WINEINFORMATICS: A QUANTITATIVE ANALYSIS OF WINE REVIEWERS

by

Valentin Dimitrov Velchev

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

Master of Science

in

Applied Computing

Conway, Arkansas
December 2017

TO THE OFFICE OF GRADUATE STUDIES:

The members of the Committee approve the thesis of

Valentin Velchev presented on November 28, 2017.

Dr. Bernard Chen, Committee Chairperson

Dr. Yu Sun

Dr. M. Emre Celebi

PERMISSION

Title	Wineinformatics: A Quantitative Analysis of Wine Reviewers
Department	Computer Science
Degree	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of Central Arkansas, I agree that the Library of this University shall make it freely available for inspections. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis/dissertation work, or, in the professor's absence, by the Chair of the Department or the Dean of the Graduate School. It is understood that due recognition shall be given to me and to the University of Central Arkansas in any scholarly use which may be made of any material in my thesis/dissertation.

Valentin Velchev

December 1, 2017

ABSTRACT

With ever increasing new technology, we have become able to measure far more information than our predecessors. However, organizing and making sense of this data becomes increasingly difficult as it grows. Data science has arisen as a new field to accomplish this task by using automation made possible by computers. This thesis seeks to apply data science to the study of wine for domain knowledge; this study is known as Wineinformatics. Collecting data about different types of wines has led to sorting characteristics about these wines, including both chemical traits such as acidity or alcohol content as well as sensory traits such as finish, body, and texture. These traits have led to the development of a computational wine wheel, a reference for various wine attributes. We have improved this computational wine wheel to include the past 10 years' Top 100 Wines from Wine Spectator, and we have created a program to process wine reviews using a given wheel. In addition, we have applied association rules based classification to region-specific wines, and we have used Naive Bayes and Support Vector Machine classification to make a quantitative evaluation of wine reviewers. Finally, we have discovered some additional attributes not included in the Top 100 Wines. In our study, we have found very promising results with specific reviewers that have significantly higher consistency than others, two with accuracy as high as 91%, and we can use this type of information in the future to grade wine reviewers based on their consistency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

As humanity has made technological strides, it has become able to manipulate exponentially increasing amounts of data, and as such, processing power and storage have grown at a phenomenal rate. A computer that would have taken up an entire room a generation ago fits neatly in our pockets today, and this trend shows no sign of deceleration. This escalation of computational strength has allowed people to use ever increasing amounts of data to measure and ascertain the realities of our world far more accurately than our ancestors. Data simply means any information that a computer can process. However, as this amount of data grows, managing it becomes ever more cumbersome. This challenge creates a demand for a way to sort through these vast data.

Data science is the new field that seeks to answer the question of this growing data. It takes advantage of the automation possible by computers, and it has clear applications for a diverse array of organizations. For instance, businesses require it to manage information about their clients and sales. Governments need it in order to protect their citizens' lives and property. Healthcare professionals store countless information about their thousands of patients in order to save lives and treat illnesses. Academics use it to inquire about and seek out answers to the social and economic issues of our day. With all of these applications, data science finds a very strong role in the background of the lives of billions of people [1, 2].

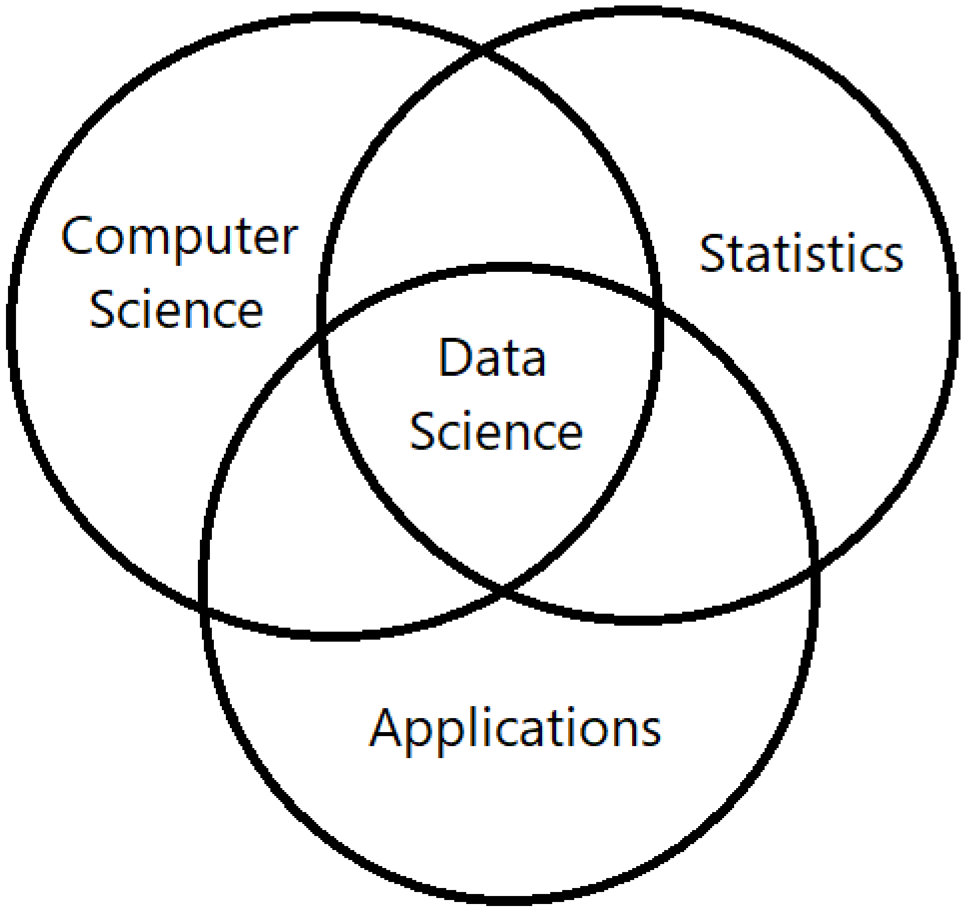


Figure 1.1: Data Science Venn Diagram

According to the data science initiative at New York University, data science in research consists of combining the fields of computer science and mathematical statistics in order to give a context for these aforementioned applications, as shown in Figure 1.1 [2]. One must have a knowledge of the computational processes involved in the data before having the ability to make sense of it. This requires working effectively with computers in order to store the data efficiently in data structures and make calculations on the data with computer algorithms. These calculations then result in statistics that we can use to make sense of the world [2].

The process through which one stores and retrieves the relevant information for use in data science is known as data warehousing. Data warehousing is described as "a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions" [3]. This requires a flexible central database, or storage space, to store this massive amount of information so that one can maximize access and analysis on the dataset, and countless businesses have made this storage of data a reality. One particular example of data warehousing involves storing a large amount of wine data from wine reviews into text files for later use and analysis, which this thesis specifically uses. Once the data is stored, we can use it to discover patterns.

Obtaining these calculations used in data science to acquire an improved understanding of the world involves data mining. A common definition of data mining includes "the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both" [4]. This involves extracting the useful information from an enormous amount of data and summarizing it in a succinct way that gives it a new meaning. For instance, a store may use data mining to determine what types of items customers commonly purchase together in order to organize these items in closer proximity so that the customers will buy more of these items. If a store finds that, for example, people buy eggs and bacon together frequently enough, the store will organize these items in such a way as to incentivize buying one with the other. It may even include a deal or coupon for buying them together.

What makes this knowledge obtained so significant? Data mining seeks to find these significant patterns within the stored data through the aforementioned analysis for various reasons, one of which includes giving producers a more adequate understanding of the interests of consumers. Specifically, retailers believe that this use of data mining will net them enough of a profit to pay off the investment. For example, credit card companies may give advertisements related to what products cardholders bought in their previous month. Rental stores for games or movies may recommend products related to previous purchases as well. The advertising revenue would offset the costs of the stock put into collecting, storing, and analyzing information about consumers. Another field that may apply these data science techniques involves the study of wine, which is the basis of this thesis.

Wine has existed for many thousands of years, but modern science has allowed us to analyze industrial wine production on a new level in our lifetimes. While people in the past saw wine as an extravagance, right now it is widely available and consumed by hundreds of millions of people. The world in total has consistently consumed between 24 and 25 billion liters of wine over the period of 2011 to 2014, which on average means more than three liters for every individual on the planet [5]. Therefore, there exists an interest in investigating the patterns of what makes people enjoy this popular product.

This investigation into quality has led to significant wine research that collects various data about the different types of wines, their regions, and other factors that would alter the wine's quality. Characteristics that certain wines exhibit include both chemical characteristics such as acidity or alcohol concentration as well as sensory attributes such as the finish, body, or texture, as shown in Figure 1.2 [6-8]. Comparing them to those of other wines in order to determine what qualities best correlate with an increase in the wine's rating would allow researchers to learn the most important types of attributes exhibited.

Kosta Browne Pinot Noir Sonoma Coast 2009

Chemical analysis

PRIMARY FERMENTATION DETAILS
HARVEST DATES
Gap's Crown: September 19, 21, 23
Terra de Promissio: September 19, 23
Walala: September 23
COLD-SOAK TIME 5 days average
FERMENTATION TIME 14 days average
FERMENTATION TEMP 86° F peak
BARREL PROGRAM
PERCENTAGE OF NEW FRENCH OAK 45%
BARREL AGING 16 months
FINISHED WINE DETAILS ALCOHOL 14.5%
PH 3.63
TITRATABLE ACIDITY 5.3 g/L
BOTTLING DATES
January 26-28, 2011

Sensory analysis

Ripe and deeply flavored, concentrated and well-structured, this full-bodied red offers a complex mix of black cherry, wild berry and raspberry fruit that's pure and persistent, ending with a pebbly note and firm tannins. Drink now through 2018. 5,818 cases made. (Spectator.)

Figure 1.2: Comparison of Chemical and Sensory Analysis in a Review

Wineinformatics has developed as a study using data science to further the understanding of wine related data for domain knowledge. The goal of this knowledge is to give other researchers a backbone of understanding for future applications in the area of wine. This form of wine research contains useful information for producers and consumers alike; producers may gain an improved understanding of the preferences of consumers so that the market provides a product more appropriately suited to their interests.

Other wine researchers back the study into flavors and aromas. According to Jorn Kleinhans, owner of Wine Elite Sommelier Co., "People talk about quality like a matter of preference and flavor, but while we've found that there are a number of personal preferences that influence what people like and think are best, there are also a number of objective factors" [9]. Kleinhans describes three factors that he uses with a team of sommeliers in blind tests to determine the quality of a wine: complexity, intensity, and balance. He describes complexity as the "distinct flavor compositions you pick up" and "descriptors of flavor profile" such as "plum, cherry, vanilla, or tobacco." He then defines intensity as in direct relation to how clear each of the individual flavors are to the drinker: "more intensely showing flavors make it easier to spot, appreciate, and recognize." Finally, he outlines balance as "the idea that an optimal wine contains a number of flavor profiles: fruits, vegetables, oak notes, the structure (which includes alcohol), and earthiness" [9]. He mentions "typicity" as a bonus factor, which means how "typical" a wine tastes compared to other wines from the same region. Obtaining the ability to analyze all of these factors, and specifically which ones contribute most to the quality of a wine, would allow producers to make more wines that their consumers enjoy.

Of particular interest are the individuals who review the wines. Wine critics gauge wines and usually assign them numerical ratings and tasting notes. The reviews these judges give have a definite significance because numerous customers make purchasing decisions based on their ratings. By discovering how judges may rate the wines from their respective regions, we gain a more informed estimation on the usefulness of their reviews. Popular wine reviewers include Wine Spectator, Wine Advocate, and Decanter [10-12].

The motivation for much of the research into the wine experts involves answering some very significant questions about the study of wine. For instance, the American Association of Wine Economics offers questions such as *"Who is a reliable wine judge? How can we aggregate the will of a tasting panel? Do wine judges agree with each* *other? Are wine judges consistent? What is the best wine in the flight?"* These questions provide the crucial basis that researchers need in order to make an informed analysis of the subject [13]. Other such questions involve what characteristics do wines from a similar or same region share, and how does this contrast with the differences between wines in other regions [14]? The climate, soil, and other factors that affect the grapes or their fermenting process that are often beyond the control of growers and researchers play a role in this analysis. However, this analysis may help provide insight into what the best regions for the wines are if certain regions have consistently high ratings.

Wine Spectator	**Wine Advocate**
95-100 Classic: a great wine	**96-100:** An **extraordinary** wine of profound and complex character displaying all the attributes expected of a classic wine of its variety. Wines of this caliber are worth a special effort to find, purchase, and consume.
90-94 Outstanding: a wine of superior character and style	**90 - 95:** An **outstanding** wine of exceptional complexity and character. In short, these are terrific wines.
85-89 Very good: a wine with special qualities	**80 - 89:** A **barely above average to very good** wine displaying various degrees of finesse and flavor as well as character with no noticeable flaws.
80-84 Good: a solid, well-made wine	**70 - 79:** An **average** wine with little distinction except that it is a soundly made. In essence, a straightforward, innocuous wine.
75-79 Mediocre: a drinkable wine that may have minor flaws	**60 - 69:** A **below average** wine containing noticeable deficiencies, such as excessive acidity and/or tannin, an absence of flavor, or possibly dirty aromas or flavors.
50-74 Not recommended	**50 - 59:** A wine deemed to be unacceptable.

Table 1.1: Wine Spectator's and Wine Advocate's rating systems.

The numerical ratings from expert wine judges at magazines such as Wine Spectator or Wine Advocate rely on a scale from 50-100 to evaluate whether a wine is unacceptable, below average, average, above average, outstanding, or extraordinary [15,16]. These fifty-point systems add more subtlety and nuance to the ratings compared to the twenty-point ratings used previously.

The tasting notes for Wine Spectator and Wine Advocate come from single-blind methodologies that test similar types of wines to provide a context, but they do not include too much information as to reveal the producer. A single-blind experiment means that the experimenters knew the outcome of the results, but the subjects, in this case the wine experts who made the reviews, did not. In a double-blind experiment, the information about the results remains hidden from both the experimenters and the subjects. Based on the rationale provided by Wine Spectator, a double blind experiment would completely eliminate some necessary context such as the "growing season, grape variety, and origin" [16,17].

One of the goals of Wineinformatics is to use the wine reviews that the judges collect for domain knowledge [14]. The reviews would be scanned into a dataset that would then be used as a basis for data mining algorithms that would attempt to make predictions based on the qualities of the wines. These data include the aforementioned "flavors and aromas" as well as "acidity, tannin, weight, finish, and structure" [14]. These attributes provide a crucial role in the analysis that Wineinformatics seeks to accomplish, and of significant note is the "ability for two people to simultaneously view the same wine differently while being able to share and detect all the same attributes." For example, data mining techniques such as clustering and association rules can find patterns that would associate a particular region or wine reviewer with various attributes that indicate a high wine rating for that region or reviewer. One of the results that Wineinformatics seeks is how each of these attributes relates to the score of the wine on the 50-100 scale mentioned earlier.

This thesis specifically seeks to determine the consistency of these wine judges through the use of data science concepts. By applying these concepts to the reviews that these judges produce, one can have a better understanding of how consistent each particular judge is as well as find any possible biases in the ratings. For example, one judge may rate wines consistently higher or lower than other judges, another judge may lean toward describing wines as having a particular type of flavor, and a third may have a completely different interpretation of the wines altogether.

CHAPTER 2: DATA

2.1 Data Collection

The research in this thesis primarily focuses on reviews collected from Wine Spectator. The reason for this is that this particular magazine has a significant influence on the culture of wine with its vast array of reviews and ratings from ten different wine reviewers, and it has been published since 1976 [10]. It has a database of hundreds of thousands of reviews on the website available to subscribers.

For this examination of wine reviews from Wine Spectator, we only use the wines from 2006 to 2015, from its ten reviewers shown in Table 2.1 and 2.2, in order to establish a set of data that reflects the recent conditions of each region to a larger extent. Table 2.1 describes the dataset that we have selected for these reviewers while Table 2.2 shows the positions of each reviewer as well as the wine regions that they taste from, which Wine Spectator describes as a “tasting beat” [10].

Reviewer\Category	**80-84**	**85-89**	**90-94**	**95-100**
James Laube (JL)	1250	7384	5168	357
Kim Marcus (KM)	1618	6690	3217	161
Thomas Matthews (TM)	1367	2981	1144	26
James Molesworth (JM)	3857	13628	6682	433
Bruce Sanderson (BS)	1148	7677	8618	451
Harvey Steiman (HS)	708	7657	5755	178
Tim Fish (TF)	1236	2531	1032	16
Alison Napjus (AN)	1510	4802	2095	33
MaryAnn Worobiec (MW)	833	3745	676	15
Gillian Sciaretta (GS)	66	355	7	0
Total	13593	57450	34394	1670

Table 2.1: Wine Spectator review metadata from 2006-2015.

Reviewer	Position	Tasting Beat
James Laube (JL)	Senior editor, Napa	California
Kim Marcus (KM)	Managing editor, New York	Argentina, Austria, Chile, Germany, Portugal
Thomas Matthews (TM)	Executive editor, New York	New York, Spain
James Molesworth (JM)	Senior editor, New York	Bordeaux, Finger Lakes, Loire Valley, Rhône Valley, South Africa
Bruce Sanderson (BS)	Senior editor, New York	Burgundy, Italy
Harvey Steiman (HS)	Editor at large, San Francisco	Australia, Oregon, Washington
Tim Fish (TF)	Senior editor, Napa	California Merlot, Zinfandel and Rhône-style wines, U.S. sparkling wines
Alison Napjus (AN)	Senior editor and tasting director, New York	Alsace, Beaujolais, Champagne, Italy
MaryAnn Worobiec (MW)	Senior editor and senior tasting coordinator, Napa	Australia, California (Petite Sirah, Sauvignon Blanc, other whites) and New Zealand
Gillian Sciaretta (GS)	Tasting coordinator, New York	France

Table 2.2: Wine Spectator reviewer profiles.

We explored how many reviews the judges made in each of our four categories, with over 107,000 in total. The middle categories contained far more reviews than the lowest and highest, and we had to accept that two of our reviewers, MaryAnn Worobiec and Tim Fish, had 15 and 16 reviews in the highest category respectively. Due to the lacking sample size of Gillian Sciaretta, specifically in the top two categories, we have decided to exclude this reviewer from our comparison of Wine Spectator judges [10].

A part of the purpose of Wineinformatics is to extract all important attribute information from the reviews, including structural information about the wine such as from the Boudreaux wine review example below: “well built” or “solid density” as well as flavor information relating to the dark plum sauce, blackberry coulis, charcoal, and tobacco, and finally the “fleshy finish.” The bold attributes hold specific interest, as they contain useful information about the individual properties of this wine relating to how it differs from other wines. This process involves an organizational hierarchical structure known as a Wine Wheel [14] that sorts these types of attributes so that they may later have computations performed on them, in order to find, for example, which attributes most often appear in higher rated wines.

DOMAINE DE L'A Castillon Côtes de Bordeaux *90 pts*

Well built*, with* ***solid density*** *for the vintage, this lets a* ***core*** *of* ***dark plum sauce****,* ***steeped currant*** *and* ***blackberry coulis*** *play out, while hints of* ***charcoal****,* ***anise*** *and* ***smoldering tobacco*** *line the* ***fleshy finish****. A* ***solid effort****. Drink now through 2018. 1,900 cases made. –JM*

2.2 Wine Aroma Wheel and Computational Wine Wheel

A fundamental concept to the field of Wineinformatics is the Wine Aroma Wheel and Computational Wine Wheel. The Aroma Wine Wheel, provided by retired professor and sensory chemist Ann C. Noble, includes a list of sensory attributes [18]. The wheel uses multiple levels and branches to separate broad categories of flavors or other attributes into more specific subcategories. For instance, the category of "FRUITY" contains other subcategories such as "TROPICAL FRUIT", "DRIED FRUIT", and "BERRY." The deepest level includes the wine tasting notes that would be attributed to a subcategory, and they include attributes such as "BLACKBERRY" in the review under the subcategory of "BERRY." This wheel provides a very useful basis for the types of attributes that one can apply data science techniques to in order to investigate their frequency and relation to the wine's other characteristics such as its rating or price [14].

The Computational Wine Wheel developed by Dr. Bernard Chen seeks to improve on the original Aroma Wheel. The Aroma wheel primarily focuses on flavors; however, many of the important non-flavor attributes, such as "tannins, acidity, body, structure, or finish" as well as descriptors that do not fit into any of the aforementioned categories, such as “dark” still give the wine a characteristic desirable to capture and compare to other attributes [14]. This adaptation to the wheel allows an increase in the quality of the measurements taken through data science techniques.

The Computational Wine Wheel has two distinct categories of attributes. The first category involves the specific attribute while the second attribute is its normalized name. The specific name describes how an attribute appears in the tasting notes while the normalized name describes the classification that the attribute is sorted into for the purposes of analysis through data mining. For instance, under the subcategory “TANNINS”, “FULL TANNINS”, “LUSH TANNINS”, and “RICH TANNINS” all represent specific names for the normalized name “TANNINS_HIGH.” The fact that such a variety of adjectives can describe one quality necessitates for several specific names to map to one normalized name. This wine wheel has been the basis for various white-box classification algorithms that would determine the consistency of the wines [14].

The Computational Wine Wheel required a second iteration, known as the Computational Wine Wheel 2.0. The purpose of this research was to enhance the vocabulary of the wheel from merely using one year, 2011, as the basis of the vocabulary, to using 10 years’ data from Wine Spectator as shown in Table 2.1, specifically from 2003 to 2013 [19]. The increase in specific and normalized terms, defined previously, to 1881 and 985 respectively, demonstrate the expansion of the wheel. Finally, I contributed to this wheel by solving an issue with plural attributes. For instance, “BLUEBERRY” and “BLUEBERRIES” would both be treated as the same specific term and normalized attribute in this new iteration, whereas the previous wheel did not account for these plural attributes [19].

	CWW 2.0	**CWW**
Data Source	Past 10 Years' Top 100 Wines from Wine Spectator	2011 Top 100 Wines from Wine Spectator
Categories	14	14
Subcategories	34	34
Specific Terms	1881	635
Normalized Attributes	985	444
Plurals	Yes	No

Table 2.3: Comparison of Old and New Computational Wine Wheel

2.3 Application of Wine Wheel on the Dataset

We programmatically extract the attributes from the description provided by the wine review using the wine wheel. The program sorts the specific attributes to detect from the wine wheel in decreasing order of length so that the program does not detect a shorter attribute that provides less useful information before a more precise longer attribute. For instance, we do not want the program to detect "BLACKBERRY" instead of "BLACKBERRY COULIS" in the Bordeaux wine example.

After the specific attributes in the wine wheel are in descending order, checks them in that order against a copy of the review. When a specific attribute is detected in the review, it is cut from the copy, and its normalized form is stored as information attached to the wine. Specifically, when "FLESHY FINISH" is detected in the Bordeaux example, "line the fleshy finish." is replaced with "line the." The final description for this review with the attributes removed is "Well built, with for the vintage, this lets a of, steeped and coulis play out, while hints of, and smoldering line the. A solid effort. Drink now through 2018. 1,900 cases made." Right now, only the first instance of a word is counted, in this case solid because it was removed the first time, but not the second.

The current Computational Wine Wheel extracts “fleshy finish” to mean “excellent finish” in this instance, while “solid”, “density”, and “core” relate to the body of the wine. Under the fruit category of the wine wheel, “blackberry”, “currant”, and “plum sauce” are detected as attributes. Other miscellaneous attributes include “charcoal” from the woody wine wheel category, “tobacco” from the herbs and vegetables category, and anise from the spicy category. Table 2.4 demonstrates an example of how we may represent the Computational Wine Wheel attributes for the Bordeaux wine; we represent each one as a binary attribute based on whether or not our scan of the Computational Wine Wheel detected that attribute for that wine.

Wine	**Berry**	**Excellent Finish**	**Blackberry**	**…**	**Score**
Castillon Côtes de Bordeaux	0	1	1		95

Table 2.4: Example of Storing Wine Attributes

However, this process can have issues. For example, the current wine wheel did not detect “well built”, which can apply to the body category, “blackberry coulis”, which describes a specific type of blackberry, and “solid effort” which describes the overall quality of the wine. Therefore, this research into the wine wheel is an ongoing process.

Right now, because the wine wheel is based on only the top 100 wines from Wine Spectator, it has a plethora of attributes for very good wines, but it may overlook other attributes that describe lower quality wines. We ran a program to detect which words appeared the most often in the 107,000 wine dataset that spanned 2006-2015, and we found that, for instance, “candy” appears extremely often, with 2,836 detections. For comparison, the word “wine” itself only appears 1,901 times. Therefore, we experimented with the wine wheel to detect these other lower quality attributes so that we can see if we have better results when attempting to predict the quality of a wine.

2.4 Program

During our research, we wanted to create a program that would easily preprocess wine data into a binary matrix in an Excel format for easier analysis of wine attributes. In this format, each wine would occupy a row, and each flavor attribute would occupy a column. So, we created a Visual Basic Wine Preprocessing form (Figure 2.1). This form would analyze a set of reviews with a given format by extracting specific keywords with accordance to the chosen wine wheel. Then, it will create a matrix to show if each of the normalized versions of those keywords are present in the review.

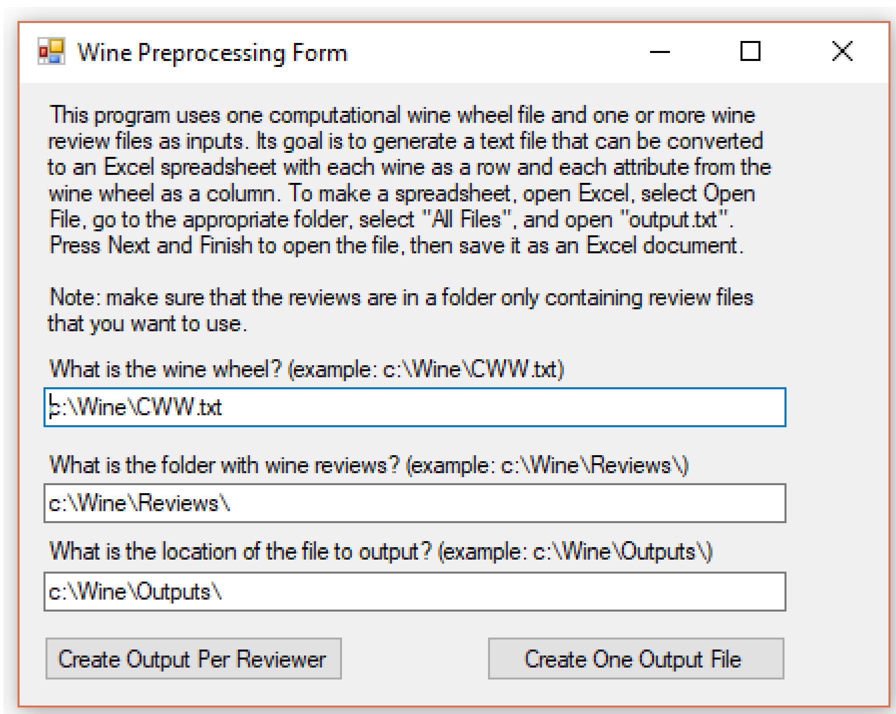
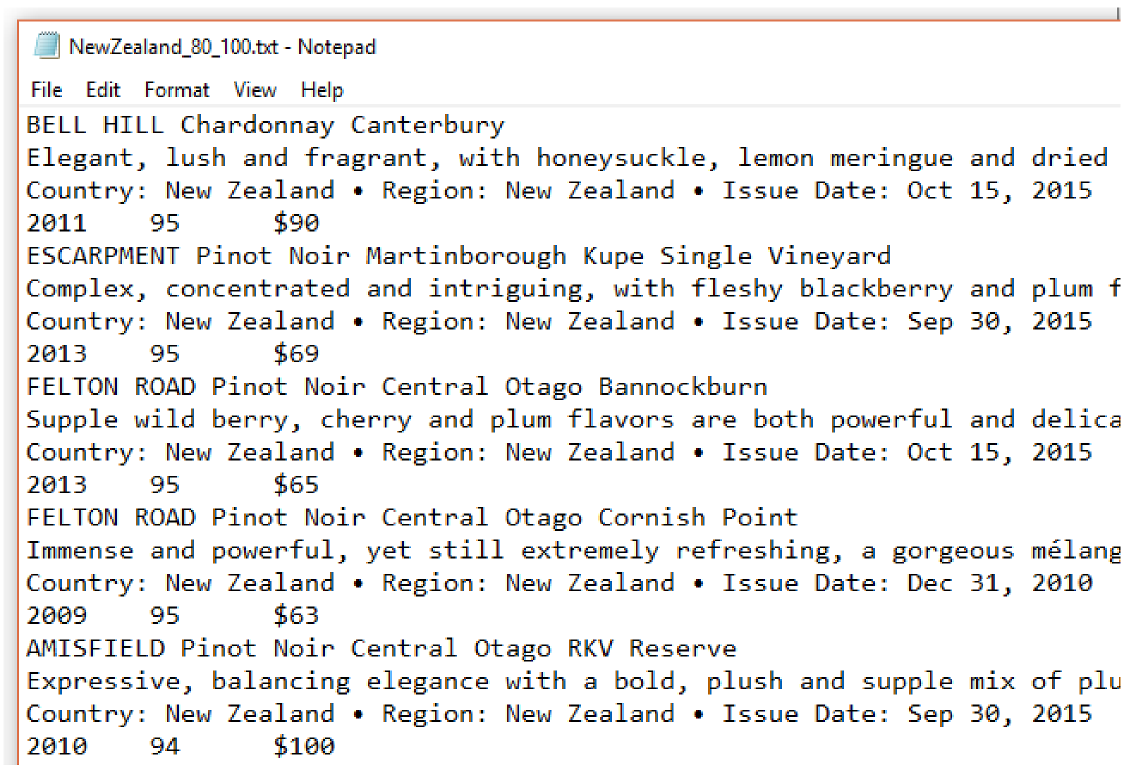


Figure 2.1: Wine Preprocessing Form

The format that this form accepts is a text file of wine reviews. For a review, the first line must have the name of the wine, the second line must have the wine’s description (the review itself), the third line must have the country and region information, and the final line must have the year, rating, and price of the wine, in order and tab delimited. The text file may have any number of subsequent reviews, as shown by the example (Figure 2.2).



```
NewZealand_80_100.txt - Notepad
File Edit Format View Help
BELL HILL Chardonnay Canterbury
Elegant, lush and fragrant, with honeysuckle, lemon meringue and dried
Country: New Zealand • Region: New Zealand • Issue Date: Oct 15, 2015
2011    95      $90
ESCARPMENT Pinot Noir Martinborough Kupe Single Vineyard
Complex, concentrated and intriguing, with fleshy blackberry and plum f
Country: New Zealand • Region: New Zealand • Issue Date: Sep 30, 2015
2013    95      $69
FELTON ROAD Pinot Noir Central Otago Bannockburn
Supple wild berry, cherry and plum flavors are both powerful and delica
Country: New Zealand • Region: New Zealand • Issue Date: Oct 15, 2015
2013    95      $65
FELTON ROAD Pinot Noir Central Otago Cornish Point
Immense and powerful, yet still extremely refreshing, a gorgeous mélang
Country: New Zealand • Region: New Zealand • Issue Date: Dec 31, 2010
2009    95      $63
AMISFIELD Pinot Noir Central Otago RKV Reserve
Expressive, balancing elegance with a bold, plush and supple mix of plu
Country: New Zealand • Region: New Zealand • Issue Date: Sep 30, 2015
2010    94      $100
```

Figure 2.2: Wine Review Format

The wine preprocessing form has a straightforward input and output. It uses the file path of the Computational Wine Wheel text file, the folder that contains the review files following the aforementioned format, and the location to output the matrix, which will be another text file. Depending on which button is pressed, the program can separate the output into a matrix for each reviewer it finds, or simply one large matrix. Afterwards, one can open Excel and then find and open the text file in order to obtain the binary matrix (Figure 2.3). In the example that includes nine wines and five attributes, two of the wines have the “citrus” normalized attribute.

	A	B	C	D	E	F
1		BLOOD OF	CITRUS	CITRUS PE	CITRUS ZE	CLEMENTI
2	ACACIA Pi	0	0	0	0	0
3	ACACIA Pi	0	0	0	0	0
4	ADRIAN F(	0	0	0	0	0
5	ADRIAN F(	0	0	0	0	0
6	ALDER SPF	0	1	0	0	0
7	ALTA MAF	0	1	0	0	0
8	ALTA MAF	0	0	0	0	0
9	AMOS CEL	0	0	0	0	0
10	ANDIS Gre	0	0	0	0	0

Figure 2.3: Preprocessing Output in Excel

The goal of this program is to allow for easier preprocessing of the wine data. Its ability to separate by reviewer helps us analyze and compare each reviewer more easily. This in turn allows us to more efficiently run data mining algorithms that would determine which types of wines more commonly contain which attributes, such as association rules and Naïve Bayes. These algorithms can detect patterns of which attributes contribute to successful wines, but it can also be used to predict unrated wines based on the attributes.

CHAPTER 3: ASSOCIATION RULES ON REGION-SPECIFIC WINES

3.1 Introduction

The unique properties of a region's soil ultimately defines a wine's quality. This chapter aims to demonstrate the connection between the attributes of region-specific wines and their quality. This region-specific analysis uses data mining to discover relevant patterns in wines specific to that particular area in order to answer such questions as, "What similarities do 95+ wines from the same region contain?" and "What combinations of attributes are common in the Napa Cabernet Sauvignon wines?" [21].

We hypothesize that wines from a similar region would share similar properties, and we aim to explore this concept in this chapter. We used as our dataset professionally written reviews for 1200 wines scoring 80 or above taken from Wine Spectator [10]. All of these wines belong to the Cabernet Sauvignon set of wines from the California Napa Valley region.

Because Cabernet Sauvignon is one of the region's most notable grapes, and because 90% of wine from the United States comes from California, we wanted to explore this specific region in depth. The purpose of this chapter is to demonstrate using patterns in wine's attributes to predict the quality of a wine not yet rated, "using only their professionally written reviews" [20].

3.2 Dataset

This chapter uses an association rule-based classification algorithm in order to predict wine quality, and it involves a wine region specific analysis of 1200 Cabernet Sauvignon wine reviews from Napa Valley, California "to construct a dataset for the Computational Wine Wheel" [21]. The dataset consists of 300 wines in each of four categories, 95+, 90-94, 85-89, and 80-84. We use the ECLAT association rules algorithm to discover the patterns in attributes that a wine rated 90 or above would hold on the 100-point scale. We base the attributes on the original Computational Wine Wheel 1.0 [22]. We also want to use the association rules to observe patterns on the entirety of the data so that we can predict a wine with an unknown rating and place it in a particular rating category.

3.3 Methods and Experimental Setup

3.3.1 Association-Based Classification

Association rules "describe if/then relationships among apparently unrelated data in a database" [21]. The underlying concept behind association rules reads that if a certain item exists within that database, other items may exist as well alongside it. Association rules are "created by analyzing data for frequent patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in a database and is found with the formula … where the rule is finding the probability that X=> Y" [21].

Confidence indicates "the number of times the if/then statements have been found to be true" [23]. The confidence has its own formula as well and "is the probability that a transaction, or an item of data, containing X also contains Y" [21].

$$Support(X \Rightarrow Y) = \frac{|X \cup Y|}{|T|} \qquad Confidence(X \Rightarrow Y) = \frac{|X \cup Y|}{|X|}$$

Each individual wine is "considered a separate transaction" [21]. We extract the attributes from the review, and these attributes are used to find a "correlation in the 'label,' or the score of the wines" [21]. This chapter, and by extension, this thesis, seeks to “predict accurately whether the wine in question is scored above or below 90, based solely on the attributes extracted from professionally written reviews” [21]. Table 3.1 provides an example [21].

Wines	**Apple**	**Berry**	**Depth**	**Wild Berry**	**Score (label)**
Wine 1	0	1	1	1	95 (+)
Wine 2	1	0	1	1	85 (-)
Wine 3	0	1	1	0	91 (+)
Wine 4	0	1	1	1	?

Table 3.1: Predicting a wine based on its attributes.

In Table 3.1, we demonstrate how we store the wines by their binary attributes as well as their scores. We use association rules to predict whether the unknown rated wine lies in the above-90 rating range or the below-90 rating range. Assume we set the minimum support at 50% and the minimum confidence at 80%; the minimum support count would be 2 (3 training wines * 0.5=1.5). Therefore, the single items that passes the minimum support count are BERRY, DEPTH, WILD BERRY and +. To generate a 2-itemset that passes minimum support count, we obtain (BERRY, DEPTH), (BERRY,+) and (DEPTH,+). Since we try to focus on the association classification problem, we only take rules with class (+,-) into account. The confidence to (BERRY => +) and (DEPTH => +) are 100% and 66%; since the user defined minimum confidence is 80%, we obtained one rule (BERRY => +) indicates “a wine with berry in their review, is a 90+ points wine” [21].

To further extend the association classification algorithm, we can produce a 3-itemset rule (BERRY, DEPTH => +) with support 66% and confidence 100%. Analyzing the above data we find a strong association rule involving “BERRY” and “DEPTH” among the above-90 class. This rule denotes that “a wine with berry and depth in their review, is a 90+ points wine." It is not known if "APPLE" and "WILD BERRY" appear together frequently in below-90 wines, since in this example we only have one member of that set, but calculations would render them as a support-passing and confidence-passing association. Now we are tasked with classifying Wine 4. We observe that Wine 4 shares the association "BERRY" and "DEPTH" with the above-90 class. We also notice that Wine 4 does not contain both "APPLE" and "WILD BERRY," and thus does not share any association rules with the below-90 class. Since Wine 4 contains more association rules shared by the above-90 class than the below-90 class, we predict that Wine 4 belongs to the above-90 class as well [21].

3.3.2 Experimental Setup

In order to obtain valid prediction accuracy, we apply five-fold cross validation in all of our measurements. This process begins by dividing our dataset into four categories. The wines are separated into the following groups based on their wine score: 95+, 90-94, 85-89, and 80-84. After the wines are separated, the individual groups are randomized to remove selection bias of any kind [21].

Following the randomization, the groups are further separated into five sections. The process iterates five times, first selecting the top section of each group as the testing dataset, and using the remaining sections of wines as the training dataset. The wines from the training dataset are examined, and all rules of association are discovered. Since the above-90 group and the below-90 group were segregated prior to this process, any association discovered among attributes is categorized as an association whose presence predicts an above-90 classification or a below-90 classification, depending on where this association was discovered. The wines in the testing dataset are treated as unknown and therefore are not examined until this phase is complete. Once all rules of association for the training wines are found, we look at the attributes from each wine in the testing set. Taking into consideration how many rules of association from each class the wine contains and how accurate each rule is, the wine is classified (e.g., suppose a wine conforms to eight association rules in the above-90 class and two from the below-90 class; it would most likely be classified as above-90) [21].

We also calculate coverage of wines because we discard the ones that we cannot accurately predict into a category for this experiment. For example, if a wine in a testing set has an equal likelihood of being above or below 90 based on the training data, we discard that wine in our results. Generally, coverage is not a major issue because only one or two percent of the total dataset would fall into this category of not being classified, but it is important to note [21].

3.4 Results

Finally, we observe the actual ratings of the testing set wines and see how often we correctly classified them. After the top group is used as the training dataset and all predictions were made, the process starts again, using the next group as the testing set and the remaining four as the training set. Once all groups have been the testing set once, the prediction accuracies are averaged, and we obtain a decent indication of how well our association rules predict the rating of a wine [21].

The advantage in using wines with known scores and treating them as unknown is that it allows us to test to see how accurately we can predict wine ratings before using our method on non-rated wines. Some wines in the testing set will have no associations in common with either the above-90 class or the below-90 class. In this case, we discard the wines that do not. As wines are discarded, our coverage, or the percentage of wines to which a rule can be applied, decreases. We have run five-fold cross validation using many combinations of minimum support and confidence parameters, and our coverage has never failed to be above ninety percent. Table 3.2 provides the prediction accuracy with coverage under different set of minimum support and confidence [21].

Confidence	**Support 1%**		**Support 2%**	
	Accuracy	**Coverage**	**Accuracy**	**Converge**
50%	60.42%	99.66%	63.69%	99.66%
60%	63.25%	99.66%	68.24%	99.41%
70%	63.34%	99.33%	70.03%	97.83%
80%	62.22%	98.75%	72.44%	93.33%
90%	61.65%	98.66%	72.56%	92.50%
100%	57.07%	98.16%	72.58%	92.50%
Confidence	**Support 3%**		**Support 4%**	
	Accuracy	**Coverage**	**Accuracy**	**Converge**
50%	64.84%	99.58%	65.53%	99.50%
60%	68.94%	99.42%	69.35%	99.25%
70%	70.85%	96.67%	71.75%	98.17%
80%	75.49%	94.25%	74.39%	97.33%
90%	76.80%	93.75%	74.77%	97.25%
100%	76.80%	93.75%	74.77%	97.25%

Table 3.2: Predicting a wine based on its attributes.

Table 3.2 pretty clearly demonstrates that higher user defined support and confidence produces higher prediction accuracy. Figure 3.1 gives a visual figure to see this trend clearly. Basically, the prediction accuracies showed in this chapter match the values reported in another study using the same association classification algorithm [6]. Since the dataset used in that study is different from this chapter [6], we do not compare their results directly. However, the major difference is the coverage; the coverage in that study is ranging from 50.90% (under minimum support=2% and minimum confidence=90%) ~ 98.40% (under minimum support=0.5% and minimum confidence=60%) [6]. The coverage in this chapter is always higher than 90%. This can be explained by the dataset we selected in this chapter is region specific; most wines produced by NAPA valley in California are somehow similar to each other [21].

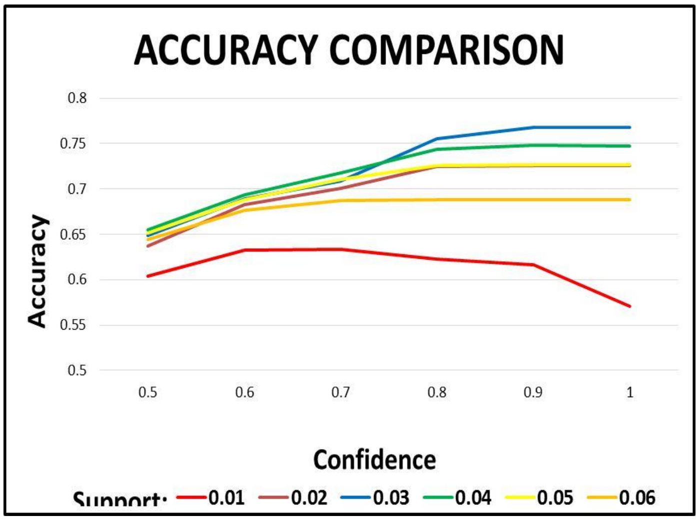


Figure 3.1. Various accuracies given varying support and confidence thresholds.

Figure 3.2 demonstrates the highest accuracy produced by different minimum support values. It is quite clear that 3% minimum support seems to be the optimal value. This finding may provide a good starting point for similar Wineinformatics datasets [21].

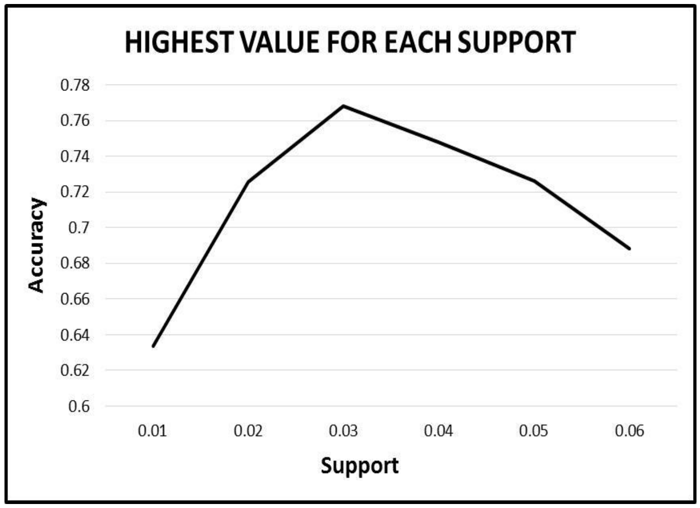


Figure 3.2. The maximum accuracy for each support value used.

We also listed some important attributes discovered by this chapter that relate to 90+ wines in Table 3.3. These attributes give the idea of what are important/characteristic attributes that make good wine in Napa Valley. It is pretty obvious that not only attributes related to fruit flavors, but also some "feelings" from the reviewers, such as ELEGANT, FINESSE, PURE, are included in Table 3.3 [21].

MATCHING ATTRIBUTES (above 90)		
BLACKBERRY	DEPTH	PERSISTENT
WILD BERRY	ELEGANT	POWER
BLACK CHERRY	FINE	PURE
ANISE	FINESSE	SMOOTH
CORE	FOCUSED	STYLE
CONCENTRATED	GRACE	SUBTLE
BALANCE	LAYERS	SUPPLE
DEEP	LONG	MOCHA
OAK	ESPRESSO	CURRANT&RICH

Table 3.3: Attributes strongly predicting a wine to be above 90.

3.5 Conclusion

This chapter has expanded Wineinformatics by creating and analyzing a Napa Valley region-specific dataset of 1200 wine reviews with association rule-based classification. We have found "that the accuracy of our predictions was satisfactory"; frequently reaching the 74% - 76% range while still maintaining above 90% coverage

[21]. This gives credence to our conclusion that region-specific wines will yield greater predictability given their inherent similarities.

These results have shown that wine reviews can in fact provide a basis for predicting wines' quality. The reviews led us to discover common rules of association among our wines, which, if for no other purpose, offers some interesting insight as to some common ground shared by wines in terms of qualitative attributes. To the best of our knowledge, this chapter is based on the first study to deal with single region wine dataset [21].

We believe it would be interesting to consider as a supplement to this research with one or more different region-specific subsets of the world's wines, such as Boudreaux in France, Pediment in Italy, or Rioja in Spain, to corroborate our results [21]. An alternative approach to explore would include further dividing this dataset of wines into the four categories and predicting which of these four categories each wine would belong to using five-fold cross validation. That would illustrate the nuances differentiating excellent wines from great wines from good wines to mediocre wines.

CHAPTER 4: EVALUATION OF JUDGES

4.1 Introduction

Wine judges provide a crucial tool for understanding the attributes that these wines exhibit. Wine judges taste a wine in order to determine which attributes that wine exhibits, and one can make a comparison among judges in order to find their collective reliability, aggregation, agreement, and consistency, as the Journal of Wine Economics alludes to [13]. It is important to keep in mind that these are simply the sensory qualifications of a particular wine, so using as many different reviewers' findings would allow us a more comprehensive understanding of their methodology and findings. For instance, if several judges can agree that a particular wine has a cherry flavor, then that solidifies the idea of associating that attribute for that wine.

The questions this thesis seeks to focus on are the consistency and reliability of the individual judges. For example, if two judges can agree with a high degree of accuracy on a particular type of flavor for a certain wine, that will increase their credibility. This section seeks to make a comparison between different judges' reviewing abilities. One of the ways in which this can be done is by evaluating each of the judges separately through use of different algorithms such as the Naïve Bayes algorithm, and a support vector machine. For this test, we use the dataset of 107,000 wines from Wine Spectator from 2006 to 2015. Another way this information may be used for various applications in Wineinformatics, is as to determine what characteristics certain judges may have a bias for and account for that, or to find characteristics that judges may share in common.

4.2 Classification Algorithms

Our goal is to use data classification in order to make a proper comparison between various wines, judges, and regions. This method separates data into several classes and uses predictive models known as classifiers in order to find out patterns and correlations in the data for further analysis. The inductive learning hypothesis, the basis for classification algorithms, describes "a heuristic search through a space of symbolic descriptions, generated by an application of certain inference rules to the initial observational statements …The inference rules include generalization rules, which perform generalizing transformations on descriptions, and conventional truth-preserving deductive rules (specialization and reformulation rules)" [24].

4.2.1 How Classification Algorithms Generate Prediction Models

The goal of classification algorithms is to use a collection of previously categorized data as a basis for categorizing new data. Therefore, the data will include a training set from which to base its classification on as well as a testing set that will predict unknown quantities into the previously established categories.

There is a difference between classification models and regression models. Classification is a discrete attribute while regression involves a continuous attribute [25]. This means that classification models will predict categorical labels, and regression models predict a value on a function. For example, we can build a classification model to categorize whether a wine will be rated above or below 90 points on a scale, or a regression model to predict exactly what the score for that wine would be given enough information. For the purposes of this thesis, we use only classification models.

4.2.2 Black-box and White-box Testing

There are several data mining algorithms that play a role in our evaluation of these judges. In this thesis, I use both black-box and white-box classification algorithms. The black-box algorithm that we shall use is the Support Vector Machine, or SVM, and the white box algorithm will be the Naïve Bayes algorithm. These will each play an important role.

The black-box algorithm offers merely a high accuracy with no way of validating the methodology. Black-box testing (also known as functional testing) is a "technique of testing without having any knowledge of the internal working of the application. It only examines the fundamental aspects of the system and has no or little relevance with the internal logical structure of the system" [26]. Black-box testing may be used in places where the people testing do not have enough understanding of the source code because of the little knowledge that it requires to use.

White-box testing (also known as structural testing) is "the detailed investigation of internal logic and structure of the code. In white box testing, it is necessary for a tester to have full knowledge of source code" [26]. White-box testing would be used where the people testing have had experience with the code, and therefore, would know how they come up the results of the program.

With regard to Wineinformatics, these both prove useful. For example, black-box testing will have a higher accuracy in its predictions, but it cannot extract the attributes that account for that accuracy in the way that white-box testing can (for example, one may find the attribute "BERRY" to correlate to above 90 rating with white-box testing, but not with black-box), so each have their unique set of advantages and disadvantages.

4.3 Evaluation Metrics

We use four different evaluation metrics to quantitatively evaluate wine reviewers. These include accuracy, precision, recall, and specificity. These each have their own unique significance and vary based on the dataset they are applied to.

In our predictions, we have four different types of results, true positives (successfully predicting a wine to be above 90), false positives (when a wine is rated below 90 but is predicted as above 90), true negatives (successfully predicting a wine to be below 90), and false negatives (when a wine is rated above 90 but is predicted as below 90). We base our use of the four evaluation metrics on these results.

The first evaluation metric, accuracy, describes the number of successful times we make a prediction that we can verify, and we verify it with through five-fold cross validation. In other words, this is the number of true cases divided by the total. These include both positive (for our test, predicting a wine to be rated above 90 points) cases and negative (predicting a wine to be rated below 90 points) cases. The second evaluation metric, precision, describes the true positives, divided by the number of true positives and false positives combined. This metric describes how many predictions are true of the wines predicted to be above 90. The third metric is recall, or specificity, which describes the number of true positives divided by the number of true positives and false negatives combined. This metric describes, of the wines rated above 90, how many were predicted accurately. The fourth metric is specificity, which is the number of true negatives divided by the number of true negatives and false positives combined. This metric describes of the wines rated below 90, how many were predicted accurately.

4.4 Naïve Bayes

4.4.1 Introduction to Naïve Bayes

Naïve Bayes is a classification algorithm with a basis on Bayes' Theorem that assumes the independence of the variables that it tests. "Naive Bayes models are so named for their "naive" assumption that all variables X_i are mutually independent given a 'special' variable C" [27]. For example, with relation to Wineinformatics, a particular wine may hold the attributes BERRY and APPLE. In this instance, the algorithm presupposes that these two attributes are completely independent of each other, which is the basis for the name "Naïve." The Naïve Bayes model is very easy to execute, and excels at working with massive datasets such as the one in this thesis relating to 107,000 wines. The Naïve Bayes algorithm is also well known for outperforming other classification algorithms in Wineinformatics, and as such, it became our algorithm of choice in this thesis [28].

Naïve Bayes' algorithm's basis, Bayes' Theorem, has four significant variables. Using our example of Wineinformatics, posterior probability could calculate how probable a wine is to be above or below 90 given its attributes. Likelihood calculates the probability of having that set of attributes given that it is above 90. The class prior probability is how probable a wine is to be above or below 90 before being given the training set on which the prediction is based, and the predictor prior probability is a measure of how probable the data within the training set is likely to be true.

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Figure 4.1: Bayes' Theorem, the basis for the Naïve Bayes algorithm.

A version of the Naïve Bayes algorithm that we must note is Laplacian smoothing. With the Naïve Bayes formula, one may have zero probabilities. For example, if we have the attribute "apple" in our testing set, but have never had that word in our training set, the probability *P(x|c)* will always be zero. This would cause us to ignore other testing attributes for this one word. Laplace smoothing alleviates this issue by adding a parameter such as one to both the numerator and the denominator so that these zero probabilities do not interfere with the probabilities of other attributes. This is important to note because we have made tests on our dataset with both the original Naïve Bayes algorithm as well as the Laplacian version.

4.4.2 Naïve Bayes Results

We ran the Naïve Bayes algorithm to determine how easy it is to predict whether a wine is above or below 90 for each of Wine Spectator's wine reviewers, as we demonstrate in Table 4.1. Across all of the reviewers, the Naïve Bayes Original algorithm trailed slightly behind the Laplace. This is due to the parameter (k = 1) adding slightly extra weight to zero attribute probabilities so that the above or below probability does not go to zero. Among all of these reviewers, there were only 49 instances of the probabilities of the original Naïve Bayes tied (specifically, with both at zero), and zero instances of Laplace's probabilities tied. These are the instances where both the above and below probabilities had zero attributes (which Laplace corrected for), so the program was forced to perform a virtual coin flip whereas the Laplace implementation did not have this problem.

Overall, Naïve Bayes has demonstrated that Naïve Bayes is the most successful white-box classification algorithm for these wine reviews. Previous research has indicated that it outclasses the decision tree, K-nearest neighbor, and association rules algorithms for wine reviews. The only algorithm that consistently surpasses Naïve Bayes' results is the black-box SVM classification [28]. We excluded the tenth reviewer, Gillian Sciaretta, as mentioned previously due to a lack of sample size.

In this case, Naïve Bayes continues to perform well. For example, the average accuracy for the original Naïve Bayes algorithm for the nine reviewers sampled is 84.2%, while the average accuracy for the Laplace version is 84.7%. Both of these are better than 5 in 6, or a dice roll with one undesirable number. These results are similar to previous results as well; for example, previous white-box testing yielded 79.6% accuracy for a Naïve Bayes that includes zero probabilities and 85.7% for the Laplace version of the algorithm.

The most reliable reviewer in this instance was Tim Fish, who had an accuracy of 87.37% with the original version of the algorithm and 88.16% with the Laplace version, as Table 4.1 demonstrates. However, MaryAnn Worobiec was a very close second with 87.36% and 88.04% respectively. Mining this type of information can give us an insight into which reviewers give more or less precise descriptions in their reviews, and with enough data collected on the reviewers, one could rank them by their reliability. In this case, by order of accuracy, these would be TF, MW, AN, JM, TM, KM, JL, BS, HS. We can use this information to create a variable for weight on their reviews; for example, a review from TF may have more weight than a review from HS.

	Original Naïve Bayes	**Laplace**
Reviewer: TF	**0.873727934**	**0.881619938**
Reviewer: MW	**0.873600304**	**0.88043272**
Reviewer: AN	**0.872867299**	**0.878909953**
Reviewer: JM	**0.868211382**	**0.870406504**
Reviewer: TM	**0.848314607**	**0.859188112**
Reviewer: KM	**0.845284956**	**0.849478008**
AVERAGE	**0.842050282**	**0.847173995**
Reviewer: JL	**0.802810933**	**0.805918497**
Reviewer: BS	**0.802615402**	**0.804794903**
Reviewer: HS	**0.791019723**	**0.793817317**

Table 4.1: Reviewers by order of Naïve Bayes accuracy

While the Laplace Correction had a slight edge over the Naïve Bayes Original (usually 0-1%), the SVM edged out Naïve Bayes by approximately 3% in all but AN (which it edged out by approximately 1%). The results of this program solidify our understanding that reviewers MW and TF have a higher rate of predictability.

Due to the skew of these datasets, with the vast majority of wines being below 90, or in the “false” prediction category, our results generally reflected high specificity, mediocre precision, and low recall. However, reviewer Bruce Sanderson remained the most consistent for our predictions despite the skew, with precision, recall, and specificity all within one percentage point for both the original Naïve Bayes and the Laplace correction.

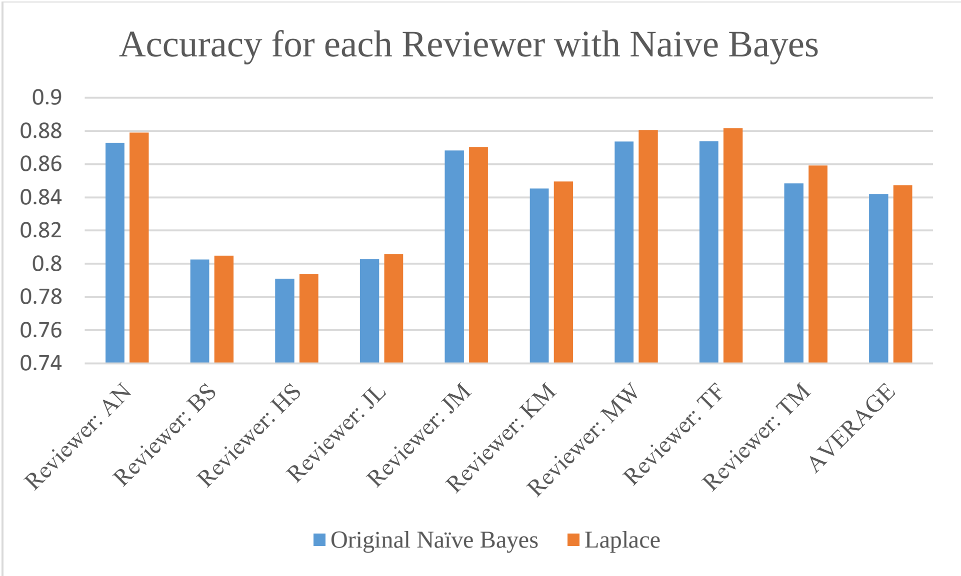


Figure 4.2: Accuracy for each reviewer with both original Naïve Bayes and Laplace.

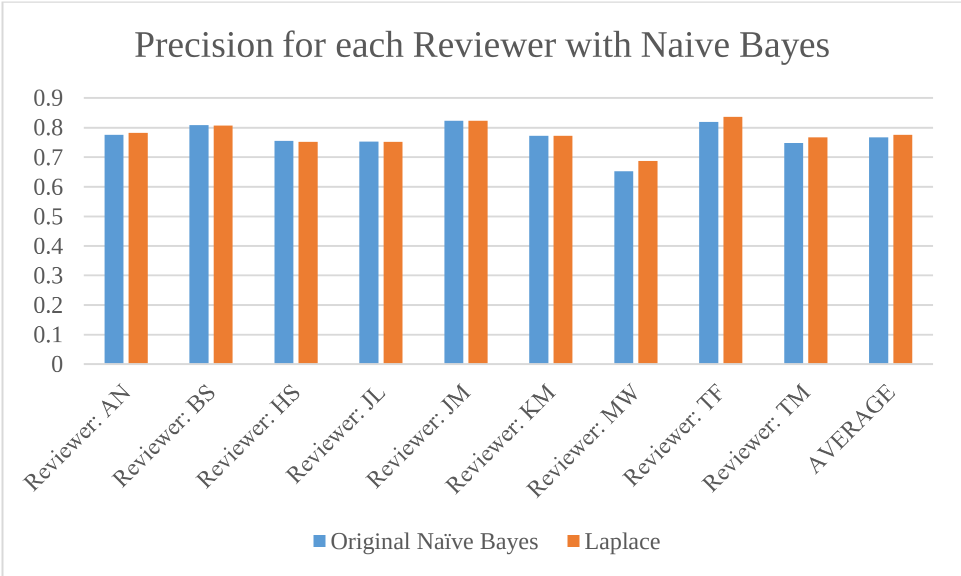


Figure 4.3: Precision for each reviewer with both original Naïve Bayes and Laplace.

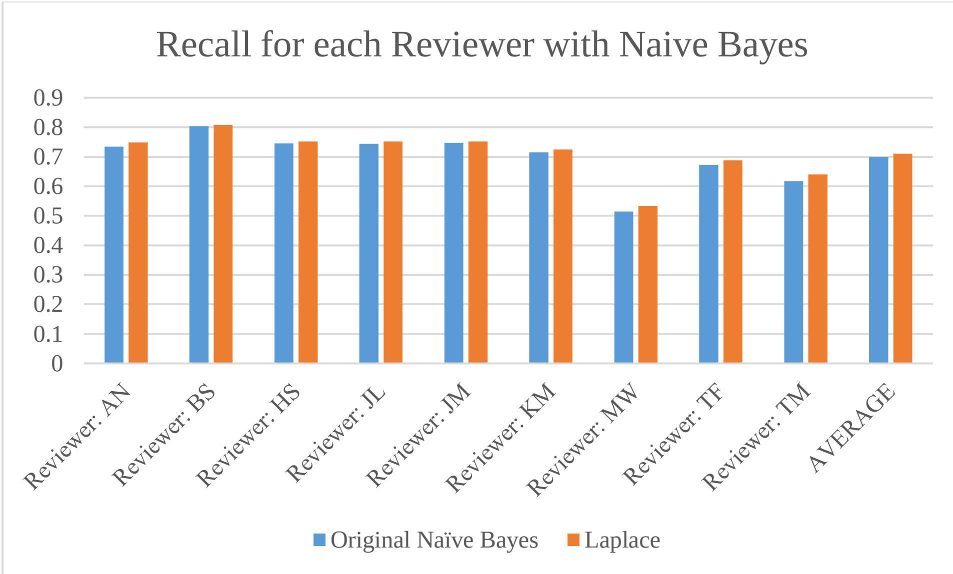


Figure 4.4: Recall for each reviewer with both original Naïve Bayes and Laplace.

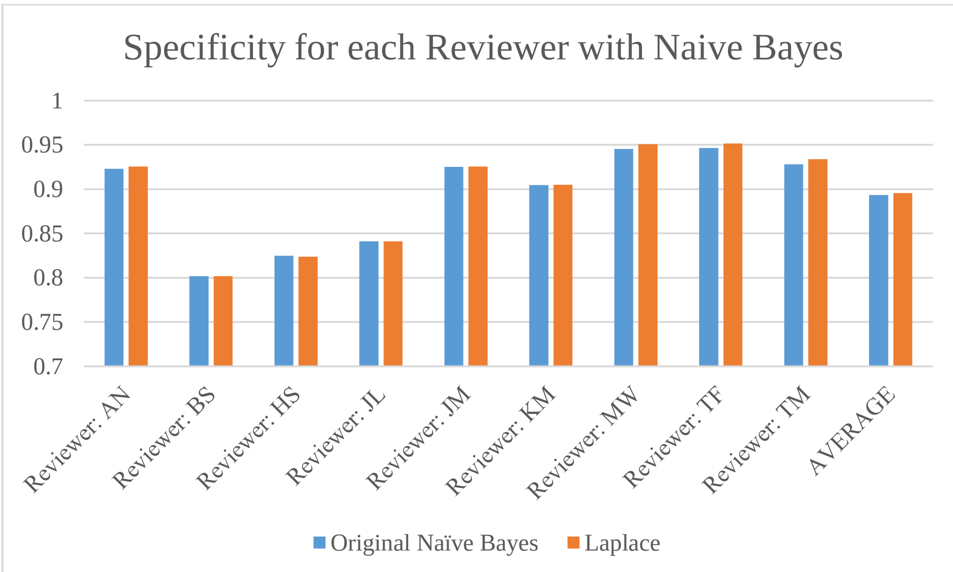


Figure 4.5: Specificity for each reviewer with both original Naïve Bayes and Laplace.

Reviewer: AN	Naïve Bayes Original	Laplace Correction
Accuracy	**0.872867299**	**0.878909953**
Precision	**0.776315789**	**0.782424812**
Recall/Sensitivity	**0.734548688**	**0.748651079**
Specificity	**0.923114198**	**0.925514801**
Reviewer: BS	Naïve Bayes Original	Laplace Correction
Accuracy	**0.802615402**	**0.804794903**
Precision	**0.808247877**	**0.807034954**
Recall/Sensitivity	**0.803463773**	**0.807658354**
Specificity	**0.801732984**	**0.801856884**
Reviewer: HS	Naïve Bayes Original	Laplace Correction
Accuracy	**0.791019723**	**0.793817317**
Precision	**0.75526715**	**0.751559076**
Recall/Sensitivity	**0.744723284**	**0.751559076**
Specificity	**0.824658858**	**0.8237896**
Reviewer: JL	Naïve Bayes Original	Laplace Correction
Accuracy	**0.802810933**	**0.805918497**
Precision	**0.753303167**	**0.751674208**
Recall/Sensitivity	**0.74441066**	**0.751130403**
Specificity	**0.840919701**	**0.841019699**
Reviewer: JM	Naïve Bayes Original	Laplace Correction
Accuracy	**0.868211382**	**0.870406504**
Precision	**0.823471539**	**0.823893183**
Recall/Sensitivity	**0.746845124**	**0.751827626**
Specificity	**0.925037302**	**0.925429983**
Reviewer: KM	Naïve Bayes Original	Laplace Correction
Accuracy	**0.845284956**	**0.849478008**
Precision	**0.772646536**	**0.772350503**
Recall/Sensitivity	**0.715068493**	**0.72492359**
Specificity	**0.904430065**	**0.904909113**
Reviewer: MW	Naïve Bayes Original	Laplace Correction
Accuracy	**0.873600304**	**0.88043272**
Precision	**0.652677279**	**0.687409551**
Recall/Sensitivity	**0.514253136**	**0.534308211**
Specificity	**0.945355191**	**0.950684932**
Reviewer: TF	Naïve Bayes Original	Laplace Correction
Accuracy	**0.873727934**	**0.881619938**
Precision	**0.81870229**	**0.836832061**
Recall/Sensitivity	**0.672413793**	**0.687304075**
Specificity	**0.946312518**	**0.951681266**
Reviewer: TM	Naïve Bayes Original	Laplace Correction
Accuracy	**0.848314607**	**0.859188112**
Precision	**0.747863248**	**0.767521368**
Recall/Sensitivity	**0.617501764**	**0.640057021**
Specificity	**0.928066325**	**0.933900365**
AVERAGE	Naïve Bayes Original	Laplace Correction
Accuracy	**0.842050282**	**0.847173995**
Precision	**0.767610542**	**0.775633302**
Recall/Sensitivity	**0.699247635**	**0.710824382**
Specificity	**0.893291905**	**0.895420738**

Table 4.2: Naïve Bayes Results

Figures 4.1, 4.2, 4.3, and 4.4 demonstrate the accuracy, precision, recall, and specificity respectively of each reviewer as summarized from Table 4.2. The graphs of accuracy and specificity highly correlate due to the greater sample size of below 90 wines. Of note is that all of the reviewers had a very high precision, generally around 80%, with the exception of MaryAnn Worobiec, one of the two higher scoring reviewers. The recall graph appears very similar to the precision graph for each reviewer.

We also ran through the program in order to test which attributes for each reviewer correlated at positively (were likely to have a 90 or higher rating) at least 90% of the time with at least 30 instances of the attribute. For example, Table 4.2 shows that the attribute Intense correlates to an above 90 rating 90.9% of the time with 33 instances, and each of the other attributes meet the same requirement. The purpose of this is to determine which words certain reviewers are likely to use when they describe a highly rated wine.

Reviewer	**Attributes correlated positively (>90 rating) with at least 30 instances**
AN	Intense 30/33, Beauty 55/58, Power 57/59, Seamless 43/44, Finesse 41/45
BS	Alluring 103/112, Excellent 182/184, Terrific 170/175, Refined 171/182, Seamless 77/80, Potential 141/149, Detailed 104/114, Beauty 285/290, Seductive 35/37, Gorgeous 33/34, Ethereal 50/53
HS	Deep 58/61, Elegant 276/305, Power 156/172, Long 765/849, Impresses 214/232, Complex 238/264, Seductive 83/92, Beauty 215/228, Tension 33/36, Remarkable 29/32, Gorgeous 71/73, Tremendous 53/53
JL	Plush 93/101, Seductive 82/88, Delicious 97/104, Wonderful 125/130, Opulent 41/44, Beauty 114/115, Remarkable 33/33, Gorgeous 65/65, Amazing 57/57
JM	Rock Solid 155/171, Seamless 146/162, Impresses 186/192, Turkish Coffee 69/75, Packed 150/159, Serious 94/98, Remarkable 51/52, Gorgeous 360/361, Terrific 104/104, Beauty 148/148, Wonderful 49/49, Backward 36/36, Stunning 57/57
KM	Complex 173/188
TM	Long 72/78

Table 4.3: Naïve Bayes Positively Correlated Attributes

Based on these results, there are several reviewers that do not have positively correlated attributes. From this, we can conclude that certain reviewers have words that they are likely to fall back on when describing quality wines, and we can use this information to make more accurate predictions about those particular reviewers and perhaps their biases. Unsurprisingly, nearly all of the attributes are generic praise such as "beauty," and we can use this in our prediction models. However, an attribute such as "Turkish Coffee" may have a high rating in general or only for the particular reviewer, and this requires further research.

4.5 SVM

4.5.1 Introduction to SVM

The support vector machine is a supervised machine learning black-box algorithm used to classify various data. In this particular algorithm, every set of coordinates represents a point in n-dimensional space. Because this algorithm is n-dimensional, it can lead to highly accurate results, but is very computationally expensive. The number n represents the total number of data points used.

According to a set of bioinformatics researchers from the University of California, University of Bristol, and University of Washington in Seattle, support vector machines (SVMs) "separate a given set of binary labeled training data with a hyper-plane that is maximally distant from them (known as 'the maximal margin hyper-plane'). For cases in which no linear separation is possible, they can work in combination with the technique of 'kernels', that automatically realizes a non-linear mapping to a feature space. The hyper-plane found by the SVM in feature space corresponds to a non-linear decision boundary in the input space" [29].

For example, for the purposes of Wineinformatics, each reviewer used has hundreds of wine reviews, which means this algorithm has hundreds of dimensions for all of them. A hyperplane then divides these data points, in this case, reviews, to predict whether the review will be above or below 90. Figure 4.6 gives a visual representation of how a hyperplane divides these dimensions.

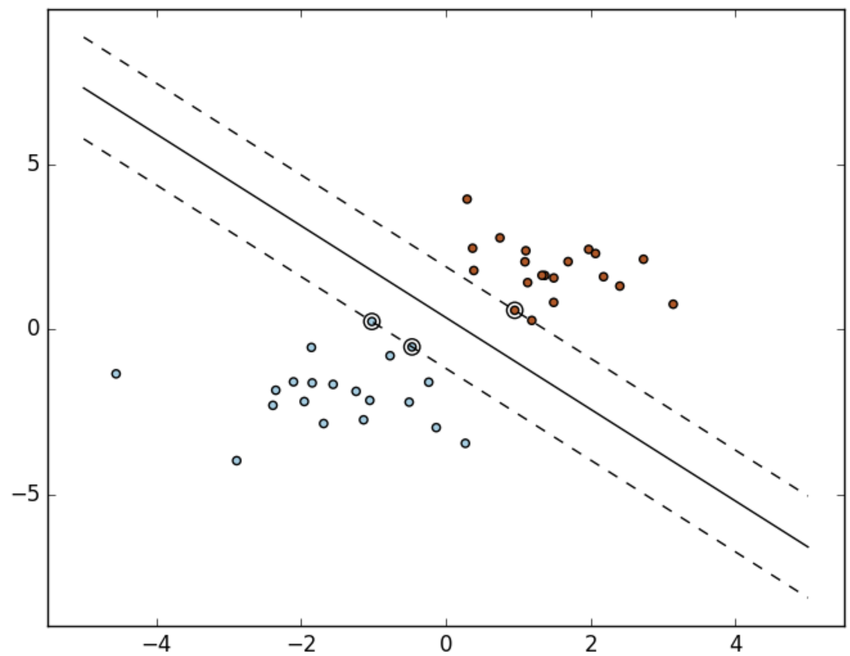


Figure 4.6: Visual implementation of using an SVM to divide data with a hyperplane.

There are also several different mathematical models with varying parameters that can be used by an SVM to compute the hyperplane. For instance, the SVC implementation takes training vectors from two different classes and a problem vector that it attempts to solve using a regularization parameter.

4.5.2 SVM Results

The SVM algorithm had the most successful results, but we cannot trace how it arrives at these results due to the black-box nature of the algorithm. For example, the average accuracy for the SVM is 87.2% for this dataset while the average accuracy for the Naïve Bayes is 84.2%, three percent higher. Again, reviewers MW and TF had the most successful results with the SVM as they did with the Naïve Bayes algorithm, both with more than 91% accuracy.

This is similar to previous tests where the SVM performs the most successfully. We can use this information as a guideline for programming our white-box classification algorithms. For example, the closer our algorithm is to the SVM ideal, the more reliable it is for us to use, and because it would be a white-box algorithm, we could trace how it arrives to its conclusions.

In our test of the accuracy of each reviewer using the SVM, we also tested only the aforementioned nine reviewers due to the tenth, Gillian Sciaretta, lacking an adequate sample size of reviews (she had only seven reviews in the 90-94 category and zero for the 95-100 category). We used LibSVM and tested with the C parameter in order to measure the rate of true positives, false positives, true negatives, and false negatives. We realized that when the parameter is a low amount (approaching zero), the SVM's dimension fails to label any of the reviews as positive. We found that the SVM gave the most accurate results when we set the C parameter between 100 and 200. Table 4.4 describes the summary of the peak performance of the results.

Most of the rest of the reviews have very high specificity, high accuracy, low precision, and very low recall. We believe this is the case because Sanderson's reviews were evenly balanced between above-90 and below-90 cases, with 49.3% of his reviews falling below 90, which demonstrates that this reviewer is more likely to rate the wines that he reviews higher.

Reviewer: AN	SVM Peak
Accuracy	**0.883649289**
Precision	**0.84085213**
Recall/Sensitivity	**0.705357143**
Specificity	**0.959759189**
Reviewer: BS	SVM Peak
Accuracy	**0.831284229**
Precision	**0.838064372**
Recall/Sensitivity	**0.827434116**
Specificity	**0.835807365**
Reviewer: HS	SVM Peak
Accuracy	**0.826479228**
Precision	**0.825583716**
Recall/Sensitivity	**0.745659869**
Specificity	**0.895517035**
Reviewer: JL	SVM Peak
Accuracy	**0.824705135**
Precision	**0.802684295**
Recall/Sensitivity	**0.733755656**
Specificity	**0.887653463**
Reviewer: JM	SVM Peak
Accuracy	**0.893536585**
Precision	**0.854740313**
Recall/Sensitivity	**0.778074491**
Specificity	**0.949613955**
Reviewer: KM	SVM Peak
Accuracy	**0.872924867**
Precision	**0.824255628**
Recall/Sensitivity	**0.72794553**
Specificity	**0.941742898**
Reviewer: MW	SVM Peak
Accuracy	**0.913835642**
Precision	**0.921428571**
Recall/Sensitivity	**0.520984081**
Specificity	**0.997597204**
Reviewer: TF	SVM Peak
Accuracy	**0.912149533**
Precision	**0.880721221**
Recall/Sensitivity	**0.71851145**
Specificity	**0.977170162**
Reviewer: TM	SVM Peak
Accuracy	**0.890540051**
Precision	**0.84057971**
Recall/Sensitivity	**0.658974359**
Specificity	**0.972171113**
AVERAGE	SVM Peak
Accuracy	**0.872122729**
Precision	**0.847656662**
Recall/Sensitivity	**0.712966299**
Specificity	**0.93522582**

Table 4.4: Reviewer Accuracies with SVM

Most of the reviewers resemble Alison Napjus's curve, as illustrated by Figure 4.7. Some of the more exaggerated versions of this curve, such as MaryAnn Worobiec in Figure 4.7, happen when the reviewer is extremely likely to rate reviews below 90 (in her instance, 86.9% of her ratings are below 90). Table 4.5 demonstrates the reviewers' rankings by SVM accuracy concisely. From our results, we found that the most consistent reviewer is MaryAnn Worobiec because her reviews consistently have 91% accuracy, as shown in Figure 4.7.

	SVM Peak
Reviewer: MW	**0.913835642**
Reviewer: TF	**0.912149533**
Reviewer: JM	**0.893536585**
Reviewer: TM	**0.890540051**
Reviewer: AN	**0.883649289**
Reviewer: KM	**0.872924867**
AVERAGE	**0.872122729**
Reviewer: BS	**0.831284229**
Reviewer: HS	**0.826479228**
Reviewer: JL	**0.824705135**

Table 4.5: Reviewers by order of SVM accuracy

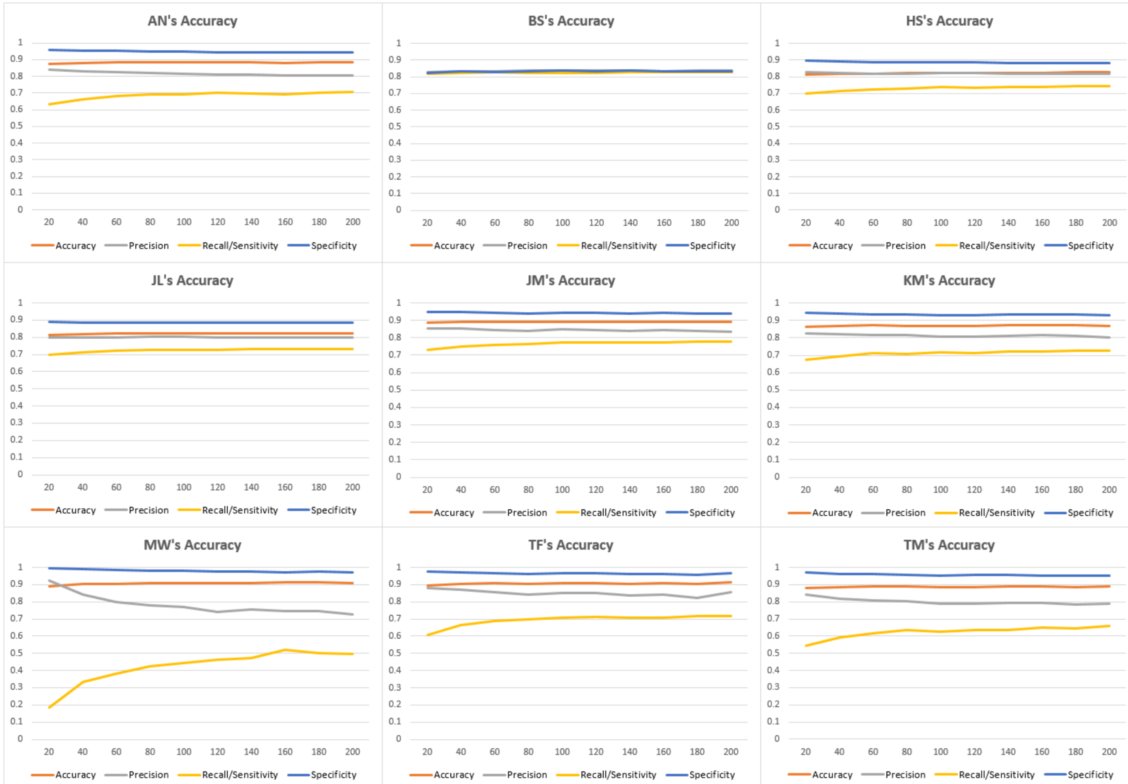


Figure 4.7: Reviewer SVM Results

4.6 Conclusion

Overall, the Naïve Bayes algorithm has provided very good results, with reviewers MaryAnn Worobiec and Tim Fish providing 88% accuracy; these reviewers reached 91% in the SVM as well. We can conclude that these two reviewers have the most reliability among the ones we have tested in determining whether the data they have sampled would have a rating higher or lower than 90. They fit the model of our computational wine wheel's choice of attributes more accurately than any of the other reviewers based on this information, and a point of expansion may be to investigate these reviewers further or examine what attributes can improve the accuracy of the rest of the reviewers.

CHAPTER 5: ADDITIONAL ATTRIBUTES

5.1 Introduction

During our research in analyzing the 107,000 reviews, we searched for some attributes that may have some significance that are not based on the top 100 wines. We wanted to learn whether adding these attributes would have an effect on predicting whether a wine has above or below 90 rating. Because they are not from the best wines, they could either increase or decrease the total accuracy of the results, and we wanted to see if there would be any significant change.

We have made a program to search through the most popular words from our dataset of 107,000 reviews. We ignored common English words such as articles like "the" and kept words that held some description of the wine. We added keywords that appeared 200 or more, a total of 77.

The most popular word included in the reviews, for example, was "candy", and Table 5.1 shows all 77 of the attributes as well as how many times they appeared in the results. We tested the Naïve Bayes algorithm again and calculated the difference between adding the new attributes and not having them.

ATTRIBUTE	COUNT	ATTRIBUTE	COUNT	ATTRIBUTE	COUNT	ATTRIBUTE	COUNT
CANDY	2836	DILL	504	COLOR	343	CHARM	237
MEDIUM-BODY	2358	COMPOTE	501	PITHY	338	SCENTED	237
RUSTIC	2023	KIWI	474	GLAZED	325	LACY	237
LEAN	1991	CAKE	472	LIP-SMACKING	319	BURST	230
FRIENDLY	1012	CHUNKY	453	BRAMBLY	314	GRAVEL	221
SIMPLE	787	FLASH	436	JUNIPER	312	PRECISE	219
EASY	777	WELL-KNIT	436	COARSE	296	UP-FRONT	214
THYME	746	GRITTY	435	POACHED	291	MANDARIN	214
STREAK	738	WOVEN	434	SMOLDERING	294	WAX	210
BLOSSOM	722	ENERGY	421	HARD	275	SALTED	208
PRESERVES	712	EDGED	416	SHARP	267	QUICK	207
GLIDE	695	GUTSY	414	GOLDEN	266	SAPPY	206
SNAP	645	GRILLED	413	UNCTUOUS	257	BREAD	203
FOREST	571	PAPAYA	412	RUBY	256	HEAVY	202
RAW	566	LACKS	387	PULP	252	SLOW	202
MEDIUM	550	CHIME	387	QUICKLY	252	VEGETAL	201
MATURE	549	MOUTHFILLING	377	COCO	246	GALA	200
SHY	549	BUILT	360	NEW	243		
EASY-DRINKING	536	ASSERTIVE	356	PROMINENT	239		
EASYGOING	533	BURLY	350	BURSTING	238		

Table 5.1: All Positively Correlated Attributes

5.2 Results

We tested for the Naïve Bayes original algorithm and the Laplace version, and we found the accuracy difference, on average, slightly worse by 0.01%, a percent of a percent. In Figure 5.1, one thing of note that we can see is that most of the reviewers had a slight increase, specifically AN, BS, HS, JL, KM, and TM. The one reviewer whose results were relatively consistent was the reviewer with the highest sample size of reviews, JM. This makes sense because with enough data, the accuracy should even out. However, the two reviewers that previously had the highest accuracy, MW and TF, saw a comparatively significant decline in accuracy. One possible conclusion is that because the attributes from the original Computational Wine Wheel correlated so well with these reviewers, adding new ones only serves as a detriment.

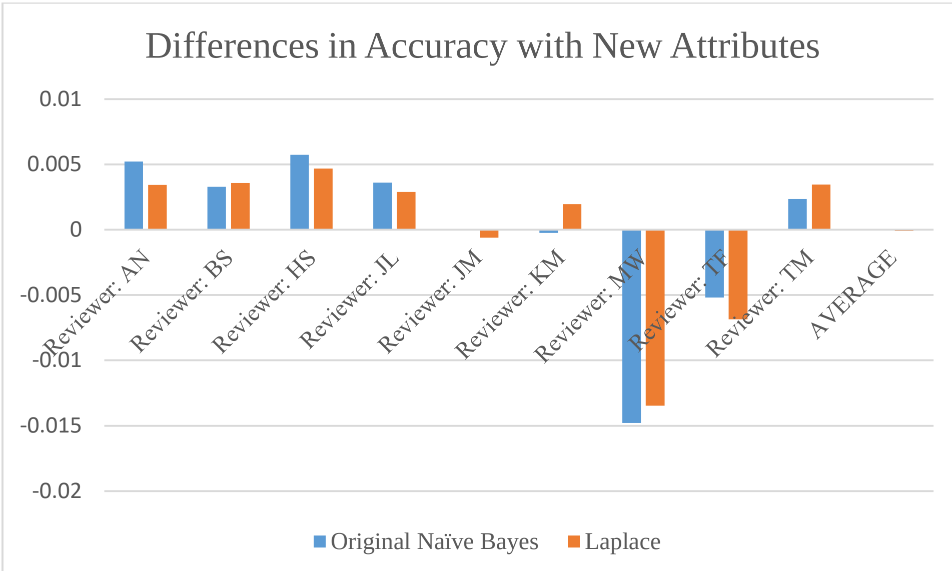


Figure 5.1: Difference in Accuracy to Naïve Bayes after New Attributes

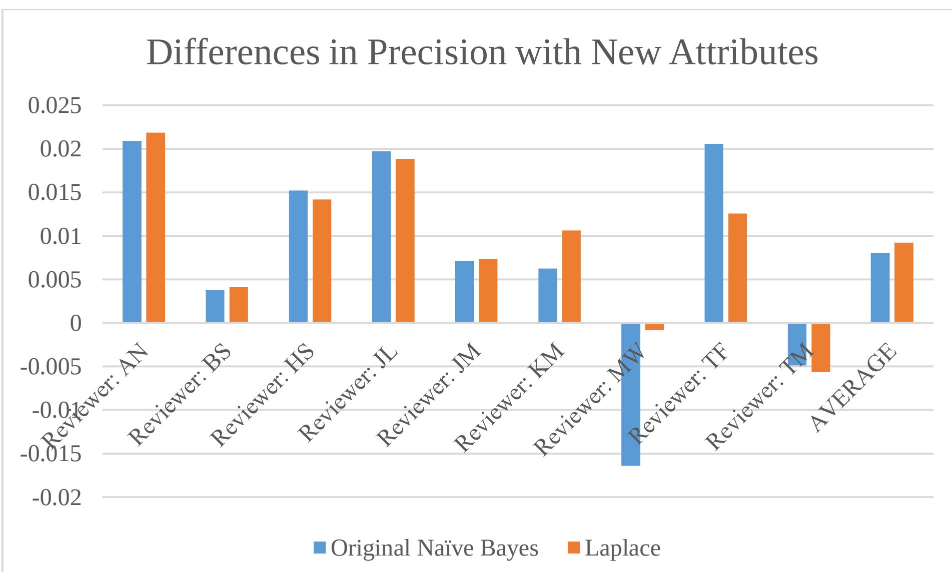


Figure 5.2: Difference in Precision to Naïve Bayes after New Attributes

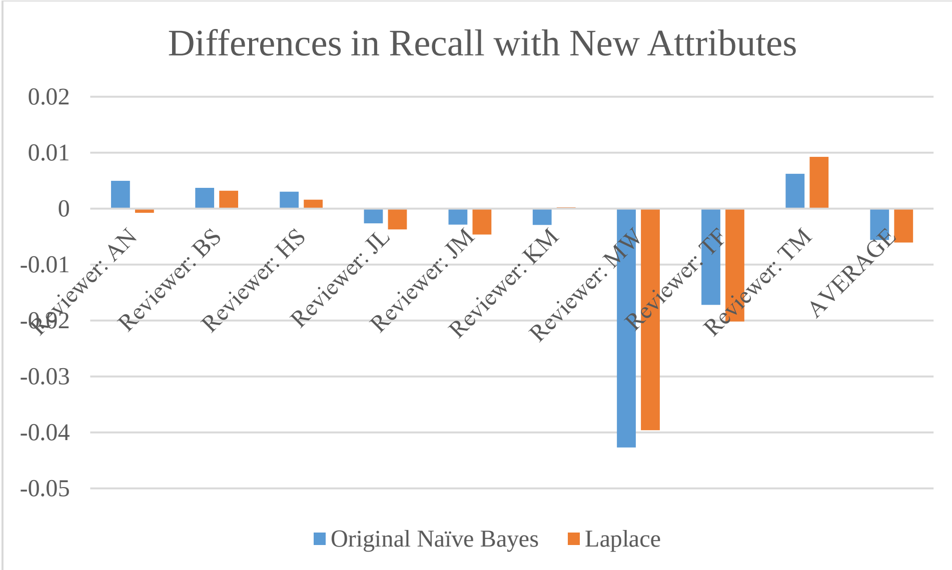


Figure 5.3: Difference in Recall to Naïve Bayes after New Attributes

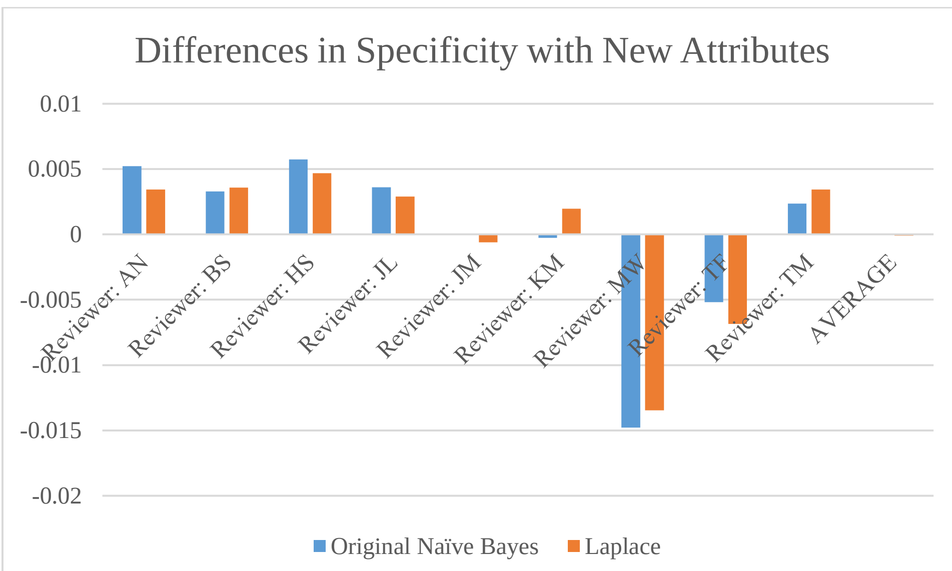


Figure 5.4: Difference in Specificity to Naïve Bayes after New Attributes

Figures 5.1, 5.2, 5.3, and 5.4 demonstrate respectively the differences in accuracy, precision, recall, and specificity for each reviewer. The differences generally fall within 1%, so they do not appear to have a significant impact. While most reviewers gained precision, the most accurate reviewer with Naïve Bayes, MW, saw a notable decrease.

Reviewer: AN	**Naïve Bayes Original**	**Difference**	**Laplace Correction**	**Difference**
Accuracy	**0.878080569**	**0.00521327**	**0.882345972**	**0.003436019**
Precision	**0.797230384**	**0.020914595**	**0.804269372**	**0.02184456**
Recall/Sensitivity	**0.739528498**	**0.00497981**	**0.747876883**	**-0.000774196**
Specificity	**0.929843512**	**0.006729314**	**0.932301646**	**0.006786845**
Reviewer: BS	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.805891376**	**0.003275974**	**0.808371058**	**0.003576155**
Precision	**0.81202344**	**0.003775563**	**0.811171783**	**0.004136829**
Recall/Sensitivity	**0.807180996**	**0.003717223**	**0.810833146**	**0.003174792**
Specificity	**0.80447876**	**0.002745776**	**0.805771731**	**0.003914847**
Reviewer: HS	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.79675408**	**0.005734357**	**0.798502504**	**0.004685187**
Precision	**0.770467499**	**0.015200349**	**0.76573328**	**0.014174204**
Recall/Sensitivity	**0.747766762**	**0.003043478**	**0.753148458**	**0.001589382**
Specificity	**0.8337134**	**0.009054542**	**0.831898315**	**0.008108715**
Reviewer: JL	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.806411903**	**0.00360097**	**0.808813682**	**0.002895185**
Precision	**0.773024429**	**0.019721262**	**0.77051981**	**0.018845602**
Recall/Sensitivity	**0.741794911**	**-0.002615749**	**0.747425336**	**-0.003705067**
Specificity	**0.850706298**	**0.009786597**	**0.850140456**	**0.009120757**
Reviewer: JM	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.868252033**	**4.0651E-05**	**0.869796748**	**-0.000609756**
Precision	**0.830603886**	**0.007132347**	**0.831261687**	**0.007368504**
Recall/Sensitivity	**0.743983074**	**-0.00286205**	**0.747200173**	**-0.004627453**
Specificity	**0.927592509**	**0.002555207**	**0.928016485**	**0.002586502**
Reviewer: KM	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.845028556**	**-0.0002564**	**0.851446383**	**0.001968375**
Precision	**0.778894516**	**0.00624798**	**0.782964139**	**0.010613636**
Recall/Sensitivity	**0.712124726**	**-0.002943767**	**0.725150757**	**0.000227167**
Specificity	**0.906351757**	**0.001921692**	**0.908639459**	**0.003730346**
Reviewer: MW	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.858804461**	**-0.014795843**	**0.866962977**	**-0.013469743**
Precision	**0.636275524**	**-0.016401755**	**0.686566656**	**-0.000842895**
Recall/Sensitivity	**0.471552838**	**-0.042700298**	**0.494721905**	**-0.039586306**
Specificity	**0.942302225**	**-0.003052966**	**0.949875324**	**-0.000809608**
Reviewer: TF	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.868535826**	**-0.005192108**	**0.874766355**	**-0.006853583**
Precision	**0.839272058**	**0.020569768**	**0.849393213**	**0.012561152**
Recall/Sensitivity	**0.655234244**	**-0.017179549**	**0.66715633**	**-0.020147745**
Specificity	**0.951439467**	**0.005126949**	**0.954693564**	**0.003012298**
Reviewer: TM	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.850669471**	**0.002354864**	**0.862629947**	**0.003441835**
Precision	**0.742977118**	**-0.00488613**	**0.761855568**	**-0.0056658**
Recall/Sensitivity	**0.623748896**	**0.006247132**	**0.649286428**	**0.009229407**
Specificity	**0.92701365**	**-0.001052675**	**0.932904659**	**-0.000995706**
AVERAGE	Naïve Bayes Original		Laplace Correction	
Accuracy	**0.842047586**	**-2.69611E-06**	**0.847070625**	**-0.00010337**
Precision	**0.775640984**	**0.008030442**	**0.784859501**	**0.009226199**
Recall/Sensitivity	**0.693657216**	**-0.005590419**	**0.704755491**	**-0.006068891**
Specificity	**0.897049064**	**0.00375716**	**0.899360182**	**0.003939444**

Table 5.2: Reviewer Accuracies with Additional Attributes

5.3 Conclusion

In our tests, two reviewers, MaryAnn Worobiec and Tim Fish, saw a decrease, and reviewer James Molesworth had almost no change, while the rest saw a slight increase with the new attributes. However, there was no significant difference after adding the new attributes, as Table 5.2 shows. This is well within the margin of error, and because the results are virtually identical, we can conclude that the current iteration of the Computational Wine Wheel based on 10 years' Top 100 Wines is robust enough for our calculations. These additional attributes these attributes do not appear to have enough significance to serve as a basis for our Computational Wine Wheel.

CHAPTER 6: CONCLUSION AND FUTURE WORKS

6.1 Conclusion

As mentioned previously, Wineinformatics has developed as a study that uses data science to further the understanding of wine related data for domain knowledge. Our goal is to expand this domain knowledge so that it may serve as a reference when others use this type of data to make predictions and analyses on different types of wines. We want to answer the questions from the American Association of Wine Economics offers questions such as *"Who is a reliable wine judge? How can we aggregate the will of a tasting panel? Do wine judges agree with each other? Are wine judges consistent? What is the best wine in the flight?"*

Our basis for answering these questions involves the computational wine wheel based on the Aroma Wine Wheel by retired professor and sensory chemist Ann C. Noble. We do this to extract attributes from reviews such as "well built" or "solid density" and analyze them to determine whether they have a positive or negative correlation with the review. The Computational Wine Wheel uses specific attributes found in the reviews, and normalized attributes that can be used for analysis; multiple specific attributes may fall into one normalized category. We have created a Wine Preprocessing Form that uses a computational wine wheel fine and one or more wine review files as inputs. This program creates a text file that can be converted to an Excel spreadsheet with a binary matrix of each attribute in each review.

We tested the association rules algorithm on region-specific wines. This algorithm gave a 77% accuracy result with 3% minimum support and 100% confidence. These rules matched various attributes to these rules, some specific such as "MOCHA" and others reflecting the "feelings" of the reviewers such as "ELEGANT." The reviews led us to find common rules of association among our wines, which, if for no other purpose, offers some interesting insight as to some common ground shared by wines in terms of qualitative attributes.

We have used Wineinformatics to investigate the reliability of different wine reviewers. This thesis focuses on predicting whether a particular wine is above or below 90 based on the reviews of Wine Spectator. We have used a white-box and a black-box classification algorithm, respectively the Naïve Bayes and SVM, for this test.

The Naïve Bayes and SVM algorithms specifically expand on previous research we have done with the association rules algorithm. Reviewers MaryAnn Worobiec and Tim Fish appear to give the most accurate results in our tests, both with higher than 91% in our SVM calculations and 88% in Naïve Bayes, so these two reviewers may have the most consistency in the attributes described by their reviews.

We have also discovered some additional attributes that were not included in the Top 100 Wines' attributes by searching for common descriptors in the text of the reviews. We used the ones that appeared at least 200 times, for a total of 77 new attributes, and found little difference in the results, leading us to the conclusion that the current iteration of the Computational Wine Wheel based on 10 years' Top 100 Wines has enough attributes to make accurate predictions.

This paper has made a quantitative analysis of each reviewer based on his or her reviews to expand the field of Wineinformatics. This analysis involves using evaluation metrics, which include accuracy, precision, recall, and specificity, to determine the consistency of each reviewer.

6.2 Future Works

There are many steps we could take to expand this research. These include more sources for wine reviews, running tests on different wine reviewers with the same wine or different regions, and using more than two rating categories. Each of these can help us to better analyze the wine reviews.

We can use more sources for reviews, and we can compare multiple reviewers for each wine. For example, in our dataset, we used 107,000 reviews from Wine Spectator, and we could expand this to using reviewers from Wine Advocate. Adding this dataset would give us a stronger sample of reviewers for our research.

In addition, we can run tests on the same type of wine with different reviewers or regions in order to create a clearer comparison of the reviewers or regions, rather than comparing the reviewers each as a whole with all of their reviews. Having these datasets would give us a better understanding of how we can use these data to determine the reliability of a particular reviewer or region.

We have only also tested whether a wine is likely to be above or below 90 with these results. This gives a straightforward but simplistic result. A possible inclusion would be a multi-class experiment with four rating categories: 80-84, 85-89, 90-94, and 95-100.

With this, we can conclude that Wineinformatics is a growing field. Many new studies can be made to discover new properties of wines, their regions, and their reviewers. Our goal is to conduct research can prove useful to the wine industry, and there are a plethora of avenues to explore for doing so.

REFERENCES

[1] “What is Data Science?” Available from Berkeley, accessed on November 20, 2017. <https://datascience.berkeley.edu/about/what-is-data-science/>.

[2] “What is Data Science?” Available from NYU, accessed November 20, 2017. <http://datascience.nyu.edu/what-is-data-science/>.

[3] Chaudhuri, S., Dayal, U. *An Overview of Data Warehousing and OLAP Technology.* ACM Sigmod Record, pp. 517-526. 1997.

[4] Raamesh, L., Uma, G.V. *Advances in Computing and Information Technology: First International Conference,* ACITY 2011, Chennai, India, pp. 92. 2011.

[5] “World Wine Consumption.” Available from the Wine Institute, accessed November 20, 2017. <http://www.wineinstitute.org/files/World_Wine_Consumption_by_Country_Revised_Nov_2015.pdf>.

[6] S. Ebeler. Flavor Chemistry - *Thirty Years of Progress, chapter Linking flavor chemistry to sensory analysis of wine.* Kluwer Academic Publishers, pp. 409-422. 1999.

[7] Chemical analysis of grapes and wine: techniques and concepts. Campbelltown, Australia: Patrick Iland wine promotions. 2004.

[8] Cortez Paulo, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. *Modeling wine preferences by data mining from physicochemical properties.* Decision Support Systems 47(4), pp. 547-553. 2009.

[9] Kane, Libby. “A Sommelier Explains How TO Tell If Your Wine Is High Quality.” Available from Libby Kane, accessed November 20, 2017. <http://www.businessinsider.com/recognize-high-quality-wine-2014-6>.

[10] “Wine Spectator Home.” Available from Wine Spectator, accessed November 20, 2017. <http://www.winespectator.com/>.

[11] Parker, Robert. “Robert Parker Wine Advocate.” Available from Robert Parker, accessed November 20, 2017. <http://www.robertparker.com/>.

[12] “Wine.” Available from Decanter, accessed November 20, 2017. <http://www.decanter.com/wine>.

[13] Masset, P., Weisskopf, J. P., & Cossutta, M. *Wine tasters, ratings, and en primeur prices*. Journal of Wine Economics, 10(01), pp. 75-107. 2015.

[14] Chen, Bernard, Rhodes, C., Crawford, A., Hambuchen, L.. Wineinformatics: *Applying Data Mining on Wine Sensory Reviews Processed by the Computational Wine Wheel.* ICDMW, pp. 142-149. 2014.

[15] “Wine Spectator’s 100-Point Scale.” Available from Wine Spectator, accessed November 20, 2017. <http://www.winespectator.com/display/show/id/scoring-scale>.

[16] Parker, Robert. “TWA Rating System.” Available from Robert Parker, accessed November 20, 2017. <https://www.robertparker.com/ratings>.

[17] “Why We Taste Blind.” Available from Wine Spectator, accessed November 20, 2017. <http://images.winespectator.com/wso/pdf/WShowwetasteLTR.pdf>.

[18] Noble, Ann C. “Wine Aroma Wheel.” Available from Ann C. Noble, accessed November 20, 2017. <http://winearomawheel.com/>

[19] Chen, Bernard, Rhodes, C., Yu, A., Velchev, V. *The Computational Wine Wheel 2.0 and the TriMax Triclustering in Wineinformatics.* ICDM 2016, pp. 223-238. 2016.

[20] “Top 10 Wine Regions of the World.” Available from Wine Folly, accessed November 20, 2017. <http://winefolly.com/update/top-wine-regions-of-the-world/>.

[21] Chen, Bernard, Velchev, V., Nicholson, B., Garrison, J., Iwamura, M., Battisto, R. *Wineinformatics: Uncork Napa's Cabernet Sauvignon by Association Rule Based Classification.* IEEE 14th ICMLA, pp. 565-569. 2015.

[22] Chen, Bernard, Rhodes, C., Crawford, A, Hambuchen, L. *Wineinformatics: Applying Data Mining on Wine Sensory Reviews Processed by the Computational Wine Wheel.* 2014 IEEE ICDMW, pp. 142-149. 2014.

[23] Ishibuchi Hisao, Tomoharu Nakashima, and Manabu Nii. *Classification and modeling with linguistic information granules: advanced approaches to linguistic data mining.* Springer. 2005.

[24] Michalski, R. *A theory and methodology of inductive learning. Artificial Intelligence*, 20(2), pp. 111-161. 1983.

[25] Fayyad, U., Piatetsky-Shapiro G., Smyth, P. *From Data Mining to Knowledge Discovery in Databases*. Al Magazine, 17(3), pp. 37-54. 1996.

[26] Khan, M., Khan, F. *A Comparative Study of White Box, Black Box and Grey Box Testing Techniques,* pp. 12. IJACRA, 3(6). 2012.

[27] Lowd, D., Domingos, P. *Naive Bayes Models for Probability Estimation*. ICML 22nd International Conference on Machine Learning, pp 529-536. 2005.

[28] Chen, Bernard, Le, H., Rhodes, C., Che, D. *Understanding the Wine judges and Evaluating the Consistency through White-box Classification Algorithms*. ICDM 2016, pp. 239-252. 2016.

[29] Furey, T., Cristianini, N., Duffy, N., Bednarski, D., Schummer, M., Haussler, D. *Support vector machine classification and validation of cancer tissue samples using microarray expression data.* Bioinformatics, 16(10), pp. 906-914. 2000.