was the result of “RPcomWS” in Experiment 4 after applying CWW2.0.

Figure 36 displays the accuracy results of Dataset 3 after implementing Naive Bayes classification. The results of “RPcomWS” were much higher than the results of “RPmixWS” in all four experiments. The applying CWW3.0 on datasets did not cause obvious changings in Figure 36. The highest accuracy result from Naive Bayes was 84.27%, which was the result of “RPcomWS” in Experiment 2 after applying CWW3.0

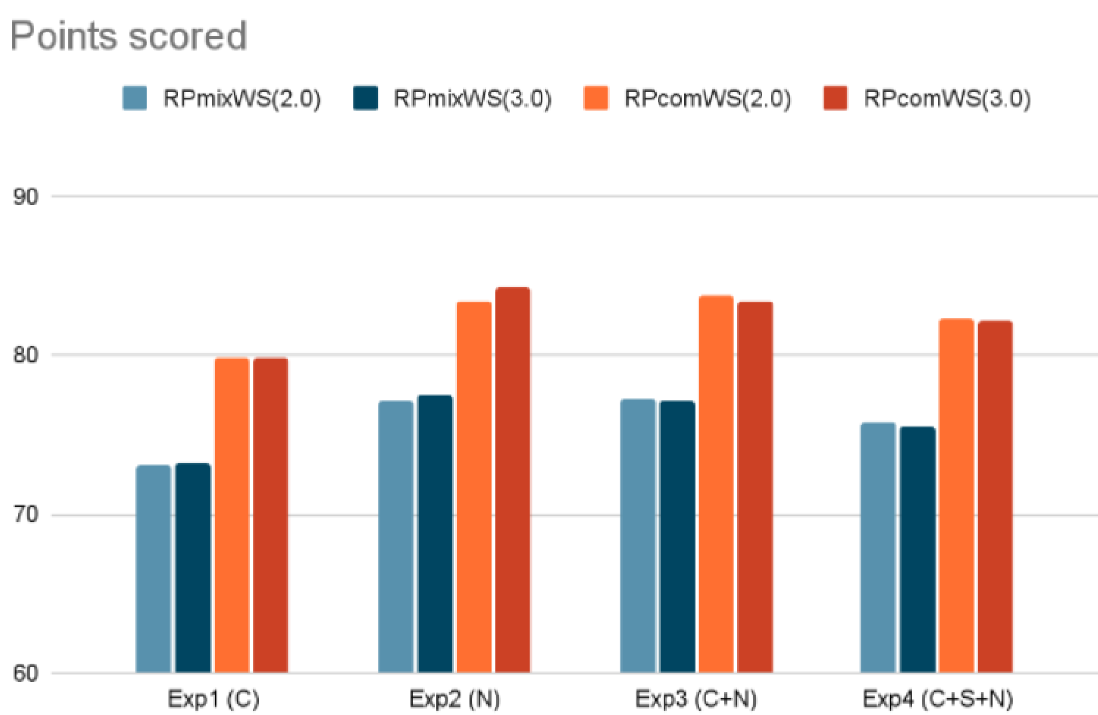

Figure 36. Naive Bayes Accuracy Result of Dataset 3

Both SVM and Naive Bayes’ highest accuracy results achieved higher than 84.27%. The results of SVM on Experiments 2, 3 and 4 were better than Naive Bayes results, but Naive Bayes achieved better accuracy than SVM in Experiment 1. The highest accuracy result of Dataset 3 was 84.68%, which was the SVM result of “RPcomWS” in Experiment 4 after applying CWW2.0.

6.2.4 Dataset 4 : California Wine

Dataset 4 contained 4180 wines that were produced in California; combination "RPmixWS" contained 8360 instances, and combination "RPcomWS" contained 4180 instances. Figure 37 shows the accuracy results for Dataset 4 after implementing SVM classification. As shown in the Figure 37, the results of all datasets increased after applying CWW3.0 compared to CWW2.0.

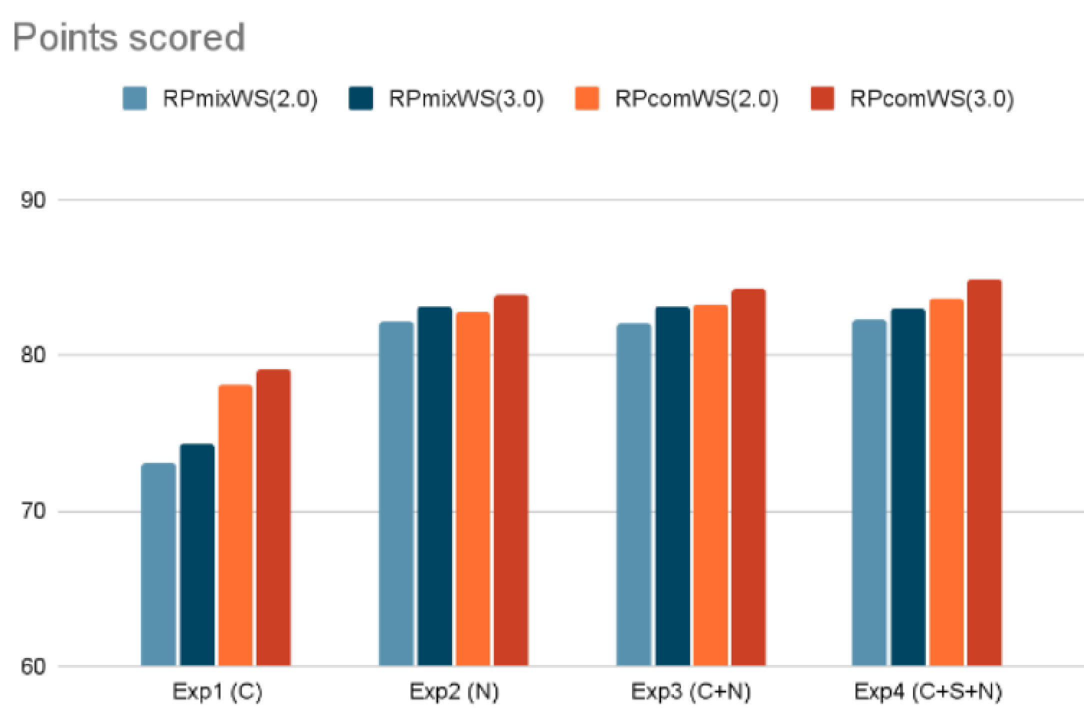

Figure 37. SVM Accuracy Results of Dataset 4

For Dataset 4, the results in Figure 37 from "RPcomWS" and "RPmixWS" were not having very big distances, both datasets achieved accuracy higher than 80% in Experiment 2, 3 and 4. The highest accuracy of Dataset 4 after implementing SVM was 84.83, which was the result of dataset "RPcomWS" in Experiment 4 after applying CWW3.0.

Figure 38 displays the accuracy results of Dataset 4 after implementing Naive Bayes classification. All dataset's results increased after applying CWW3.0 compared to applying CWW2.0. Experiments 2, 3 and 4 had very similar movement and accuracy. The highest result from Figure 38 was 81.75, which was the result of dataset "RPcomWS" in Experiment 3 after applying CWW3.0.

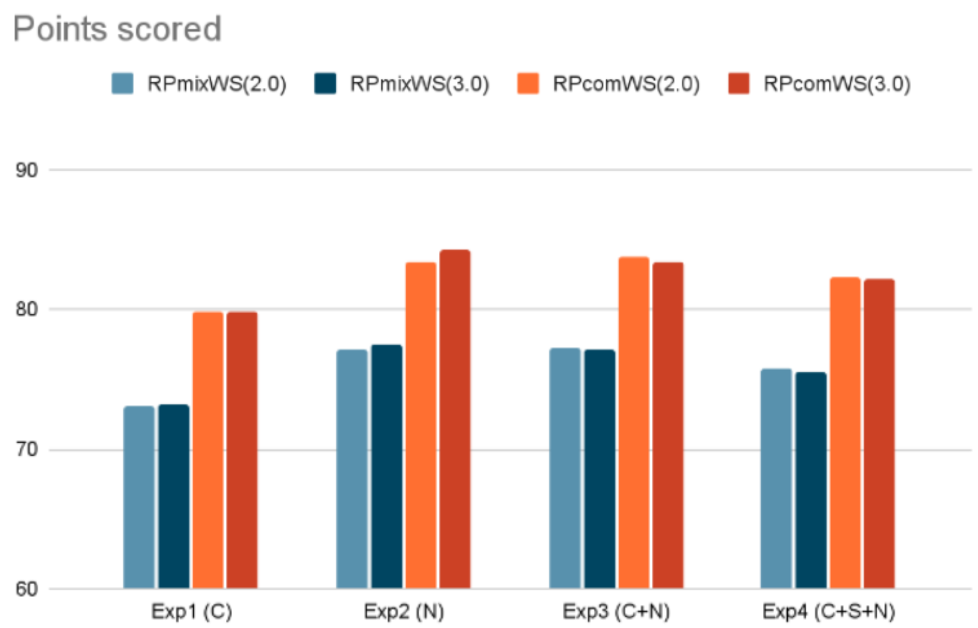

Figure 38. Naive Bayes Accuracy Result of Dataset 4

Both SVM and Naive Bayes accuracy results had similar flow movement on the figures. However, all four experiments on SVM classification achieved higher accuracy results than Naive Bayes classification. The highest accuracy result of Dataset 4 was 84.83, which was the SVM result of dataset "RPcomWS" in Experiment 4 after applying CWW3.0.

6.2.5 Dataset 5 : All Datasets Combination

Dataset 5 contained 10232 wines that were formed by the former four datasets; combination "RPmixWS" contained 20464 instances, and combination "RPcomWS" contained 10232 instances. Dataset 5 contained all instances in this research, which was the largest dataset out of all datasets. Figure 39 displays the accuracy results of Dataset 5 after implementing SVM classification. As shown in the figure, all datasets increased the accuracy after applying CWW3.0 compared to applying CWW2.0.

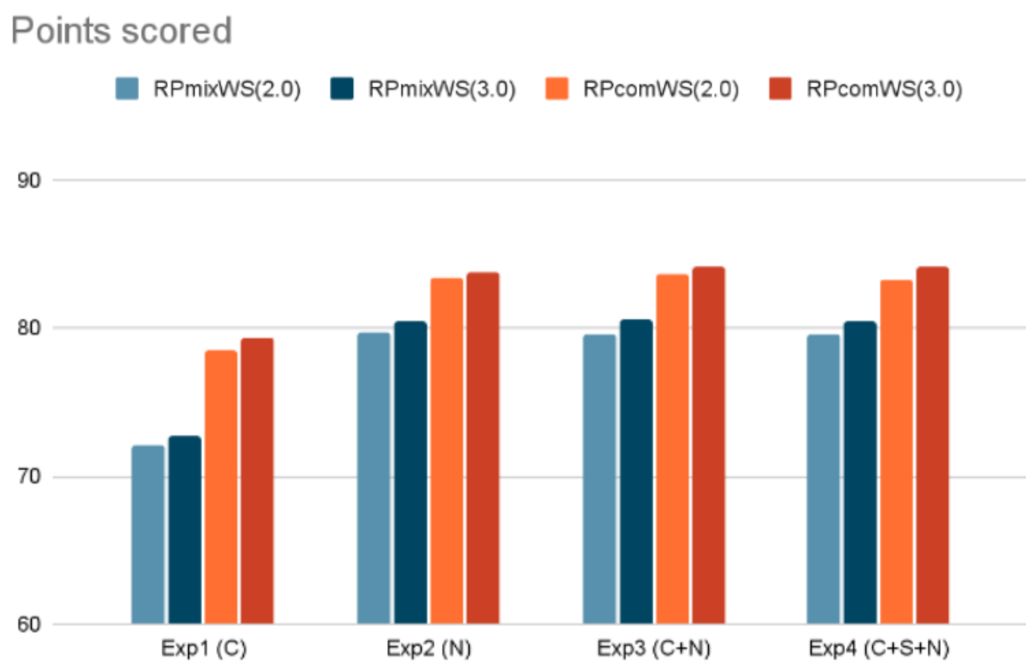

Figure 39. SVM Accuracy Results of Dataset 5

All datasets from the "RPcomWS" achieved higher accuracies than the "RPmixWS" in each experiment in Figure 39. Especially Experiments 2, 3 and 4 from "RPcomWS" datasets achieved accuracies higher than 83%, and same experiments from "RPmixWS" datasets that after applying CWW3.0 achieved accuracies higher than 80%. The highest accuracy result from SVM was 84.15%, which was the result of "RPcomWS" in Experiment 4 after applying CWW3.0.

Figure 40 shows the accuracy results of Dataset 5 after implementing Naive Bayes classification. This figure shows that datasets of "RPcomWS" achieved better results than "RPmixWS" and applying CWW3.0 helped "RPcomWS" datasets achieved better performance more than it helped "RPmixWS." Datasets of "RPcomWS" had the best results on Experiment 3, while datasets of "RPmixWS" had the best results on Experiment

2. The highest accuracy results of Dataset 5 after implementing Naive Bayes was 80.21, which was the result of “RPcomWS” in Experiment 3 after applying CWW3.0.

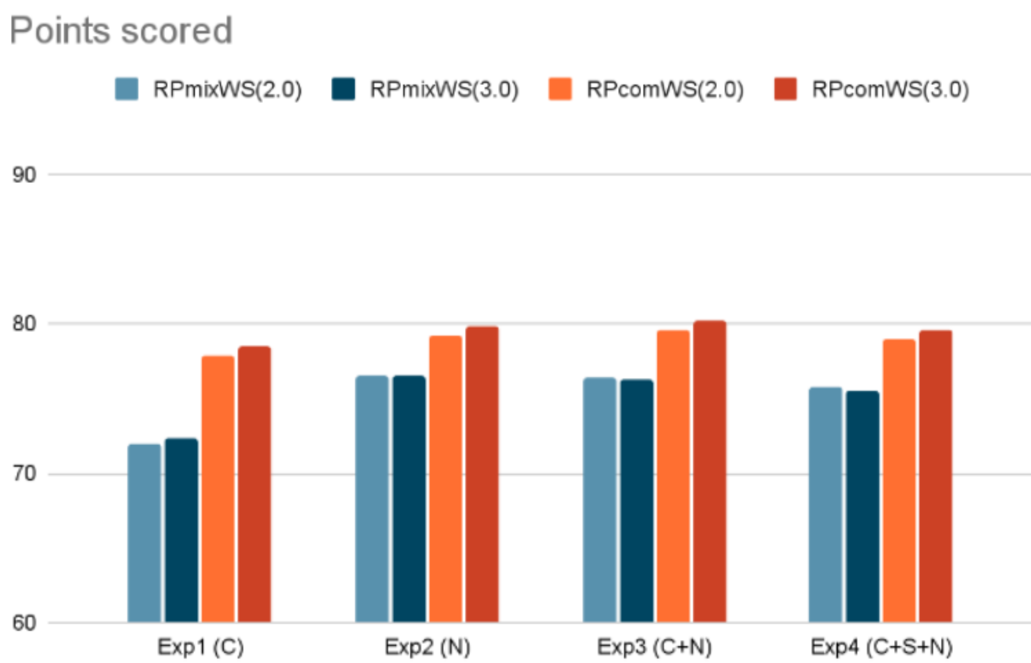

Figure 40. Naive Bayes Accuracy Result of Dataset 5

SVM has achieved better results on all four experiments than Naive Bayes, and the difference between SVM results and Naive Bayes results was as high as 4.6%. Most datasets achieved better results after applying CWW3.0 compared to CWW2.0. The highest accuracy result of Dataset 5 was 84.15%, which was the SVM result of “RPcomWS” in Experiment 4 after applying CWW3.0.

6.2.6 Dataset Results of Accuracy and Balance Statistic

All accuracy results of each dataset were demonstrated in the above. However, the results could have been affected by all kinds of reasons, one of them was the balance status of the dataset. Table 17 displays the detailed balance status of each dataset. This figure calculates how many positive and negative cases in datasets of “RPmixWS” and “RPcomWS”, and the percentage of them in each dataset.

Table 17. Positive and Negative cases count for all datasets

	RPmixWS 95+	RPmixWS 94-	RPcomWS 95+	RPcomWS 94-
Dataset 1	380 (37%)	646 (63%)	175 (34%)	338 (66%)
Dataset 2	1786 (38%)	2896 (62%)	797 (34%)	1544 (66%)
Dataset 3	1767 (28%)	4629 (72%)	684 (21%)	2514 (79%)
Dataset 4	2622 (31%)	5738 (69%)	1135 (27%)	3045 (73%)
Dataset 5	6555 (32%)	13909 (68%)	2791 (27%)	7441 (73%)

In Table 17, the most imbalanced dataset for combination "RPmixWS" was Dataset 3, which contained 28% positive (95+) cases and 72% negative (94-) cases, and the most balanced dataset for "RPmixWS" was Dataset 2, which contained 38% positive (95+) cases and 62% negative (94-) cases.

Table 18 displays the results of four different measurement metrics from the most balanced dataset (Dataset 2), the most imbalanced dataset (Dataset 3), and the dataset that had the average balance (Dataset 5). Dataset 5 is listed in Table 18 because it contained all datasets together, which represented the average results of all datasets. The accuracy of Dataset 2, 78.86%, was the highest accuracy out of all subsets in Dataset 2; the accuracy of Dataset 3, 79.13%, was the highest accuracy out of all subsets in Dataset 3; the accuracy of Dataset 5, 80.51%, was the highest accuracy out of all subsets in Dataset 5. The highest accuracy results from all three datasets were picked as an example; therefore, the differences of all three datasets on the balance problem were more obvious. Dataset 5 had higher accuracy and precision than Dataset 2 and Dataset 3. Dataset 2 had higher recall and f-score than Dataset 3 and 5. Since the Dataset 3 was the most imbalanced dataset, it received the lowest recall, precision, and f-score out of all three datasets.

Table 18. Results Comparison for Balanced and Imbalanced Dataset for "RPmixWS"

	RPmixWS 95+	RPmixWS 94-	RPcomWS 95+	RPcomWS 94-
Dataset 1	380 (37%)	646 (63%)	175 (34%)	338 (66%)
Dataset 2	1786 (38%)	2896 (62%)	797 (34%)	1544 (66%)
Dataset 3	1767 (28%)	4629 (72%)	684 (21%)	2514 (79%)
Dataset 4	2622 (31%)	5738 (69%)	1135 (27%)	3045 (73%)
Dataset 5	6555 (32%)	13909 (68%)	2791 (27%)	7441 (73%)

From Table 17, the most balanced dataset for combination "RPcomWS" was Dataset 1 and Dataset 2, which contained 34% positive (95+) cases and 66% negative (94-) cases; while the most imbalanced dataset for "RPcomWS" was Dataset 3, which contained 21% positive (95+) cases and 79% negative (94-) cases. Table 19 displays the results of four different measurement metrics from the most balanced dataset (Dataset 2), the most imbalanced dataset (Dataset 3), and the dataset that has the average balance (Dataset 5). The accuracy of Dataset 2, 84.89%, was the highest accuracy out of all subsets in Dataset 2; the accuracy of Dataset 3, 84.68%, was the highest accuracy out of all subsets in Dataset 3; the accuracy of Dataset 5, 84.15%, was the highest accuracy out of all subsets in Dataset 5.

Table 19. Results Comparison for Balanced and Imbalanced Dataset for "RPcomWS"

	RPcomWS	Accuracy	Sensitivity	Precision	F-score
Most Balanced	Dataset2	84.89%	68.13%	84.64%	75.49%
Most Imbalanced	Dataset3	84.68%	39.16 %	78.31%	52.21%
Average Balanced	Dataset5	84.15%	55.5%	80.29%	65.61%

The accuracy of Dataset 2 had the highest accuracy out of all three datasets in Table 19. Because Dataset 3 contained more imbalance data, which can lead to over-predicting on negative classes (since Dataset 3 contained more negative classes) to cause abnormal high accuracy. However, the accuracy of Dataset 2 was higher than Dataset 3, which could be that this combination type “RPcomWS” actually extracted the most amount of attributes from wine reviews to predict results more accurately. The Recall, Precision, and F-score of Dataset 3 were the lowest because it had the most imbalanced dataset. The Recall, Precision, and F-score of Dataset 5 were almost the averages of Dataset 2 and Dataset 3, which probably because of its balance percentage was also almost the average of Dataset 2 and Dataset 3.

All measurement results (Accuracy, Precision, Recall, and F-score) of all datasets are in the appendix for reference purpose.

6.2.7 Summary

The accuracy results in section 6.2 in Chapter 6 came from five datasets (1-Elite Bordeaux, 2-Bordeau, 3-Italy, 4-California, 5-All Combination) that were combined into two types of combination (“RPmixWS” and “RPcomWS”) to be applied to two versions of Computational Wine Wheel (CWW2.0 and CWW3.0), and separated into four experiment methods (1CATEGORY, 2NORMALIZED, 3CATEGORY+ NORMALIZED, 4CATEGORY+SUBCATEGORY+NORMALIZED), then implemented by using two classification (SVM and Naive Bayes).

Dataset 1’s highest accuracy was 82.85, which was the SVM accuracy result of “RPcomWS” in Experiment 3 after applying CWW2.0. Dataset 2’s highest accuracy was 84.89, which was the SVM result of “RPcomWS” in Experiment 4 after applying CWW3.0. Dataset 3’s was 84.68%, which was the SVM result of “RPcomWS” in Experiment 4 after applying CWW2.0. Dataset 4’s was 84.83, which was the SVM result of dataset

"RPcomWS" in Experiment 4 after applying CWW3.0. Dataset 5's was 84.15%, which was the SVM result of "RPcomWS" in Experiment 4 after applying CWW3.0. These highest accuracies were all of the SVM results of "RPcomWS" combination, which proved that SVM classification did better performance than Naive Bayes in this research and "RPcomWS" combination performed better than the "RPmixWS" combination. Out of all four experiment methods, Experiment 4 had the higher chance to produce the highest accuracy than others. Out of the massive amount of results, the highest accuracy result was 84.89%, which was the SVM result of "RPcomWS" in Experiment 4 after applying CWW3.0 from Dataset 2.

All measurement metrics results of "RPcomWS" and "RPmixWS" can be found in Appendix B.

6.3 Result Comparison of Chapter 5 and Chapter 6

In sum, Chapter 5 discussed the results of reviews from Robert Parker and reviews from Wine Spectator separately; Chapter 6 discussed the results of reviews from Robert Parker and Wine Spectator combinations. Therefore, for research purposes, comparing Chapter 5 and Chapter 6 is an important exploration for the domain of Wineinformatics.

Table 20 displays a comparison table for Chapter 5 and Chapter 6. The column "Highest Accuracy" displays the highest accuracy of each dataset in its chapter. The column "CWW" shows which Computational Wine Wheel was used to achieve the highest accuracy; the column "Reviewer" contains which reviewer (RP-Robert Parker, WS-Wine Spectator, RPcomWS, RPmixWS) were for the highest accuracy; the column "Classification" shows which classification method was implemented to achieved the highest accuracy; the column "Experiment" contains the experiment of the highest accuracy.

Table 20. Comparison of Chapter 5 and Chapter 6

Chapter 5	Highest Accuracy	CWW	Reviewer	Classification	Experiment
Dataset 1	78.36%	CWW3.0	WS	SVM	Exp 4 (C+S+N)
Dataset 2	79.97%	CWW3.0	WS	SVM & NB	Exp 3 (C+N)
Dataset 3	82.8%	CWW2.0	WS	NB	Exp 2 (N)
Dataset 4	86.8%	CWW3.0	WS	SVM	Exp 4 (C+S+N)
Dataset 5	83.48%	CWW3.0	WS	SVM	Exp 4 (C+S+N)

Chapter 6	Highest Accuracy	CWW	Reviewer	Classification	Experiment
Dataset 1	82.85%	CWW2.0	RPcomWS	SVM	Exp 3 (C+N)
Dataset 2	84.89%	CWW3.0	RPcomWS	SVM	Exp 4 (C+S+N)
Dataset 3	84.68%	CWW2.0	RPcomWS	SVM	Exp 4 (C+S+N)
Dataset 4	84.83%	CWW3.0	RPcomWS	SVM	Exp 4 (C+S+N)
Dataset 5	84.15%	CWW3.0	RPcomWS	SVM	Exp 4 (C+S+N)

As shown in Table 20, the highest accuracies of Chapter 6 were better than Chapter 5, except Dataset 4 was 86.8% in Chapter 5. However, "WS" in Dataset 4 in Chapter 5 contained the most imbalanced dataset (16% positive and 84% negative), which made the high accuracy doubtable. The accuracies of the "WS" in Chapter 5 were not consistent, which could be caused by the imbalance situation. The accuracies of "RPcomWS" in Chapter 6 were fairly consistent even though not all datasets were balanced. Therefore, "RPcomWS" should be the best performing out of all four types (RP, WS, RPcomWS, RPmixWS). Computational Wine Wheel 3.0 had a higher chance to achieve high accuracy than CWW2.0 as shown in Table 20. The SVM classification was performing better to achieve higher accuracy than Naive Bayes classification. From all four experiments, Experiment 4 (CATEGORY+SUBCATEGORY+NORMALIZED) had the best performance for achieving high accuracy.

6.4 Combination "RPcomWS" with Separate Class Label

As summarized in the sections above, the combination of "RPcomWS" achieved the higher accuracy in each dataset over the combination of "RPmixWS." As introduced at the beginning of Chapter 6, combination "RPcomWS" was combining review from Robert Parker and review from Wine Spectator into one review, and the class label of "RPcomWS" was taking the average score of Robert Parker and Wine Spectator. To have a deeper studying on this type of combination, another research has been done with the combination of "RPcomWS," which was using the reviews of "RPcomWS" to predict the class label of Robert Parker and Wine Spectator separately. Therefore, two additional classifier training have been developed and evaluated: for one, the reviews of "RPcomWS" trained on the WS label; for the other one, the reviews of "RPcomWS" trained on the RP label.

Table 21 displays an example of the "RPcomWS" review and the class label to predict. The first column "Robert Parker Review - RP" displays the review of Robert Parker and the class label is "RP Score 91"; the second column "Wine Spectator Review - WS" displays the review of Wine Spectator and the class label is based on the "WS Score 95"; the third column "RPcomWS Review - RP" displays the combined review of Robert Parker and Wine Spectator and the class label is based on the "RP score 91"; the fourth column "RPcomWS - WS" displays the review of combined Robert Parker and Wine Spectator and its label is based on the "WS Score 95."

Table 21. Example of "RPcomWS" Reviews and Class label

Robert Parker Review - RP	Wine Spectator Review - WS	RPcomWS Review - RP	RPcomWS Review - WS
Deep garnet color. Aromas of warm cassis, dried plum, cloves and pencil shavings. Medium+ acidity and medium to firm, fine tannins. Long finish. Tasted August 2009.	This shows lots of mulled spice, warm tobacco leaf and well-roasted cedar accents, but isn't short on fruit, offering enticing layers of red currant, plum and blackberry confiture. The long finish is riddled with sweet smoke, black tea and iron notes. A gorgeous wine from an overlooked vintage.	Deep garnet color. Aromas of warm cassis, dried plum, cloves and pencil shavings. Medium+ acidity and medium to firm, fine tannins. Long finish. Tasted August 2009. This shows lots of mulled spice, warm tobacco leaf and well-roasted cedar accents, but isn't short on fruit, offering enticing layers of red currant, plum and blackberry confiture. The long finish is riddled with sweet smoke, black tea and iron notes. A gorgeous wine from an overlooked vintage.	Deep garnet color. Aromas of warm cassis, dried plum, cloves and pencil shavings. Medium+ acidity and medium to firm, fine tannins. Long finish. Tasted August 2009. This shows lots of mulled spice, warm tobacco leaf and well-roasted cedar accents, but isn't short on fruit, offering enticing layers of red currant, plum and blackberry confiture. The long finish is riddled with sweet smoke, black tea and iron notes. A gorgeous wine from an overlooked vintage.
RP Score : 91	WS Score : 95	RP Score : 91	WS Score : 95

For this "RPcomWS" study, the "RPcomWS" combination from all five dataset (1-Elite Bordeaux, 2-Bordeau, 3-Italy, 4-California, 5-All Combination) were used for predicting the score of Robert Parker and Wine Spectator separately. Figure 41 contains all five datasets with different sources and labels. The first result in Figure 41 is named as "RP - RP", which is using the Robert Parker's review only to predict the Robert Parker's class labels; the second result is named as "RPcomWS - RP", which is using the "RPcomWS" reviews to predict the Robert Parker's class labels; the third result is named as "WS - WS", which is using Wine Spectator's review only to predict the Wine Spectator's class labels; the fourth result is named as "RPcomWS - WS", which is using the "RPcomWS" reviews to predict Wine Spectator's class labels; the last result is named as "RPcomWS - Average", which is using the "RPcomWS" reviews to predict the average score of Robert Parker and Wine Spectator. All results were the SVM results of datasets after applying Computational Wine Wheel 3.0 in Experiment 4 (CATEGORY+SUBCATEGORY+NORMALIZED).

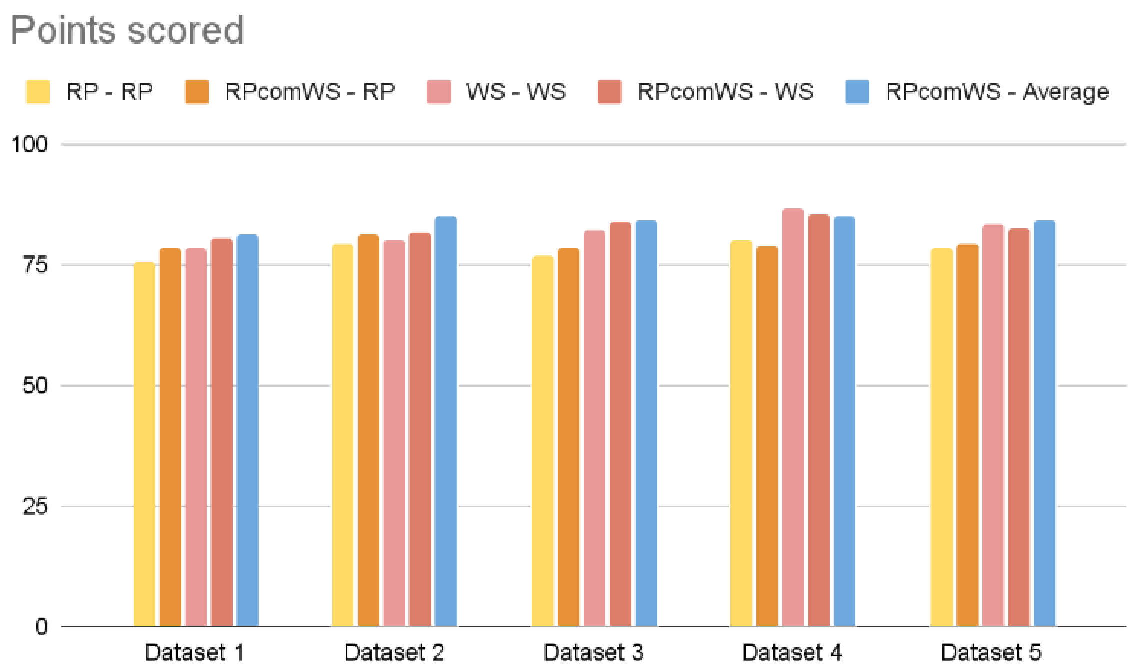

Figure 41. Accuracy of All Five Datasets with Different Class Labels and Sources

The reason for Figure 41 showing all of this information is for comparing the results of predicted class labels of Robert Parker and Wine Spectator separately by using "RPcomWS" reviews with using other reviews. For example, when trying to compare the accuracy of predicting Robert Parker's class label by using only Robert Parker's reviews or using the "RPcomWS" reviews is to compare the results of "RP - RP" and "RPcomWS - RP" in Figure 41; when trying to compare the accuracy of predicting Wine Spectator's class label by using only Wine Spectator's reviews or using the "RPcomWS" reviews is to compare the result of "WS - WS" and "RPcomWS - WS" in Figure 41. It shows that the result of "RPcomWS - RP" was better than "RP - RP" in Dataset 1, 2, 3 and 5, and the result of "RPcomWS - WS" was better than "WS - WS" in dataset 1, 2, 3, which could mean that reviews of "RPcomWS" were able to generate more wine attributes for prediction than the reviews of only Robert Parker or the reviews of only Wine Spectator. The result of "RPcomWS - Average" was higher than all other results in dataset 1, 2, 3, and 5, which meant using reviews of “RPcomWS” to predict the average score of Robert Parker and Wine Spectator performed the best.

All measurement metrics results of “RPcomWS” with different labels can be found in Appendix C.

6.5 Important Key Attributes

Two classification methods were implemented in this project: Support Vector Machines (SVM) and Naive Bayes Classification. From the results of section 6.3, the SVM classification performs better than Naive Bayes classification when it comes to achieving high accuracy, which was expected. SVMs are black-box classification, which are considered as the models that are very difficult to explain and to be understood but come with high performance. Naive Bayes classification is white-box classification, which can be understood and explained in a practical way. Therefore, using SVM classification to achieve high results and using Naive Bayes classification to study the inner logic of the data are the ideal usage for this project.

For this project, wine attributes from wine reviews were the important elements for predicting class labels. Using Naive Bayes classification helps finding the essential attributes that are having high probability to affect classification. The reviews of “RPcomWS” in Dataset 2 (Bordeaux Wine) after applying the Computation Wine Wheel 3.0 was the example dataset in this section for extracting important key attributes. Table 22’s first column displays 10 key attributes extracted from instances that were predicted as positive (95+) class and the second column displays another 10 distinguished key attributes extracted from instances that were predicted as negative (94-) class.

Table 22. Ten Key Attribute from Instance with Positive Labels and Negative Labels

Key Attributes from Positive Label (95+)	Key Attributes from Negative Label (94-)
'GREAT', 'BLACK CURRANT', 'PURPLE', 'FULL-BODIED', 'LAYER', 'RIPE', 'FRUIT', 'FLORAL', 'FIRM', 'BLACK FRUIT'	'MEDIUM-BODIED', 'WELL-BALANCED', 'VELVET', 'YOUNG', 'PURE', 'AROMA', 'DENSE', 'STYLE', 'TANNINS_DECENT', 'TOBACCO'

As shown in Table 22, the attribute "FULL-BODIED" from first column and the attribute "MEDIUM-BODIED" from second column formed a particular set of contrasts: a Bordeaux wine is more likely to be scored higher than 95 when it has the feature of "FULL-BODIED" mouthfeel, while the wine is more likely to be scored lower than 94 when it has the feature of ""MEDIUM-BODIED" mouthfeel. The attribute "FLORAL" from the first column and the attribute "AROMA" from the second column are also an interesting set of contrasts. Both "FLORAL" and "AROMA" are describing scents, but "FLORAL" is more specific to flower scent, which implies that 95+ wines are more likely to have flower scent.

Studying the key attributes of different classes is important to the domain of Wineinformatic. Key attributes provide detailed information about what formed positive instances and what formed instances, which is meaningful for wine industries in a practical way.

CHAPTER 7 CONCLUSION AND FUTURE WORK

7.1 Summary

This project focused on three major goals: 1. Enhancing the Computational Wine Wheel by analyzing and processing Robert Parker's reviews; 2. Evaluating and comparing the consistency of Wine Spectator and Robert Parker's reviews using the new Computational Wine Wheel 3.0; 3. Proposing new methods to combine different sources of reviews using the new Computational Wine Wheel 3.0 for achieving better performance.

Chapter 1 introduced the concept of the Natural language processing tool Computational Wine Wheel that is utilized for studying wine reviews under the domain of Wineinformatics. Chapter 2 described the wine review datasets used in this project and how they were generated by the Computational Wine Wheel. Chapter 3 introduced the basic concept of classification and the two classification methods used in this project.

Chapter 4 included a detailed process description of enhancing the Computational Wine Wheel. For upgrading the Computational Wine Wheel, a new resource, Robert Parker's wine reviews were analyzed and processed for extracting new attributes to add into the vocabulary of the new CWW.

Chapter 5 evaluated and compared the consistency of Wine Spectator and Robert Parker's reviews using both Computational Wine Wheel 2.0 and 3.0. The wine attribute datasets were generated by extracting the attributes from five datasets through the Natural Language Processing Tool Computational Wine Wheel 2.0 and 3.0. Out of 40 Robert Parker subsets from all five datasets, 33 subsets increased the accuracy after applying the CWW 3.0 compared to CWW 2.0; out of 40 Wine Spectator subsets from all five datasets, 27 subsets have increased the accuracy after applying the CWW 3.0 compared to CWW 2.0.

Chapter 6 demonstrated the detailed process of proposing the new method to combine two different review sources then using the new Computational Wine Wheel to achieve better results. Two customized combination type reviews, “RPcomWS” and “RPmixWS”, were evaluated and compared by using the Computational Wine Wheel 2.0 and 3.0. The combination “RPcomWS” achieved much higher accuracy compared to the combination “RPmixWS.” The SVM classification performed better than the Naive Bayes classification in this research. Applying the Computational Wine Wheel 3.0 has a higher possibility to achieve higher accuracy compared to the CWW2.0. The combination attributes of CATEGORY, SUBCATEGORY, and NORMALIZED had the highest chance out of all four experiments to achieve the highest accuracy. Since the combination “RPcomWS” is performing well, the combined reviews from “RPcomWS” were used for predicting the class labels of Robert Parker and Wine Spectator separately. The results of the prediction were better than only using one reviewer’s reviews.

These three goals have been accomplished in this project. The Computational Wine Wheel has been upgraded, and its upgraded version CWW 3.0 has performed better than the older version CWW 2.0 on both Robert Parker and Wine Spectator’s datasets. The new method for combining two reviews into one has achieved ideal results that are better than all other review formats.

7.2 Future Works

This research also provides new research directions in Wineinformatics. About the three major goals of this research, each one of them has a lot of improvement potential.

First, newer versions of the Computational Wine Wheel may be developed by including wine reviews from other wine experts or wine publications. For example, James Suckling as a famous American wine critic, and Wine Enthusiast as an influential wine magazine, all can be found on the website wine.com. As a Natural Language Processing tool, the CWW should be expanded to include more variety of attributes, especially attributes such as word-groups or two words that are not adjacent to each other but the meaning is related. Upgrading the CWW has an essential effect on wine review studying.

Second, collecting more reviews from different sources can be used to evaluate and compare by using the upgraded Computational Wine Wheel. Wines in this project that have been studied are mostly the classic (95+) and outstanding (90-95) level wines, including different levels of wines will help discover more strength of the Computational Wine Wheel.

Last but not least, from the result of “RPcomWS” in this research, it shows that more attributes extracted from one wine can help predict the class label better. Therefore, the combination type “RPcomWS” can be an important future study. The combination method of “RPcomWS” can be expanded by collecting wine reviews from more wine sources, and then combine all wine reviews that are describing the same wine but from different sources into one review. The instance from “RPcomWS” can then be even stronger since it contains all attributes from different sources. The class label of “RPcomWS” is also an essential part of this future study. The class label can be the average of all sources or having a set of class labels that predict all sources labels separately, or the majority label out of all sources.

REFERENCES

[1] Smith, E.A. (2001), "The role of tacit and explicit knowledge in the workplace", Journal of Knowledge Management, Vol. 5 No. 4, pp. 311-321. https://doi.org/10.1108/13673270110411733

[2] Aggarwal C.C. (2011) An Introduction to Social Network Data Analytics. In: Aggarwal C. (eds) Social Network Data Analytics. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-8462-3_1

[3] A. Gyrard, C. Bonnet and K. Boudaoud, "Enrich machine-to-machine data with semantic web technologies for cross-domain applications," 2014 IEEE World Forum on Internet of Things (WF-IoT), 2014, pp. 559-564, doi: 10.1109/WF-IoT.2014.6803229.

[4] Lawrence Harris, "A transaction data study of weekly and intradaily patterns in stock returns," Journal of Financial Economics, Volume 16, Issue 1, 1986, pp. 99-117, ISSN 0304-405X, https://doi.org/10.1016/0304-405X(86)90044-9.

[5] Cunningham P., Cord M., Delany S.J. (2008) Supervised Learning. In: Cord M., Cunningham P. (eds) Machine Learning Techniques for Multimedia. Cognitive Technologies. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-75171-7_2

[6] Ghahramani, Zoubin, "Unsupervised learning." Summer school on machine learning. Springer, Berlin, Heidelberg, 2003.

[7] Sadaaki Miyamoto. 2012. An Overview of Hierarchical and Non-hierarchical Algorithms of Clustering for Semi-supervised Classification. Modeling Decisions for Artificial Intelligence, 1-10.

[8] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.

[9] N. V. Ganapathi Raju, K. Prasanna Lakshmi, K. G. Praharshitha and C. Likhitha, "Prediction of chronic kidney disease (CKD) using Data Science," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 642-647, doi: 10.1109/ICCS45141.2019.9065309.

[10] N. K. Gyamfi and J. Abdulai, "Bank Fraud Detection Using Support Vector Machine," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 37-41, doi: 10.1109/IEMCON.2018.8614994.

[11] Linoff, Gordon S., and Michael JA Berry. Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2011.

[12] A. Agnihotri, P. Saraf and K. R. Bapnad, "A Convolutional Neural Network Approach Towards Self-Driving Cars," 2019 IEEE 16th India Council International Conference (INDICON), 2019, pp. 1-4, doi: 10.1109/INDICON47234.2019.9030307.

[13] James,Simpson, Wine: a short history (2014)

[14] Pau Roca, "State Of The Vitiviniculture World In 2020", International Organization of Vine and Wine, 2021

[15] Busby, J., Journal of a Tour through Some of the Vineyards of Spain and France (Sydney, 1833).

[16] wine spectator

[17] Dong, Zeqing, Travis Atkison, and Bernard Chen. 2021. "Wineinformatics: Using the Full Power of the Computational Wine Wheel to Understand 21st Century Bordeaux Wines from the Reviews" Beverages 7, no. 1: 3.

[18] Kwabla, William, Falla Coulibaly, Yerkebulan Zhenis, and Bernard Chen. 2021. "Wineinformatics: Can Wine Reviews in Bordeaux Reveal Wine Aging Capability?" Fermentation 7, no. 4: 236.

[19] Chen, Bernard, Christopher Rhodes, Aaron Crawford, and Lorri Hambuchen. "Wineinformatics: applying data mining on wine sensory reviews processed by the computational wine wheel." In 2014 IEEE International Conference on Data Mining Workshop, pp. 142-149. IEEE, 2014.

[20] Chen, B.; Rhodes, C.; Yu, A.; Velchev, V. The Computational Wine Wheel 2.0 and the TriMax Triclustering in Wineinformatics. In Industrial Conference on Data Mining; Springer: Cham, Switzerland, 2016; pp. 223–238.

[21] About Us | Wine.com

[22] Alasadi, Suad A., and Wesam S. Bhaya. “Review of data preprocessing techniques in data mining.” *Journal of Engineering and Applied Sciences* 12.16 (2017): 4102-4107.

[23] Zaki, Mohammed J., Wagner Meira Jr, and Wagner Meira. Data mining and analysis: *fundamental concepts and algorithms*. Cambridge University Press, 2014.

[24] Yabing, Jiao. "Research of an improved apriori algorithm in data mining association rules." International Journal of Computer and Communication Engineering 2.1 (2013): 25.

[25] Berkhin, Pavel. "A survey of clustering data mining techniques." Grouping multidimensional data. Springer, Berlin, Heidelberg, 2006. 25-71.

[26] Gorade, Sudhir M., Ankit Deo, and Preetesh Purohit. "A study of some data mining classification techniques." *International Research Journal of Engineering and Technology* 4.4 (2017): 3112-3115.

[27] Refaeilzadeh, Payam, Lei Tang, and Huan Liu. "Cross-validation." Encyclopedia of database systems 5 (2009): 532-538.

[28] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," 2016 IEEE 6th International Conference on Advanced Computing (IACC), 2016, pp. 78-83, doi: 10.1109/IACC.2016.25.

[29] Loyola-Gonzalez, Octavio. "Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view." IEEE Access 7 (2019): 154096-154113.

[30] Cherkassky, Vladimir, and Sauptik Dhar. "Interpretation of black-box predictive models." *Measures of Complexity*. Springer, Cham, 2015. 267-286.

[31] Abraham, Ajith. "Artificial neural networks." Handbook of measuring system design (2005).

[32] Chen, Yen-Liang, and Lucas Tzu-Hsuan Hung. "Using decision trees to summarize associative classification rules." *Expert Systems with Applications* 36.2 (2009): 2338-2351.

[33] Weisberg, Sanford. Applied linear regression. Vol. 528. John Wiley & Sons, 2005.

[34] Ben-Gal, Irad. "Bayesian networks." Encyclopedia of statistics in quality and reliability 1 (2008).

[35] Webb, Geoffrey I., Eamonn Keogh, and Risto Miikkulainen. "Naïve Bayes." Encyclopedia of machine learning 15 (2010): 713-714.

[36] Pintelas, E.; Livieris, I.E.; Pintelas, P. A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability. Algorithms 2020, 13, 17. https://doi.org/10.3390/a13010017

[37] Howar. Falk, Bengt Jonsson, and Frits Vaandrager. “Combining black-box and white-box techniques for learning register automata.” Computing and Software Science. Springer, Cham, 2019. 563-588.

[38] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. “A training algorithm for optimal margin classifiers.” *Proceedings of the fifth annual workshop on Computational learning theory*. 1992.

[39] Karimi, N., et al. "On a generalized Gaussian radial basis function: Analysis and applications." Engineering Analysis with Boundary Elements 112 (2020): 46-57.

[40] T. Joachims, Making large-Scale SVM Learning Practical. Advances in Kernel Methods – Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

[41] Raizada, Rajeev DS, and Yune-Sang Lee. "Smoothness without smoothing: why Gaussian naive Bayes is not naive for multi-subject searchlight studies." PloS one 8.7 (2013): e69566.

[42] Agostini, H. & Guichard, M.-F. 2007. Robert Parker: anatomie d'un mythe. Paris: Scali. Amoraritei, L. 2002. La métaphore en oeologie. Metaphorik.de 3:1–12. Available at http://www.metaphorik.de/03/amoraritei.htm. Date of access 3 March 2011.

[43] Hommerberg, Charlotte. Persuasiveness in the discourse of wine: The rhetoric of Robert Parker. Diss. Linnaeus University Press, 2011.

[44] Langewiesche, W. 2000. The million-dollar nose. The Atlantic Monthly Dec 2000:42–70. Available at http://www.erobertparker.com/info/milliondollarnose.asp. Date of access 1 March 2011.

[45] Goodhue, Rachael, et al. "California wine industry evolving to compete in 21st century." California Agriculture 62.1 (2008).

[46] Chen, B.; Velchev, V.; Palmer, J.; Atkison, T. Wineinformatics: A Quantitative Analysis of Wine Reviewers. Fermentation 2018, 4, 82.

[47] Bodington, J.C. Evaluating wine-tasting results and randomness with a mixture of rank preference models. J. Wine Econ. 2015, 10, 31–46.

q

APPENDIX A. MEASUREMENT METRICS RESULTS OF ROBERT PARKER AND WINE SPECTATOR

Abbreviation Notes

SVM -- Support Vector Machine Classification
NB -- Naive Bayes Classification
CATEGORY -- CATEGORY attributes
NORMALIZED -- NORMALIZED attributes
C+N -- CATEGORY + NORMALIZED attributes combination
C+S+N -- CATEGORY + SUBCATEGORY + NORMALIZED attributes combination
2.0 -- Computational Wine Wheel 2.0
3.0 -- Computational Wine Wheel 3.0

ACCURACY

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
Robert Parker 2.0 CATEGORY	73.1%	71.35 %	72.66 %	71.17 %	68.48 %	69.48 %	64.45 %	63.18 %	67.76 %	66.33 %
Robert Parker 2.0 NORMALIZED	73.29%	74.85 %	79.5 %	76.8 %	76.21 %	77.52 %	77.61 %	76.48 %	77.71 %	75.1 %
Robert Parker 2.0 C+N	75.63%	76.02 %	79.41 %	76.12 %	77.3 %	76.83 %	78.37 %	76.67 %	77.8 %	74.02 %
Robert Parker 2.0 C+S+N	75.63%	74.85 %	79.84 %	75.61 %	76.89 %	76.89 %	77.8 %	75.72 %	77.49 %	73.22 %
Robert Parker 3.0 CATEGORY	74.07%	71.93 %	74.97 %	72.62 %	68.48 %	69.79 %	65.53 %	64.95 %	69.28 %	77.32 %
Robert Parker 3.0 NORMALIZED	74.46%	75.83 %	78.9 %	78.9 %	76.67 %	77.83 %	78.9 %	76.96 %	78.76 %	81.07 %
Robert Parker 3.0 C+N	75.63%	75.44 %	79.97 %	78.34 %	76.67 %	77.74 %	79.35 %	77.89 %	78.44 %	81.17 %
Robert Parker 3.0 C+S+N	74.27%	75.24 %	79.32 %	77.79 %	76.89 %	75.89 %	79.88 %	77.06 %	78.48 %	79.2 %
Wine Spectator 2.0 CATEGORY	71.35%	71.93 %	69.8 %	69.54 %	76.27 %	76.67 %	83.8 %	84.55 %	76.25 %	67.8 %
Wine Spectator 2.0 NORMALIZED	75.44%	74.07 %	79.45 %	79.92 %	81.84 %	82.8 %	86.43 %	85.33 %	83.09 %	75.88 %
Wine Spectator 2.0 C+N	74.46%	75.83 %	79.75 %	79.45 %	81.95 %	82.11 %	86.7 %	84.71 %	83.1 %	75.62 %
Wine Spectator 2.0 C+S+N	76.02%	74.46 %	79.41 %	78.13 %	82.49 %	79.74 %	86.79 %	83.18 %	83.33 %	74.78 %
Wine Spectator 3.0 CATEGORY	71.35%	71.15 %	71.38 %	71.04 %	76.27 %	76.67 %	83.8 %	73.22 %	72.08 %	77.85 %
Wine Spectator 3.0 NORMALIZED	77%	74.03 %	78.68 %	79.45 %	81.96 %	82.71 %	86.51 %	78.01 %	79.76 %	81.2 %
Wine Spectator 3.0 C+N	78.17%	77.52 %	79.92 %	79.97 %	81.8 %	81.99 %	86.79 %	78.29 %	79.58 %	81.24 %
Wine Spectator 3.0 C+S+N	78.36%	77.13 %	79.97 %	78.6 %	82.06 %	80.43 %	86.8 %	77.66 %	79.6 %	79.88 %

PRECISION

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
Robert Parker 2.0 CATEGORY	79.21%	66.67%	74.27%	68.84%	0	52.91%	65.87%	62.44%	68.0%	60.35%
Robert Parker 2.0 NORMALIZED	72.06%	77.42%	80.24%	75.06%	69.57%	69.61%	78.89%	74.2%	77.63 %	70.85 %
Robert Parker 2.0 C+N	76.47 %	76.43%	79.55%	74.53%	72.4%	66.54%	79.91%	75.55 %	77.79%	69.37 %
Robert Parker 2.0 C+S+N	78.12%	73.61%	79.57%	72.44 %	71.64%	61.96%	78.63%	74.29 %	77.25%	67.84%
Robert Parker 3.0 CATEGORY	77.59%	68.03%	76.67%	70.77%	0	53.75%	67.26%	64.76 %	69.72%	62.63%
Robert Parker 3.0 NORMALIZED	74.26%	79.67 %	79.41%	78.13 %	71.73 %	70.34%	80.7 %	75.26%	78.85%	72.17%
Robert Parker 3.0 C+N	76.09%	76.69 %	80.3%	77.53 %	70.75%	68.32%	81.38%	77.08%	78.49%	71.72%
Robert Parker 3.0 C+S+N	74.81%	74.65 %	80.2 %	76.12 %	71.83%	63.3 %	81.57%	76.0 %	78.0%	70.1%
Wine Spectator 2.0 CATEGORY	70.0%	65.6%	66.25%	58.45%	0	51.27%	0	56.43%	0	54.62 %
Wine Spectator 2.0 NORMALIZED	70.21%	77.17%	75.57%	74.89%	75.4%	69.9 %	76.56%	55.78 %	77.99%	60.92%
Wine Spectator 2.0 C+N	69.7%	78.1 %	75.96%	72.6 %	75.42%	64.06 %	77.81%	52.84%	77.59%	60.2 %
Wine Spectator 2.0 C+S+N	73.08%	73.21%	75.42%	68.62 %	76.27%	56.78%	79.17 %	48.41%	77.47 %	55.48%
Wine Spectator 3.0 CATEGORY	67.65 %	62.07%	67.54%	60.28%	0	51.33%	0	55.4%	0	56.79%
Wine Spectator 3.0 NORMALIZED	72.11 %	78.82 %	75.0 %	75.54%	75.47%	70.52 %	77.84%	58.16%	77.88%	61.71 %
Wine Spectator 3.0 C+N	75.54 %	83.84 %	76.58%	74.36%	72.88 %	64.73 %	78.67%	54.57%	77.64%	60.68%
Wine Spectator 3.0 C+S+N	75.71%	77%.78	77.34%	70.67%	74.88%	58.56 %	77.31%	49.24%	77.48%	57.01%

RECALL

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
Robert Parker 2.0 CATEGORY	40.61%	50.76%	52.62%	56.21%	0	28.87 %	49.0%	52.39%	37.84 %	48.05%
Robert Parker 2.0 NORMALIZED	49.75%	48.73%	67.49 %	66.36%	43.66%	50.89 %	70.85%	75.84 %	62.84%	64.95%
Robert Parker 2.0 C+N	52.79 %	54.31%	68.21%	64.82%	45.33%	53.27%	71.57 %	73.73%	62.88%	63.68 %
Robert Parker 2.0 C+S+N	50.76%	53.81%	69.44%	66.87 %	44.03%	55.75 %	71.83%	73.11%	62.64 %	63.83 %
Robert Parker 3.0 CATEGORY	45.69 %	50.76%	57.33%	58.36 %	0	29.86%	50.49 %	54.14%	42.16 %	49.89%
Robert Parker 3.0 NORMALIZED	51.27%	49.75%	66.67 %	68.51%	42.86%	51.29%	71.93%	75.22%	64.65%	65.38%
Robert Parker 3.0 C+N	53.3 %	51.78%	68.82 %	67.59%	44.14%	54.76%	72.24%	74.7%	64.12%	65.24 %
Robert Parker 3.0 C+S+N	49.75 %	53.81%	66.87%	68.0%	43.96%	55.95%	73.37%	74.09%	64.97%	65.31 %
Wine Spectator 2.0 CATEGORY	34.43%	44.81%	26.52 %	41.8%	0	34.52 %	0	20.09 %	0	26.5%
Wine Spectator 2.0 NORMALIZED	54.1 %	38.85%	60.05%	63.26%	35.05%	48.35%	23.18%	45.64 %	40.25%	56.58%
Wine Spectator 2.0 C+N	50.27%	44.81%	60.91 %	65.35%	35.81 %	56.13%	24.81%	52.29 %	40.74 %	61.07%
Wine Spectator 2.0 C+S+N	51.91%	44.81%	60.17%	67.94%	38.22 %	61.26%	25.11 %	58.35%	42.02%	62.92 %
Wine Spectator 3.0 CATEGORY	37.7%	49.18 %	33.91%	48.09%	0	33.07 %	0	17.43 %	0	28.23 %
Wine Spectator 3.0 NORMALIZED	57.92%	36.61%	57.7%	60.17%	35.3%	46.64%	23.48%	44.76%	40.78%	54.86 %
Wine Spectator 3.0 C+N	57.38%	45.36 %	60.67%	64.36%	37.29 %	52.96%	25.86%	51.11%	42.96 %	59.63 %
Wine Spectator 3.0 C+S+N	57.92 %	49.73%	59.68%	65.35%	37.3%	59.95%	26.16%	57.46 %	42.96 %	62.1%

F-SCORE

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
Robert Parker 2.0 CATEGORY	53.69%	57.64%	61.6%	61.89%	0	37.36%	56.2%	56.98%	48.61%	53.5%
Robert Parker 2.0 NORMALIZED	58.86%	59.81%	73.31%	70.44%	53.65%	58.8%	74.65%	75.01%	69.44%	67.77%
Robert Parker 2.0 C+N	62.46%	63.5%	73.44%	69.34%	55.75%	59.17%	75.51%	74.63%	69.53%	66.41%
Robert Parker 2.0 C+S+N	61.54%	62.17%	74.16%	69.54%	54.54%	58.69%	75.08%	73.7%	69.17%	65.78%
Robert Parker 3.0 CATEGORY	57.51%	58.14%	65.6%	63.97%	0	38.39%	57.68%	58.98%	52.53%	55.54%
Robert Parker 3.0 NORMALIZED	60.66%	61.25%	72.48%	73.0%	53.66%	59.32%	76.06%	75.24%	71.05%	68.61%
Robert Parker 3.0 C+N	62.69%	61.82%	74.12%	72.22%	54.36%	60.79%	76.54%	75.87%	70.58%	68.33%
Robert Parker 3.0 C+S+N	59.76%	62.54%	72.93%	71.83%	54.54%	59.4%	77.25%	75.03%	70.88%	67.62%
Wine Spectator 2.0 CATEGORY	46.16%	53.25%	37.88%	48.74%	0	41.26%	0	29.63%	0.0	35.69%
Wine Spectator 2.0 NORMALIZED	61.11%	51.64%	66.92%	68.59%	47.85%	57.16%	35.59%	50.2%	53.05%	58.67%
Wine Spectator 2.0 C+N	58.41%	56.95%	67.61%	68.78%	48.56%	59.83%	37.62%	52.56%	53.34%	60.63%
Wine Spectator 2.0 C+S+N	60.7%	55.59%	66.94%	68.28%	50.92%	58.93%	38.13%	52.92%	54.47%	58.97%
Wine Spectator 3.0 CATEGORY	48.42%	54.88%	45.15%	53.5%	0	40.22%	0	26.52%	0	37.71 %
Wine Spectator 3.0 NORMALIZED	64.24%	50.0%	65.22%	66.98%	48.1%	56.15%	36.085	50.59%	53.53%	58.08%
Wine Spectator 3.0 C+N	65.22%	58.87%	67.7%	69.0%	49.34%	58.26%	38.92%	52.78%	55.28%	60.15%
Wine Spectator 3.0 C+S+N	65.63%	60.67%	67.37%	67.91%	49.8%	59.25%	39.09%	53.03%	55.24%	59.44%

APPENDIX B. MEASUREMENT METRICS RESULTS OF COMBINATIONS OF ROBERT PARKER AND WINE SPECTATOR

ACCURACY

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS 2.0 CATEGORY	78.36%	76.02 %	77.92 %	76.55 %	78.61 %	79.89 %	78.09 %	77.89 %	78.49 %	77.92 %
RPcomWS 2.0 NORMALIZED	79.92%	80.9 %	83.51 %	83.26 %	83.65 %	83.36 %	82.75 %	80.26 %	83.41 %	79.26 %
RPcomWS 2.0 C+N	82.85%	81.09 %	83.98 %	82.7 %	84.49 %	83.83 %	83.23 %	80.43 %	83.59 %	79.65 %
RPcomWS 2.0 C+S+N	82.26%	78.75 %	83.98 %	80.73 %	84.68 %	82.24 %	83.66 %	79.0 %	83.27 %	78.92 %
RPcomWS 3.0 CATEGORY	78.17%	77.97 %	78.81 %	77.49 %	78.61 %	79.8 %	79.14 %	78.4 %	79.36 %	78.46 %
RPcomWS 3.0 NORMALIZED	79.92%	80.9 %	84.11 %	82.79 %	83.9 %	84.27 %	83.9 %	80.86 %	83.75 %	79.89 %
RPcomWS 3.0 C+N	79.14 %	81.09 %	84.54 %	83.26 %	83.99 %	83.43 %	84.21 %	81.75 %	84.09 %	80.21 %
RPcomWS 3.0 C+S+N	81.09%	81.48 %	84.89 %	81.42 %	84.11 %	82.11 %	84.83 %	80.57 %	84.15 %	79.55 %
RPmixWS 2.0 CATEGORY	73.49%	71.15 %	70.16 %	70.25 %	72.37 %	73.06 %	73.07 %	73.22 %	72.08 %	71.92 %
RPmixWS 2.0 NORMALIZED	73.59%	75.44 %	78.32 %	75.2 %	78.55 %	77.13 %	82.12 %	78.01 %	79.76 %	76.46 %
RPmixWS 2.0 C+N	75.15%	75.24 %	78.09 %	74.37 %	78.56 %	77.22 %	82.02 %	78.29 %	79.58 %	76.44 %
RPmixWS 2.0 C+S+N	75.34%	75.34 %	78.3 %	73.52 %	78.52 %	75.83 %	82.32 %	77.66 %	79.6 %	75.84 %
RPmixWS 3.0 CATEGORY	70.08%	68.91 %	71.49 %	70.23 %	72.37 %	73.22 %	74.32 %	73.53 %	72.74 %	72.31 %
RPmixWS 3.0 NORMALIZED	75.34%	73.98 %	78.02 %	74.48 %	78.64 %	77.5 %	83.13 %	78.58 %	80.4 %	76.55 %
RPmixWS 3.0 C+N	73.49%	74.37 %	78.36 %	74.13 %	78.92 %	77.11 %	83.19 %	78.67 %	80.56 %	76.29 %
RPmixWS 3.0 C+S+N	73.88%	74.66 %	78.86 %	73.92 %	79.13 %	75.52 %	82.97 %	78.12 %	80.51 %	75.54 %

PRECISION

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS 2.0 CATEGORY	78.57%	69.4%	75.28%	68.73%	0	56.9%	75.0%	64.24 %	76.05%	64.07%
RPcomWS 2.0 NORMALIZED	78.57%	81.3%	82.02%	77.55%	75.82%	65.02 %	77.01 %	61.98%	78.36 %	62.15 %
RPcomWS 2.0 C+N	82.71%	80.47%	82.85%	76.06%	78.19 %	64.42 %	78.37 %	62.57%	78.83 %	62.72 %
RPcomWS 2.0 C+S+N	83.33%	73.91%	82.1%	72.29%	78.31%	58.43 %	79.08%	59.67%	77.73%	60.85 %
RPcomWS 3.0 CATEGORY	80.58%	74.6 %	76.05%	70.71 %	0	56.83%	75.08 %	64.01 %	78.49%	65.81%
RPcomWS 3.0 NORMALIZED	80.0%	82.91%	82.75 %	77.21%	77.16 %	67.99%	78.27 %	62.97 %	79.45%	63.29 %
RPcomWS 3.0 C+N	76.56%	82.5 %	84.01%	77.55%	77.48%	63.8%	79.04%	64.86%	80.3 %	63.71 %
RPcomWS 3.0 C+S+N	80.95%	79.85%	84.64%	74.2%	77.25%	58.56%	80.16 %	62.13%	80.29 %	62.09 %
RPmixWS 2.0 CATEGORY	76.21%	65.11%	72.14%	65.08%	0	52.54%	72.0 %	61.1 %	74.49%	59.79 %
RPmixWS 2.0 NORMALIZED	68.35 %	73.36 %	76.63%	70.82 %	72.23%	60.67 %	78.75%	64.38%	77.34%	64.63 %
RPmixWS 2.0 C+N	71.33%	72.18%	75.92 %	69.64%	71.8%	59.95 %	78.64%	64.96%	76.8 %	64.37 %
RPmixWS 2.0 C+S+N	71.82 %	70.16%	76.29 %	67.37 %	72.42%	56.0%	79.37 %	63.52%	76.34 %	62.71%
RPmixWS 3.0 CATEGORY	70.39%	60.34%	73.12%	64.58 %	0	53.03%	72.79 %	60.77%	72.19%	59.98%
RPmixWS 3.0 NORMALIZED	71.38 %	71.0%	76.52 %	70.17 %	72.64%	61.53 %	81.45 %	65.49%	78.8%	65.0%
RPmixWS 3.0 C+N	68.37 %	70.82%	76.92%	69.21%	72.95%	59.62 %	81.64%	65.66%	78.76%	64.24%
RPmixWS 3.0 C+S+N	69.18 %	69.11%	77.69%	68.28%	73.16%	55.44%	81.24%	64.5%	78.27%	62.33%

RECALL

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS 2.0 CATEGORY	50.29%	53.14%	52.46%	57.09%	0	24.71%	29.25%	41.94 %	31.03%	43.39 %
RPcomWS 2.0 NORMALIZED	56.57%	57.14%	65.99%	71.52 %	34.68%	48.1 %	52.07%	70.66%	54.21%	61.3 %
RPcomWS 2.0 C+N	62.86%	58.86 %	67.0 %	71.77 %	38.16%	54.53%	52.95%	69.52%	54.53 %	62.63%
RPcomWS 2.0 C+S+N	60.0 %	58.29%	67.63%	70.39%	39.16 %	58.77 %	54.19%	69.87%	54.21%	63.7 %
RPcomWS 3.0 CATEGORY	47.43%	53.71%	55.34%	57.84 %	0	23.1%	34.71 %	46.7%	33.5%	43.78 %
RPcomWS 3.0 NORMALIZED	54.86%	55.43%	67.38%	70.14%	35.37%	50.0%	56.3 %	71.63 %	54.53%	62.52 %
RPcomWS 3.0 C+N	56.0%	56.57%	67.64%	71.52%	35.66 %	52.05%	57.0 %	71.54%	55.21%	63.78%
RPcomWS 3.0 C+S+N	58.29%	61.14 %	68.13%	69.64%	36.53%	55.99%	58.59%	72.86%	55.5 %	64.31%
RPmixWS 2.0 CATEGORY	41.32%	47.63%	35.5 %	47.48 %	0	25.75%	23.15%	40.2 %	19.51%	37.64 %
RPmixWS 2.0 NORMALIZED	53.42 %	52.89%	62.32%	59.52 %	36.34%	48.9%	58.89%	66.93%	52.07%	58.57 %
RPmixWS 2.0 C+N	55.0%	53.95%	62.32%	58.17%	36.67%	52.86%	58.62 %	66.82%	51.96%	59.22 %
RPmixWS 2.0 C+S+N	55.0 %	58.16%	62.66%	59.29%	35.94%	58.35%	58.96%	67.54 %	52.65 %	60.58 %
RPmixWS 3.0 CATEGORY	33.16%	46.84%	40.15%	48.6 %	0	26.71%	28.99%	44.01%	24.23%	40.76 %
RPmixWS 3.0 NORMALIZED	55.79 %	50.26%	61.2%	57.56%	36.61%	49.52%	59.84%	67.01%	53.1%	58.05 %
RPmixWS 3.0 C+N	52.89%	52.37 %	61.87%	58.01 %	37.69 %	53.14%	59.88%	67.09%	53.81%	58.64 %
RPmixWS 3.0 C+S+N	53.16%	57.11%	62.54 %	59.07%	38.71 %	58.01%	59.46%	67.28%	54.19%	59.71%

F-SCORE

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS 2.0 CATEGORY	61.33%	60.19%	61.83%	62.37%	0	34.46%	42.09%	50.75%	44.04%	51.74%
RPcomWS 2.0 NORMALIZED	65.78%	67.11%	73.14%	74.41%	47.59%	55.29%	62.13%	66.04%	64.07%	61.72 %
RPcomWS 2.0 C+N	71.43%	67.99%	74.09%	73.85%	51.29%	59.06%	63.2%	65.86%	64.45%	62.67%
RPcomWS 2.0 C+S+N	69.77%	65.18%	74.17%	71.33%	52.21%	58.6%	64.31%	64.37%	63.86%	62.24%
RPcomWS 3.0 CATEGORY	59.71%	62.45%	64.06%	63.63%	0	32.85%	47.47%	54.0%	46.95%	52.58%
RPcomWS 3.0 NORMALIZED	65.09%	66.44%	74.28%	73.51%	48.51%	57.62%	65.49%	67.02%	64.67%	62.91%
RPcomWS 3.0 C+N	64.69%	67.12%	74.94%	74.41%	48.84%	57.33%	66.23%	68.04%	65.41%	63.74 %
RPcomWS 3.0 C+S+N	67.78%	69.25%	75.49%	71.85%	49.6%	57.25%	67.7%	67.07%	65.61%	63.18 %
RPmixWS 2.0 CATEGORY	53.59%	55.01%	47.58%	54.9%	0	34.56 %	35.04%	48.49%	30.92%	46.19%
RPmixWS 2.0 NORMALIZED	59.97%	61.47%	68.74%	64.68%	48.35%	54.15 %	67.39%	65.63%	62.22%	61.45%
RPmixWS 2.0 C+N	62.11%	61.75%	68.45%	63.39%	48.48%	56.18 %	67.17%	65.88%	62.22%	61.69 %
RPmixWS 2.0 C+S+N	62.29%	63.6%	68.81%	63.07%	48.0%	57.15 %	67.66%	65.47%	62.31%	61.63%
RPmixWS 3.0 CATEGORY	45.08%	52.74%	51.84%	55.46%	0	35.53 %	41.47%	51.05%	36.27%	48.54%
RPmixWS 3.0 NORMALIZED	62.63%	58.865	68.01%	63.24%	48.64%	54.88 %	68.99%	66.24%	63.44%	61.33%
RPmixWS 3.0 C+N	59.64%	60.21%	68.58%	63.12%	49.7%	56.19 %	69.09%	66.37%	63.93%	61.31%
RPmixWS 3.0 C+S+N	60.12%	62.54%	69.3%	63.34%	50.57 %	56.69 %	68.66%	65.86%	64.04%	60.99%

APPENDIX C. MEASUREMENT METRICS RESULTS OF COMBINATIONS OF ROBERT PARKER WITH DIFFERENT LABEL

ACCURACY

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS - RP 2.0 CATEGORY	74.27 %	72.9 %	73.94 %	72.83 %	68.82 %	71.45 %	67.25 %	66.53 %	70.28 %	69.45 %
RPcomWS - RP 2.0 NORMALIZED	79.92 %	77.78 %	80.39 %	80.01 %	76.52 %	77.92 %	77.15 %	73.88 %	77.83 %	74.28 %
RPcomWS - RP 2.0 C+N	79.92 %	79.92 %	80.69 %	78.47 %	77.02 %	77.02 %	77.51 %	74.74 %	77.69 %	74.6 %
RPcomWS - RP 2.0 C+S+N	78.96 %	79.92 %	81.12 %	78.64 %	77.61 %	75.27 %	76.84%	73.83 %	78.01 %	73.95 %
RPcomWS - RP 3.0 CATEGORY	76.01 %	74.85 %	75.65 %	73.22 %	71.04 %	71.51 %	68.47 %	67.51 %	71.3 %	70.56 %
RPcomWS - RP 3.0 NORMALIZED	75.43 %	80.31 %	80.35 %	80.73 %	78.27 %	78.36 %	78.37 %	75.38 %	79.27 %	75.15 %
RPcomWS -RP 3.0 C+N	75.45 %	78.56 %	80.31 %	80.31 %	78.11 %	78.58 %	78.97 %	75.84 %	79.49 %	75.67 %
RPcomWS -RP 3.0 C+S+N	78.38 %	78.95 %	81.42 %	80.31 %	78.36 %	77.61 %	78.64 %	75.55 %	79.27 %	74.89 %
RPcomWS - WS 2.0 CATEGORY	73.69 %	71.93 %	75.82 %	74.11 %	78.61 %	76.61 %	83.8 %	83.42 %	76.25 %	78.19 %
RPcomWS - WS 2.0 NORMALIZED	76.21 %	77.58 %	80.14 %	78.64 %	84.09 %	81.27 %	85.12 %	82.18 %	82.12 %	79.17 %
RPcomWS - WS 2.0 C+N	76.6 %	77.39 %	78.98 %	78.64 %	83.71 %	80.36 %	85.19 %	81.41 %	82.06 %	78.63 %
RPcomWS - WS 2.0 C+S+N	77.99 %	77.97 %	80.05 %	77.53 %	83.96 %	79.02 %	85.21%	79.86 %	82.21 %	77.66 %
RPcomWS - WS 3.0 CATEGORY	73.5 %	75.24 %	76.5 %	74.37 %	78.61 %	77.83 %	83.8 %	84.11 %	76.25 %	78.26 %
RPcomWS - WS 3.0 NORMALIZED	76.02 %	77.97 %	82.36 %	79.54 %	83.77 %	81.46 %	85.38 %	81.91 %	82.55 %	79.28 %
RPcomWS - WS 3.0 C+N	77.2 %	78.36 %	81.67 %	78.73 %	83.9 %	81.08 %	85.19 %	81.94 %	82.49 %	78.59 %
RPcomWS - WS 3.0 C+S+N	80.32 %	77.39 %	81.89 %	77.87 %	83.8 %	79.8 %	85.43 %	80.6 %	82.48 %	78.05 %

PRECISION

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS - RP 2.0 CATEGORY	76.11 %	67.9 %	74.61 %	70.7 %	51.43 %	58.68 %	69.2 %	66.35 %	71.09 %	65.87 %
RPcomWS - RP 2.0 NORMALIZED	80.17 %	79.02 %	80.41 %	78.58 %	71.42 %	67.36 %	77.13 %	70.73 %	77.17 %	69.11 %
RPcomWS - RP 2.0 C+N	80.77 %	83.1 %	80.36 %	76.61 %	73.54 %	64.54 %	77.32 %	72.69 %	76.5 %	69.8 %
RPcomWS - RP 2.0 C+S+N	80.81 %	80.92 %	81.29 %	76.07 %	75.39 %	60.67 %	76.85 %	71.48 %	77.12 %	68.63 %
RPcomWS - RP 3.0 CATEGORY	81.77 %	72.67 %	75.87 %	71.12 %	74.76 %	58.43 %	70.13 %	67.82 %	71.71 %	67.41 %
RPcomWS - RP 3.0 NORMALIZED	74.75 %	85.29 %	81.21 %	79.98 %	75.23 %	68.72 %	79.04 %	72.61 %	79.05 %	70.33 %
RPcomWS -RP 3.0 C+N	76.5 %	81.29 %	81.12 %	78.88 %	74.08 %	68.04 %	79.77 %	74.06 %	79.55 %	71.45 %
RPcomWS -RP 3.0 C+S+N	79.03 %	81.12 %	82.37 %	78.24 %	74.8 %	64.84 %	79.07 %	73.37 %	79.37 %	70.03 %
RPcomWS - WS 2.0 CATEGORY	74.13 %	63.83 %	74.0 %	65.79 %	0.0 %	51.83 %	0.0 %	46.52 %	0.0 %	57.65 %
RPcomWS - WS 2.0 NORMALIZED	71.03 %	74.64 %	78.07 %	71.87 %	75.79 %	63.99 %	75.92 %	45.32 %	75.25 %	56.73 %
RPcomWS - WS 2.0 C+N	73.17 %	73.1 %	76.0 %	70.99 %	73.79 %	60.41 %	75.11 %	43.65 %	74.99 %	55.07 %
RPcomWS - WS 2.0 C+S+N	74.37 %	72.15 %	77.25 %	68.48 %	75.52 %	55.85 %	79.05 %	40.96 %	74.89 %	52.77 %
RPcomWS - WS 3.0 CATEGORY	70.54 %	70.29 %	74.63 %	66.51 %	0.0 %	59.26 %	0.0 %	52.81 %	0.0 %	58.19 %
RPcomWS - WS 3.0 NORMALIZED	73.01 %	75.74 %	81.51 %	72.49 %	74.98 %	65.15 %	74.59 %	44.8 %	76.28 %	56.82 %
RPcomWS - WS 3.0 C+N	74.66 %	75.0 %	81.05 %	71.01 %	77.58 %	62.88 %	68.49 %	45.13 %	75.65 %	54.95 %
RPcomWS - WS 3.0 C+S+N	79.18 %	71.07 %	79.9 %	69.05 %	76.86 %	58.01 %	75.22 %	42.78 %	75.31 %	53.57 %

RECALL

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS - RP 2.0 CATEGORY	48.66 %	55.84 %	56.72 %	59.38 %	2.89 %	31.85 %	53.47 %	56.97 %	44.29 %	50.25 %
RPcomWS - RP 2.0 NORMALIZED	63.96 %	57.36 %	70.05 %	71.49 %	42.65 %	58.13 %	72.39 %	74.81 %	63.95 %	65.45 %
RPcomWS - RP 2.0 C+N	62.86 %	59.9 %	71.08 %	69.54 %	42.46 %	60.12 %	73.11 %	73.21 %	64.48 %	65.21 %
RPcomWS - RP 2.0 C+S+N	59.96 %	62.44 %	70.97 %	71.08 %	43.15 %	61.21 %	71.88 %	72.8 %	64.65 %	65.19 %
RPcomWS - RP 3.0 CATEGORY	48.71 %	55.33 %	61.03 %	60.1 %	13.48 %	33.33 %	56.14 %	57.43 %	47.64 %	52.24 %
RPcomWS - RP 3.0 NORMALIZED	54.81 %	58.88 %	68.72 %	71.69 %	46.32 %	57.54 %	72.9 %	75.63 %	66.08 %	66.33 %
RPcomWS -RP 3.0 C+N	53.73 %	57.36 %	68.82 %	72.0 %	47.32 %	60.42 %	73.52 %	73.98 %	66.13 %	66.06 %
RPcomWS -RP 3.0 C+S+N	58.94 %	58.88 %	70.46 %	73.03 %	47.31 %	63.29 %	73.52 %	74.5 %	65.67 %	65.94 %
RPcomWS - WS 2.0 CATEGORY	43.03 %	49.18 %	47.09 %	52.65 %	0.0 %	20.55 %	0.0 %	15.81 %	0.0 %	30.7 %
RPcomWS - WS 2.0 NORMALIZED	56.24 %	56.28 %	59.55 %	63.01 %	38.01 %	48.22 %	11.96 %	48.6 %	36.83 %	51.89 %
RPcomWS - WS 2.0 C+N	53.97 %	57.92 %	57.83 %	64.86 %	37.16 %	50.07 %	12.7 %	50.81 %	36.79 %	54.32 %
RPcomWS - WS 2.0 C+S+N	59.32 %	62.3 %	60.04 %	65.1 %	37.17 %	55.34 %	11.66 %	55.24 %	37.82 %	56.5 %
RPcomWS - WS 3.0 CATEGORY	44.87 %	53.01 %	48.95 %	52.4 %	0.0 %	21.08 %	0.0 %	18.02 %	0.0 %	30.12 %
RPcomWS - WS 3.0 NORMALIZED	53.12 %	56.28 %	63.76 %	65.97 %	36.09 %	47.04 %	14.77 %	50.22 %	38.6 %	53.13 %
RPcomWS - WS 3.0 C+N	55.9 %	59.02 %	61.9 %	65.23 %	35.22 %	49.54 %	15.51 %	53.32 %	38.64 %	54.57 %
RPcomWS - WS 3.0 C+S+N	60.73 %	61.75 %	63.99 %	65.47 %	34.93 %	53.89 %	15.36 %	58.64 %	39.01 %	56.87 %

F-SCORE

	Dataset 1 SVM	Dataset 1 NB	Dataset 2 SVM	Dataset 2 NB	Dataset 3 SVM	Dataset 3 NB	Dataset 4 SVM	Dataset 4 NB	Dataset 5 SVM	Dataset 5 NB
RPcomWS - RP 2.0 CATEGORY	59.16 %	61.28 %	64.42 %	64.55 %	4.93 %	41.29 %	60.3 %	61.3 %	54.55 %	57.01 %
RPcomWS - RP 2.0 NORMALIZED	71.05 %	66.47 %	74.84 %	74.87 %	53.35 %	62.41 %	74.66 %	72.71 %	69.92 %	67.23 %
RPcomWS - RP 2.0 C+N	70.56 %	69.62 %	75.41 %	72.9 %	53.59 %	62.25 %	75.15 %	72.95 %	69.97 %	67.43 %
RPcomWS - RP 2.0 C+S+N	68.61 %	70.49 %	75.75 %	73.49 %	54.85 %	60.94 %	74.28 %	72.13 %	70.32 %	66.87 %
RPcomWS - RP 3.0 CATEGORY	60.82 %	62.82 %	67.61 %	65.15 %	22.18 %	42.45 %	62.36 %	62.19 %	57.23 %	58.86 %
RPcomWS - RP 3.0 NORMALIZED	62.98 %	69.67 %	74.41 %	75.61 %	57.32 %	62.63 %	75.82 %	74.09 %	71.99 %	68.27 %
RPcomWS -RP 3.0 C+N	62.53 %	67.26 %	74.42 %	75.28 %	57.71 %	64.0 %	76.49 %	74.02 %	72.21 %	68.65 %
RPcomWS -RP 3.0 C+S+N	67.36 %	68.24 %	75.94 %	75.54 %	57.89 %	64.06 %	76.19 %	73.93 %	71.86 %	67.92 %
RPcomWS - WS 2.0 CATEGORY	53.21 %	55.56 %	57.31 %	58.49 %	0	29.43 %	0.0 %	23.59 %	0.0 %	40.06 %
RPcomWS - WS 2.0 NORMALIZED	62.58 %	64.17 %	67.41 %	67.15 %	50.57 %	55.0 %	20.63 %	46.9 %	49.45 %	54.2 %
RPcomWS - WS 2.0 C+N	61.93 %	64.63 %	65.53 %	67.78 %	49.24 %	54.76 %	21.67 %	46.96 %	49.33 %	54.69 %
RPcomWS - WS 2.0 C+S+N	65.23 %	66.86 %	67.55 %	66.75 %	49.68 %	55.59 %	20.24 %	47.04 %	50.23 %	54.57 %
RPcomWS - WS 3.0 CATEGORY	54.15 %	60.44 %	59.09 %	58.62 %	0	31.1 %	0.0 %	26.87 %	0.0 %	39.7 %
RPcomWS - WS 3.0 NORMALIZED	61.17 %	64.58 %	71.45 %	69.08 %	48.72 %	54.63 %	24.57 %	47.35 %	51.24 %	54.91 %
RPcomWS - WS 3.0 C+N	63.23 %	66.06 %	70.09 %	67.99 %	48.32 %	55.42 %	25.11 %	48.88 %	51.12 %	54.76 %
RPcomWS - WS 3.0 C+S+N	68.11 %	66.08 %	70.98 %	67.22 %	48.01 %	55.87 %	25.41 %	49.47 %	51.4 %	55.17 %