

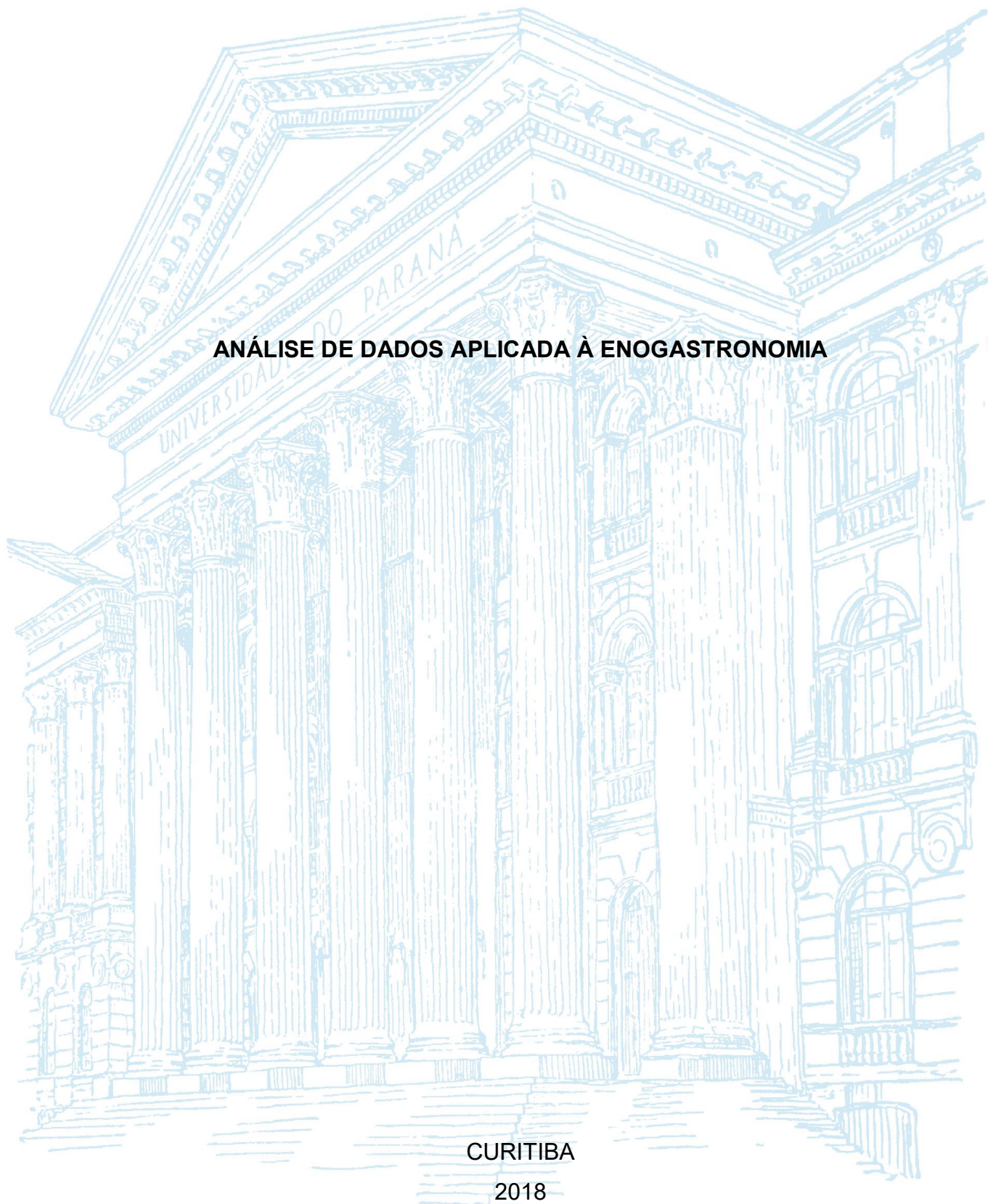
UNIVERSIDADE FEDERAL DO PARANÁ

ANDRÉIA KAROLINA LAPKOUSKI

ANÁLISE DE DADOS APLICADA À ENOGASTRONOMIA

CURITIBA

2018



ANDRÉIA KAROLINA LAPKOUSKI

ANÁLISE DE DADOS APLICADA À ENOGASTRONOMIA

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção de grau de Bacharel no Curso de Gestão da Informação, Departamento de Ciência e Gestão da Informação, Setor de Ciências Sociais Aplicadas, da Universidade Federal do Paraná.

Orientadora: Prof.^a Dr.^a Denise Fukumi Tsunoda

CURITIBA

2018

RESUMO

Estudo de natureza qualitativa que objetiva realizar uma análise de dados a partir de técnicas estatísticas e mineração de dados, a fim de identificar relações e padrões de ocorrência em uma base de dados Enogastronômica. Visa a concepção de uma base de dados que reúna elementos da Gastronomia e Enologia delimitando-se a pizzarias de Curitiba premiadas nos concursos Bom Gourmet 2018 e Comer&Beber 2017/2018, cuja finalidade é comparar os resultados de 2 (dois) algoritmos de mineração de dados, *Apriori* e *Prism*, voltados para a descoberta de padrões a partir de regras de associação e classificação; e, aplicar 2 (dois) métodos estatísticos, Qui-quadrado e Análise de Correspondência Múltipla (ACM), com o propósito de realizar uma análise descritiva da base de dados. Apresenta resultados satisfatórios na execução dos algoritmos de mineração de dados, em que a aplicação do *Apriori* resulta em regras com grau de interesse (*lift*) superior a 1,45, apresentando uma correlação positiva entre os atributos, além de compor regras com dois ou mais atributos no consequente; enquanto para o *Prism* considerou os atributos que possuem uma taxa de acerto inicial de 85,00% entre as instâncias, onde o maior índice de acerto foi na aplicação cujo atributo-meta é o ingrediente molho de tomate com 96,89%. Valida as regras geradas nestas aplicações junto aos especialistas com a finalidade de verificar a ocorrência destas no ato de uma harmonização ou a geração de um novo conhecimento. Revela que na inferência estatística, o teste Qui-quadrado apresenta diferenças estatisticamente significativas em um dos atributos e no ACM é possível explicar o método em 2 (duas) dimensões e visualizar 4 (quatro) agrupamentos distintos. Verifica a existência de relação entre um agrupamento da ACM e as regras do algoritmo *Prism*. Sugere como trabalhos futuros, a ampliação da base de dados para a realização de uma análise qualitativa dos dados; a concepção de outras bases de dados Enogastronômicas voltadas a demais estilos de Gastronomia; e, a criação de um guia de harmonização a partir dos dados coletados, visando auxiliar a escolha do cliente dos estabelecimentos participantes do estudo.

Palavras-chave: Mineração de Dados. Análise Descritiva. Pizzas e Vinhos.

ABSTRACT

Qualitative research that aims to perform a data analysis from statistical techniques and data mining, in order to identify relations and patterns of occurrence in an Enogastronomic database. Its objective is to design a database that gathers elements of Gastronomy and Oenology delimiting the pizzerias of Curitiba awarded in the competitions “Bom Gourmet 2018” and “Comer & Beber 2017/2018”, whose purpose is to compare the results of two (2) data mining algorithms, Apriori and Prism, aimed at the discovery of standards from rules of association and classification; and also apply two (2) statistical methods, Chi-square and Multiple Correspondence Analysis (MCA), in order to perform a descriptive analysis of the database. It presents satisfactory results in the execution of the data mining algorithms, in which the Apriori application accrues in rules with a degree of interest (lift) superior to 1.45, presenting a positive correlation between the attributes, besides composing rules with two or more attributes consequent; while on Prism was considered the attributes that have an initial hit rate of 85.00% between the instances, where the highest hit rate was in the application whose target attribute was the tomato sauce ingredient with 96.89%. It validates the rules generated in these applications with the specialists for the purpose of verifying their occurrence in the act of a harmonization or the generation of a new knowledge. It reveals that in statistical inference, the chi-square test presents statistically significant differences in one of the attributes and in the MCA it is possible to explain the method in 2 (two) dimensions and to visualize 4 (four) distinct groupings. It verifies the existence of a relation between a grouping of MCA and the rules of the Prism algorithm. It suggests as future actions the expansion of the database to perform a qualitative analysis of the data; the design of other Enogastronomic databases focused on different styles of Gastronomy; and, the creation of a harmonization guide from the data collected, aiming to help the client to choose the establishments participating in the study.

Keywords: Data Mining. Descriptive Analysis. Pizzas and Wines.

LISTA DE ABREVIATURAS E SIGLAS

ACM – Análise de Correspondência Múltipla

GI – Gestão da Informação

KDD – *Knowledge Discovery in Databases*

RI – Recuperação da Informação

SciELO – *Scientific Eletronic Library Online*

SiBi – Sistema de Bibliotecas da Universidade Federal do Paraná

UFPR – Universidade Federal do Paraná

UFU – Universidade Federal de Uberlândia

UFG – Universidade Federal de Goiás

UFPE – Universidade Federal de Pernambuco

WWW – *World Wide Web*

WEKA – *Waikato Environment for Knowledge Analysis*

WOS – *Web Of Science*

XML – *eXtensible Markup Language*

LISTA DE FIGURAS

FIGURA 1 – MODELO ECOLÓGICO PARA O GERENCIAMENTO DE INFORMAÇÃO	22
FIGURA 2 – CICLO INFORMACIONAL.....	22
FIGURA 3 – MODELO PROCESSUAL DA GESTÃO DA INFORMAÇÃO	23
FIGURA 4 – REPRESENTAÇÃO DO PROCESSO DE RECUPERAÇÃO DA INFORMAÇÃO (RI)	25
FIGURA 5 – TAXONOMIA DOS MODELOS DE RI.....	27
FIGURA 6 – PROCESSO DA ESTATÍSTICA DESCRITIVA.....	29
FIGURA 7 – ETAPAS DO KDD (Knowledge Discovery in Database)	31
FIGURA 8 – DIVISÃO DE NÍVEIS DA METODOLOGIA CRISP-DM.....	33
FIGURA 9 – FASES DO MODELO DE REFERÊNCIA.....	34
FIGURA 10 – HARMONIZAÇÃO POR CONCORDÂNCIA.....	37
FIGURA 11 – HARMONIZAÇÃO POR CONTRAPOSIÇÃO	38
FIGURA 12 – MODELO FISAR DE HARMONIZAÇÃO	42
FIGURA 13 – MÉTODO MERCADINI.....	45
FIGURA 14 – CARACTERIZAÇÃO DA PESQUISA	51
FIGURA 15 – PROCEDIMENTOS METODOLÓGICOS.....	52
FIGURA 16 – PERCENTUAL ACUMULADO ENTRE AS DIMENSÕES	78

LISTA DE GRÁFICOS

GRÁFICO 1 - DIMENSÕES DA APLICAÇÃO ACM	79
GRÁFICO 2 - AGRUPAMENTO DAS VARIÁVEIS GERADAS	80

LISTA DE QUADROS

QUADRO 1 – NÚMERO DE RESULTADOS DE BUSCA NO SISTEMA DE BIBLIOTECAS DA UFPR.....	14
QUADRO 2 – NÚMERO DE RESULTADOS DE BUSCA NO PORTAL DA CAPES	15
QUADRO 3 – NÚMERO DE RESULTADOS DE BUSCA NA BASE <i>WEB OF SCIENCE</i>	15
QUADRO 4 – NÚMERO DE RESULTADOS DE BUSCA NA BASE SCOPUS	16
QUADRO 5 – NÚMERO DE RESULTADOS DE BUSCA DA BASE SCIELO	16
QUADRO 6 – REPRESENTAÇÃO DOS ESTUDOS COM TEMAS RELACIONADOS À <i>DATA MINING</i> , MINERAÇÃO DE DADOS E ENOGASTRONOMIA.....	17
QUADRO 7 – DISTINÇÃO ENTRE DADOS, INFORMAÇÃO E CONHECIMENTO	21
QUADRO 8 – CARACTERÍSTICAS DE HARMONIZAÇÃO ENTRE COMIDA E VINHO	39
QUADRO 9 – ASPECTOS DE AVALIAÇÃO DE COMIDAS.....	44
QUADRO 10 – ASPECTOS DE AVALIAÇÃO DE VINHOS.....	44
QUADRO 11 – MODELO DE AVALIAÇÃO DAS SENSações DAS COMIDAS ...	46
QUADRO 12 – MODELO DE AVALIAÇÃO DAS SENSações DOS VINHOS	47
QUADRO 13 – COMBINAÇÕES COMUNS ENTRE PRATOS E VINHOS.....	49
QUADRO 14 – DESCRIÇÃO DOS ATRIBUTOS DO VINHO INCLUINDO SEUS TIPOS E VARIAÇÕES.....	53
QUADRO 15 – DESCRIÇÃO DOS ATRIBUTOS DOS PRATOS INCLUINDO SEUS TIPOS E VARIAÇÕES.....	55
QUADRO 16 – TRANSFORMAÇÃO DOS TIPOS DE ATRIBUTOS DOS VINHOS	56
QUADRO 17 – TRANSFORMAÇÃO DOS TIPOS DE ATRIBUTOS DOS PRATOS	58
QUADRO 18 – PIZZARIAS PREMIADAS NO CONCURSO VEJA BOM GOURMET 2017/2018.....	59
QUADRO 19 – PIZZARIAS PREMIADAS NO CONCURSO BOM GOURMET GAZETA DO POVO 2018.....	60
QUADRO 20 – CONJUNTO DE ATRIBUTOS SIMILARES DE INGREDIENTES DAS PIZZAS	62

QUADRO 21 – CONJUNTO DE ATRIBUTOS SIMILARES AOS AROMAS POSSIVEIS DO VINHO	62
QUADRO 22 – ATRIBUTOS DESCONSIDERADOS DA BASE DE DADOS	63
QUADRO 23 – SABORES DAS PIZZAS HARMONIZADAS	70
QUADRO 24 – OCORRÊNCIA NAS INSTÂNCIAS QUANTO AOS INGREDIENTES NA COMPOSIÇÃO DAS PIZZAS	71
QUADRO 25 – FAIXA DE PREÇO DE COMERCIALIZAÇÃO DOS VINHOS	72
QUADRO 26 – ESTILO DO VINHO.....	72
QUADRO 27 – VARIEDADE DE UVAS UTILIZADAS NA FABRICAÇÃO DOS VINHOS.....	73
QUADRO 28 – CARACTERISTICAS DE COMPOSIÇÃO DO VINHO	74
QUADRO 29 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE EMBUTIDOS.....	76
QUADRO 30 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE MOLHO DE TOMATE	76
QUADRO 31 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE MUÇARELA	77
QUADRO 32 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE OUTROS QUEIJOS	77

SUMÁRIO

1	INTRODUÇÃO	11
1.1	PROBLEMATIZAÇÃO	12
1.2	OBJETIVOS	13
1.2.1	Objetivo Geral	13
1.2.2	Objetivos Específicos	13
1.3	JUSTIFICATIVAS	13
1.3.1	Justificativa para a Ciência	13
1.3.2	Justificativa para a Área da Gestão da Informação	17
1.3.3	Justificativa Pessoal	18
1.4	DELIMITAÇÃO DA PESQUISA	18
1.5	ESTRUTURA DO DOCUMENTO	19
2	REFERENCIAL TEÓRICO	20
2.1	GESTÃO DA INFORMAÇÃO	20
2.1.1	Recuperação da Informação	25
2.2	ESTATÍSTICA	28
2.2.1	Estatística Descritiva	29
2.3	KDD (<i>KNOWLEDGE DISCOVERY IN DATABASES</i>)	30
2.3.1	Mineração de Dados	32
2.4	ENOGASTRONOMIA	35
2.4.1	Modelo Mercadini	43
2.4.2	Combinações Comuns Entre Vinhos e Comidas	48
3	ENCAMINHAMENTOS METODOLÓGICOS	50
3.1	CARACTERIZAÇÃO DA PESQUISA	50
3.2	MATERIAIS E MÉTODOS	51
3.2.1	Base de Dados	52
3.2.2	Testes Estatísticos da Base de Dados	64
3.2.3	Mineração de Dados	65
4	RESULTADOS E ANÁLISES	70
4.1	BASE DE DADOS	70
4.1.1	Análise Estatística da Base de Dados	75

4.1.1.1	Qui-quadrado (χ^2)	75
4.1.1.2	Análise de Correspondência Múltipla (ACM)	78
4.2	MINERAÇÃO DE DADOS	80
4.2.1	Algoritmo <i>Apriori</i>	81
4.2.2	Algoritmo <i>Prism</i>	83
4.2.3	Validação da Proposta.....	86
4.3	SÍNTESE DOS RESULTADOS	89
5	CONSIDERAÇÕES FINAIS	92
5.1	ALCANCE DOS OBJETIVOS	92
5.2	CONTRIBUIÇÕES.....	94
5.3	TRABALHOS FUTUROS.....	94
	REFERÊNCIAS	95
	APÊNDICE A – QUESTIONÁRIO DE VALIDAÇÃO DA PESQUISA	102
	APÊNDICE B – VARIEDADE DE VINHOS	107
	APÊNDICE C – PAÍSES E REGIÕES DE PRODUÇÃO DOS RÓTULOS	109
	APÊNDICE D – TEMPO DE AMADURECIMENTO (EM MESES) E TIPO DE BARRICA PARA MATURAÇÃO DO VINHO	110
	APÊNDICE E – AROMAS POSSÍVEIS NO ATO DA DEGUSTAÇÃO DO VINHO	111
	APÊNDICE F – RESULTADO DA APLICAÇÃO DO QUESTIONÁRIO DE VALIDAÇÃO DA PROPOSTA	112

1 INTRODUÇÃO

Desde a “domesticação da uva” pelos povos que viviam na região da atual Geórgia, Europa, entre 7000-5000 a.C, a humanidade desenvolveu uma relação próxima com o vinho, originado a partir da fermentação da fruta (DUARTE, 2016). Por garantir uma bebida relativamente livre de microrganismos nocivos, o vinho foi a fonte principal de hidratação para os habitantes de regiões produtoras durante tempos antigos, nos quais a contaminação por água era um problema corriqueiro (CURRY, 2017).

Além das suas características químicas, o consumo de vinho foi estimulado e difundido ao longo de diferentes regiões também graças ao seu efeito inebriante, apesar de originalmente não ser uma bebida agradável ao paladar. Buscando dar ao vinho um sabor mais apazível, as formas rudimentares utilizadas na produção original deram lugar a técnicas aprimoradas que se diferenciavam entre si, a partir de variações no tipo da uva utilizada, tempo de maturação, adição de especiarias e/ou açúcares (PINTO, 2009). Essas variações nos processos produtivos culminaram no surgimento dos diversos tipos de vinho que até hoje são presentes na mesa do consumidor.

A relação gastronômica entre as variedades de vinhos existentes no mercado com os diversos tipos de alimentos pode ser considerada de forma harmônica, contraposição ou não harmônica. A Enogastronomia dedica-se a estudar estas interações, com base nos critérios técnicos de avaliação e as tradições culturais associadas à bebida e alimento (QUEIROZ, 2009). Desta forma, é possível estabelecer infinitas possibilidades de interpretação quando se realiza uma harmonização, visto que o gosto pessoal é um dos fatores que influenciam na escolha da combinação, porém existem regras simples e básicas que são levadas em conta, como a combinação de peixes com vinho branco.

Desta forma, esta pesquisa tem como propósito coletar informações essenciais disponíveis na literatura pertinente para realizar uma harmonização equilibrada entre comida e vinho, a fim de construir uma base de dados Enogastronômica e, posteriormente realizar a análise estatística e mineração de dados com o intuito de investigar a possível existência de padrões e regras para se realizar uma harmonização.

1.1 PROBLEMATIZAÇÃO

A gastronomia possui variações entre as diferentes culturas espalhadas no mundo, e o Brasil em especial, apresenta diversas culturas gastronômicas. Pelo Brasil existem diversos roteiros gastronômicos que incluem os pratos típicos com visitas a fazendas e vinhedos, e pratos regionais estes acompanhados por música e artesanatos de cada região (MONTECINOS *et. al.*, 2012). A gastronomia pertence a riqueza cultural de regiões e abre oportunidades para conhecer novos costumes e hábitos a partir do paladar.

Um fato pouco explorado, está relacionado a combinação entre sabores e incrementos que agregam valor à gastronomia, como por exemplo, a harmonização com vinhos. Harmonizar pratos e vinhos exige um alto grau de conhecimento entre cores, acidez, sabores, variedades, aromas e texturas para que seja possível compor combinações agradáveis. Esta relação é conhecida como “Enogastronomia”.

Com a globalização a arte da cozinha evoluiu, onde surgiram novos equipamentos, tendências, técnicas, regras e normas, acompanhamento nutricional e demais fatores, que agregam valor ao mercado gastronômico (MONTECINOS *et. al.*, 2012). Neste sentido a Enogastronomia ainda não é popularizada, mas está a caminho de se popularizar, já que a apreciação do vinho vem se tornando mais expressiva dentro do mercado gastronômico.

Cada literatura engloba fatores diferenciados e expressam suas harmonizações de forma simplória ou de caráter bem específico, onde não são colocados explicitamente a existência de padrões pré-estabelecidos sobre como realizar uma harmonização gastronômica equilibrada. Segundo Rocha (2009), a Enogastronomia não deve apenas tratar a harmonização entre e vinho como algo constante, já que a experiência Enogastrônômica envolve o gosto pessoal de cada indivíduo.

Desta forma, a questão de pesquisa deste estudo consiste em: **Quais as potencialidades da análise de dados (estatística e mineração de dados) para descoberta de relações e padrões em uma base de dados de Enogastronomia?**

1.2 OBJETIVOS

A fim de responder o problema de pesquisa deste estudo, foram definidos os objetivos a serem alcançados com este projeto. Estes foram desmembrados em objetivos geral e específicos.

1.2.1 Objetivo Geral

O objetivo geral deste estudo consiste em realizar uma análise de dados a partir da aplicação de técnicas de estatística e mineração de dados, a fim de identificar relações e padrões de ocorrência em uma base de dados Enogastronômica.

1.2.2 Objetivos Específicos

Os objetivos específicos definidos para alcançar o objetivo geral deste projeto são:

- a) construir uma base de dados que reúna elementos da Gastronomia e Enologia;
- b) comparar os resultados obtidos a partir de dois métodos de mineração de dados;
- c) realizar a análise descritiva da base de dados e analisar os resultados obtidos.

1.3 JUSTIFICATIVAS

Com o propósito de fundamentar os motivos para o desenvolvimento desta pesquisa, subdividiu-se em 3 (três) aspectos do saber, sendo eles: científica, acadêmica e pessoal.

1.3.1 Justificativa para a Ciência

No dia 7 de abril de 2018, foi realizado um levantamento no Sistema de Bibliotecas (SiBi) Portal da Capes, e nas bases *Web of Science* (WOS), *Scopus* e *Scientific Eletronic Library Online* (SciELO), com o intuito de identificar a produção científica com o tema que compõe este trabalho.

Em primeiro momento, houve a aplicação direta ao Sistema de Bibliotecas com a aplicação dos termos em inglês. Ao total foram aplicadas oito buscas, considerando todos os índices da base. Os resultados desta aplicação estão descritos no quadro 1.

QUADRO 1 – NÚMERO DE RESULTADOS DE BUSCA NO SISTEMA DE BIBLIOTECAS DA UFPR

BUSCA	TERMOS	TOTAL DE REGISTROS
1	<i>"Data Analysis"</i>	47.158.001
2	"Análise de Dados"	517.964
3	<i>"Data Mining"</i>	186.639
4	"Mineração de dados"	9.536
5	<i>"Enogastronomic"</i>	104
6	"Enogastronomia"	103
7	<i>"Data Analysis" AND "Enogastronomic"</i>	4
8	<i>"Data Mining" AND "Enogastronomic"</i>	0

FONTE: A autora (2018).

Também foram realizadas buscas no repositório do Sistema de Bibliotecas (SiBi) da UFPR, voltado aos trabalhos de graduação na área de turismo, visto que é o ambiente que investiga a Gastronomia nas diferentes regiões, a fim de verificar a existência de aplicações estatísticas e de mineração de dados aplicadas a Enogastronomia, porém não foram encontrados nenhum resultado.

Posteriormente, realizou-se a aplicação dos termos no Portal da Capes. Os resultados referentes a esta aplicação estão apresentados no quadro 2.

QUADRO 2 – NÚMERO DE RESULTADOS DE BUSCA NO PORTAL DA CAPES

BUSCA	TERMOS	TOTAL DE REGISTROS
1	"Data Analysis"	12.661.516
2	"Data Mining"	521.466
3	"Análise de Dados"	42.283
4	"Mineração de Dados"	809
5	"Enogastronomia"	14
	"Enogastronomic"	13
7	"Data Analysis" AND "Enogastronomic"	5
8	"Data Mining" AND "Enogastronomic"	0

FONTE: A autora (2018).

Seguidamente, realizou-se uma pesquisa em inglês nas bases de dados *Web Of Science* e *Scopus*, considerando o volume de acervo disponível. Os resultados obtidos na base *Web Of Science* são representados no quadro 3.

QUADRO 3 – NÚMERO DE RESULTADOS DE BUSCA NA BASE *WEB OF SCIENCE*

BUSCA	TERMOS	TOTAL DE REGISTROS
1	"Data Mining"	53.093
2	"Enogastronomic"	5
3	"Data Mining" AND "Enogastronomic"	0

FONTE: A autora (2018).

Já os resultados obtidos na base *Scopus*, é apresentado no quadro 4.

QUADRO 4 – NÚMERO DE RESULTADOS DE BUSCA NA BASE SCOPUS

BUSCA	TERMOS	TOTAL DE REGISTROS
1	"Data Mining"	130.551
2	"Enogastronomic"	2
3	"Data Mining" AND "Enogastronomic"	0

FONTE: A autora (2018).

Foi possível observar que nestas bases o volume de estudos relacionados à "Mineração de Dados" é vasto, enquanto o termo "Enogastronomia" é pouco explorado. Percebe-se também a não ocorrência de estudos que possuem relação entre Mineração de Dados e Enogastronomia.

Outra pesquisa foi realizada na base *Scientific Eletronic Library Online* (SciELO) com o intuito de verificar a ocorrência de publicações nacionais relacionadas ao tema de pesquisa. Os resultados são descritos no quadro 5.

QUADRO 5 – NÚMERO DE RESULTADOS DE BUSCA DA BASE SCIELO

BUSCA	TERMOS	TOTAL DE REGISTROS
1	"Análise de Dados"	24.742
2	"Mineração de dados"	182
3	"Enogastronomia"	0
4	"Mineração de dados" e "Enogastronomia"	0

FONTE: A autora (2018).

Percebe-se que a ocorrência do termo "Mineração de dados" em publicações nacionais é mínima perante as demais bases, enquanto não houve nenhuma ocorrência de estudos voltados à Enogastronomia.

Vale ressaltar que em nenhuma das bases pesquisadas se obteve resultados relacionados ao cruzamento dos termos "Mineração de Dados" e "Enogastronomia", indicando uma escassez de estudos na área da pesquisa que está sendo desenvolvida, reforçando a relevância da contribuição do desenvolvimento deste estudo para a ciência.

1.3.2 Justificativa para a Área da Gestão da Informação

No dia 7 de abril de 2018, realizou-se uma pesquisa nos repositórios das 4 (quatro) universidades que ofertam o curso de Gestão da Informação no Brasil, sendo elas: Universidade Federal do Paraná (UFPR); Universidade Federal de Uberlândia (UFU); Universidade Federal de Pernambuco (UFPE); e Universidade Federal de Goiás (UFG). O intuito desta busca envolveu identificar trabalhos com a temática deste estudo. O quadro 6 representa os resultados colhidos a partir da pesquisa dos termos *Data Mining*, Mineração de Dados e Enogastronomia empregados a área da Gestão da Informação.

QUADRO 6 – REPRESENTAÇÃO DOS ESTUDOS COM TEMAS RELACIONADOS À *DATA MINING*, MINERAÇÃO DE DADOS E ENOGASTRONOMIA

UNIVERSIDADE	<i>Data Mining</i>	Mineração de Dados	Enogastronomia
UFPR	7	10	0
UFU	-	-	-
UFPE	0	6	0
UFG	0	0	0

FONTE: A autora (2018).

Dos resultados das pesquisas realizadas, destaca-se:

- o repositório da Universidade Federal de Uberlândia (UFU) está em construção, assim, não responde a nenhum tipo de pesquisa;
- o resultado trazido da Universidade Federal de Pernambuco (UFPE) é resultante do Departamento de Artes e Comunicação;
- os estudos relacionados aos temas “*Data Mining*” e “Mineração de Dados” na Universidade Federal de Goiás (UFG) não pertencem ao curso de Gestão da Informação, mas a área de Sistemas de Informação.

Considerando a base que concentra as monografias já desenvolvidas no curso de Gestão da Informação ofertado pela Universidade Federal do Paraná, local onde este estudo está sendo desenvolvido, utilizou-se termos como: “*Data Mining*”, onde possui sete estudos documentados; “Mineração de Dados” com dez estudos

documentados; e o termo “Enogastronomia” nenhum resultado foi obtido, além disso, nenhum possuía temática semelhante ao tema proposto.

1.3.3 Justificativa Pessoal

A proposta do tema se deu como forma de aplicar técnicas estatística e de mineração de dados aplicados a Enogastronomia, visto que é uma área pouco explorada. Com esta aplicação é possível aliar duas áreas de interesse pessoal de atuação futura. Desta forma, notou-se a falta de trabalhos relacionados a temática de estudo, trazendo o interesse em aplicar técnicas e ferramentas de mineração de dados e análise descritiva, a fim de descobrir regras e padrões quanto a harmonização de vinhos e gastronomia.

1.4 DELIMITAÇÃO DA PESQUISA

Inicialmente buscou-se em repositórios digitais base de dados que contemplassem dados relacionados à Enogastronomia, porém não foi encontrada nenhuma base que combinasse ingredientes e propriedades de vinho, ou ainda, pratos e bebidas de forma geral. Foi encontrada uma única base de dados que relaciona as propriedades com a qualidade dos vinhos. Então, decidiu-se pela concepção de uma base de dados Enogastrônômica, limitando-se a partir das pizzarias de Curitiba premiadas nos concursos Comer&Beber 2017/2018 ofertado pela revista Veja e Bom Gourmet 2018 oferecido pelo Jornal Gazeta do Povo que possuíssem adegas e profissionais Enólogo¹ e/ou *Sommeliers*². Os rótulos de vinhos participantes do estudo estão de acordo com a carta de vinhos que a pizzaria trabalha, bem como, a sugestão de harmonização definida pelo profissional responsável pela adega.

¹ O enólogo é o profissional que compreende as características do vinho, sendo responsável pela elaboração dos vinhos.

² O *sommelier* é o profissional que desempenha funções em restaurantes; ele é capaz de propor sugestões aos clientes quanto ao tipo de bebida (ISTITUTO EUROPEO DI DESIGN, 2011).

1.5 ESTRUTURA DO DOCUMENTO

Este documento está dividido em 5 (cinco) seções. A primeira seção corresponde à introdução do estudo que está sendo realizado, sendo apresentado o problema de pesquisa, os objetivos traçados, justificativa de sua realização e a delimitação do escopo de pesquisa.

A segunda seção engloba o referencial teórico pertinente ao assunto que está sendo abordado, onde se apresenta a Gestão da Informação, Recuperação da Informação, Estatística, Estatística Descritiva, Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Database – KDD*), Mineração de Dados e Enogastronomia.

Na terceira seção são apresentados os encaminhamentos metodológicos, onde se descreve a caracterização de pesquisa e os materiais e métodos que foram utilizados.

Na quarta seção, estão descritos os resultados e análises obtidos, envolvendo a base de dados e a análise estatística da mesma e a aplicação dos algoritmos de mineração de dados *Apriori* e *Prism*.

Por fim, a quinta seção apresenta as considerações finais, detalhando o alcance dos objetivos e os potenciais trabalhos futuros.

2 REFERENCIAL TEÓRICO

Nesta seção é apresentada a fundamentação teórica, cujo objetivo é conduzir a pesquisa de acordo com a problemática investigada e os objetivos traçados. Desta forma, são abordados como temas: Gestão da Informação, Recuperação da Informação, Estatística, Estatística Descritiva, *Knowledge Discovery in Database* (KDD), Mineração de Dados e Enogastronomia.

2.1 GESTÃO DA INFORMAÇÃO

Antes de tratar da Gestão da Informação (GI) em si, é necessário diferenciar dado, informação e conhecimento. De acordo com Moreira (2015), dados são fatos em sua forma primária, sendo que quando estes dados são organizados de forma significativa, estes se transformam em informação. Neste sentido, a informação pode ser considerada uma propriedade interior de um indivíduo ou recebido por este, onde pode estar inserida em sua esfera mental ou ser recebida por meio da representação simbólica, como textos, figuras, animações etc. (SETZER, 1999). Já o conhecimento, segundo Barreto (1999, p.3) “acontece em cada indivíduo em sua individualidade, manipulando sua sensibilidade e percepção no trato com a informação”.

Para Valentim (2008), a informação é considerada um objeto, pois quando de forma explícita em algum suporte, conseguimos visualizar o conhecimento construído. Além disso, pode ser considerada como um fenômeno, pois ocorre a partir de um processo cognitivo variável para se obter um relatório, visto as etapas de relacionar, isolar, associar, analisar e sintetizar.

Davenport (1998), em seu livro denominado “Ecologia da Informação”, define a Gestão da informação (GI) como um conjunto estruturado de atividades que incluem o modo com que se obtêm, distribuem e utilizam a informação e o conhecimento. Além desta contribuição, Davenport (1998) realiza a distinção entre dados, informação e conhecimento, como descrito no quadro 7.

QUADRO 7 – DISTINÇÃO ENTRE DADOS, INFORMAÇÃO E CONHECIMENTO

DADOS	INFORMAÇÃO	CONHECIMENTO
Simples observações sobre o estado do mundo <ul style="list-style-type: none"> - Facilmente estruturado - Facilmente obtido por máquina - Frequentemente quantificado - Facilmente transferível 	Dados dotados de relevância e propósito <ul style="list-style-type: none"> - Requer unidade de análise - Exige consenso em relação ao significado - Exige necessariamente a medição humana 	Informação valiosa na mente humana <ul style="list-style-type: none"> - Inclui reflexão, síntese, contexto - De difícil estruturação - De difícil captura em máquinas - Frequentemente tácito - De difícil transferência

FONTE: Davenport (1998, p.18).

Neste contexto, McGee e Prusak (1994) criaram um modelo de processo de gerenciamento de informações, focado no valor estratégico de informações, onde este é definido em 3 (três) tarefas, sendo elas: identificação de necessidades e requisitos de informação; classificação, armazenamento, tratamento e apresentação da informação; e, o desenvolvimento de produtos e serviços de informação.

Davenport (1998), explica que o modelo de gestão é composto por 3 (três) ambientes, sendo eles: externo, organizacional e informacional. O ambiente externo é considerado os mercados sejam de negócios, tecnológicos ou de informação; enquanto o ambiente organizacional é composto por negócios, espaço físico e tecnologia; por fim, o ambiente informacional é constituído por componentes críticos, como estratégia, política, cultura, *staff*, processos e arquitetura. A figura 1 apresenta o modelo descrito pelo autor.

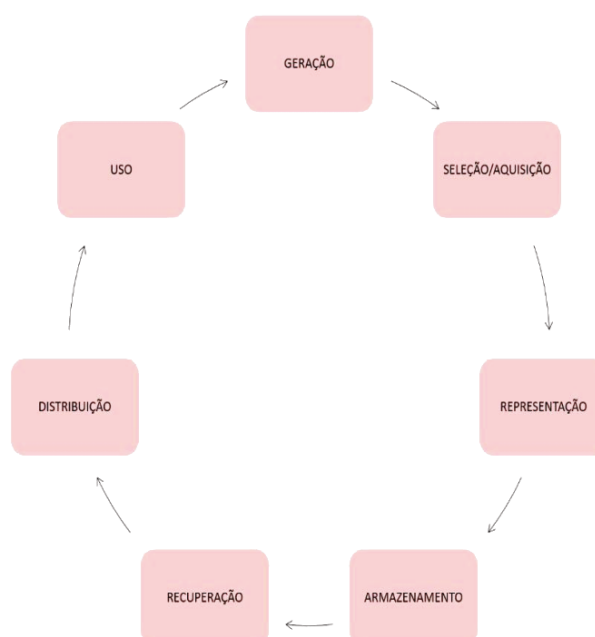
FIGURA 1 – MODELO ECOLÓGICO PARA O GERENCIAMENTO DE INFORMAÇÃO



FONTE: Davenport (1998, p.51).

Ponjuán Dante (1998), afirma que a gestão da informação deve incluir dimensões estratégicas e operacionais, bem como, os mecanismos para a obtenção e utilização dos recursos humanos, tecnológicos, financeiros, materiais e físicos para o gerenciamento de informações, tal que se transforme em insumo útil e estratégico para indivíduos, grupos e organizações. Deste modo, a figura 2 apresenta a aplicação do ciclo da informação.

FIGURA 2 – CICLO INFORMACIONAL



FONTE: Ponjuán Dante (1998, p.47).

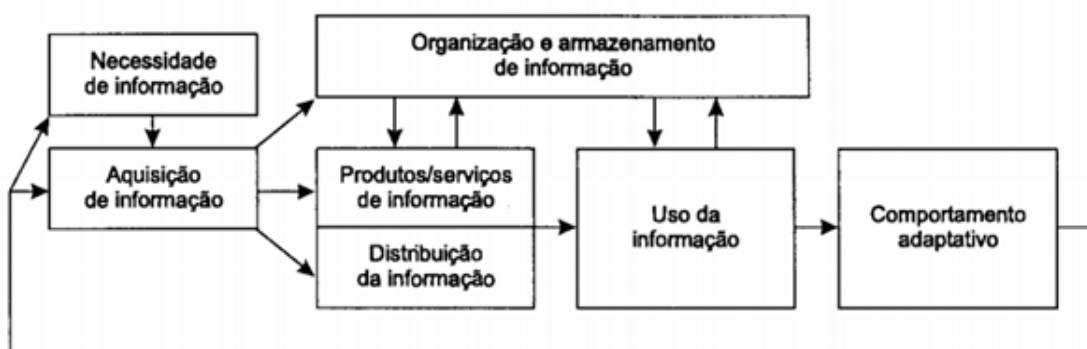
Pode-se considerar sobre o ciclo de informação:

O ciclo informacional é iniciado quando se detecta uma necessidade informacional, um problema a ser resolvido, uma área ou assunto a ser analisado. É um processo que se inicia com a busca da solução a um problema, da necessidade de obter informações sobre algo, e passa pela identificação de quem gera o tipo de informação necessária, as fontes e o acesso, a seleção e aquisição, registro, representação, recuperação, análise e disseminação da informação, que, quando usada, aumenta o conhecimento individual e coletivo (TARAPANOFF, 2006, p. 23).

Para Miranda e Streit (2007), a gestão da informação engloba a fase de processos e práticas fragmentadas, onde o mapeamento de processos auxilia na organização do ambiente de gestão, sendo este, baseado no ciclo de vida da informação. Marchiori (2002), explica a grande expectativa do mundo interconectado, onde é necessário utilizar de padrões e procedimentos para que os conteúdos informativos sejam estruturados. Posteriormente, a autora aborda como a gestão da informação é vista sob o enfoque da tecnologia, considerando-a um recurso otimizado, visto que se utilizam arquiteturas de *hardware*, *software* e redes de comunicação inserida em diferentes sistemas de informação (MARCHIORI, 2002).

Choo (2003) traz outra abordagem relacionada à gestão da informação, onde para que ela se torne estratégica é necessário transformá-la em conhecimento. Para que a gestão da informação seja aplicada de forma efetiva, o autor propõe a GI como um conjunto de relação em 6 (seis) processos, sendo eles: identificação das necessidades de informação; aquisição da informação; organização e armazenamento da informação; desenvolvimento de produtos e serviços de informação; distribuição da informação e o uso da informação. Este processo é representado na figura 3.

FIGURA 3 – MODELO PROCESSUAL DA GESTÃO DA INFORMAÇÃO



FONTE: Choo (2003, p.404).

A etapa de identificação das necessidades de informação, surgem a partir de problemas encontrados em determinadas situações, sejam estes causados por incertezas ou ambiguidade. Desta maneira, devem-se observar os padrões e regras de uso, bem como, as condições propostas para que a informação se torne significativa às necessidades do usuário (CHOO, 2003).

Segundo Choo (2003), a aquisição da informação engloba 2 (duas) demandas opostas. A primeira se dá a partir da variedade de informações disponíveis, enquanto a outra acontece a partir da capacidade cognitiva humana para organizar e compreender a mensagem. O autor enfatiza que a seleção e uso das fontes de informação devem ser planejados, monitorado e avaliado continuamente para que se torne um recurso vital dentro da organização (CHOO, 2003).

A etapa seguinte é a organização e armazenamento da informação, onde Choo (2003) ressalta que a maneira como a informação é armazenada, bancos de dados e sistemas de informação influenciam no processo de recuperação da informação. Considerando que este processo é utilizado para a frequente realização de consultas da memória da organização, visando a resolução de problemas e a interpretação de soluções.

O desenvolvimento de produtos e serviços de informação consiste em garantir que as necessidades informacionais dos usuários sejam atendidas, de modo, a responder todas as suas perguntas e promover possíveis soluções aos problemas, proporcionando auxílio aos membros da organização no processo de tomada de decisão (CHOO, 2003).

De acordo com Choo (2003), a fase de distribuição da informação compreende no processo de disseminação da informação para os usuários, de forma que a informação correta chegue para a pessoa certa, no momento certo e no formato adequado. Vale ressaltar, que o propósito desta etapa é facilitar o compartilhamento de informações, sendo essencial para a geração de significado, construção do conhecimento e auxiliar o processo de tomada de decisão.

Por fim, Choo (2003) aborda o uso da informação, sendo considerado um processo que resulta na geração de significado; construção de conhecimento em tácito, explícito e cultural; e descoberta de padrões de ações. Este é considerado um processo repetitivo e recíproco.

2.1.1 Recuperação da Informação

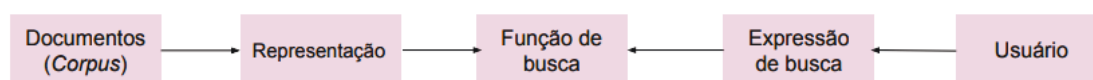
Para Ferneda (2003), o termo denominado Recuperação da Informação (RI) no âmbito da ciência da informação, para alguns pode significar a operação de documentos a partir do acervo em função de uma demanda. Enquanto para outros, este consiste no fornecimento a partir de uma demanda já definida pelos usuários.

Lancaster (1978 *apud* FIGUEIREIDO, 2006, p.15), trata a denominação da recuperação da informação como um sinônimo de busca, sendo considerado o processo para buscar um conjunto de documentos. Baeza-Yates e Ribeiro-Neto (2013), partem do princípio que a Recuperação da Informação engloba a ciência da computação, sendo capaz de providenciar informações aos usuários de forma mais simplificada de acordo com seu interesse. Baeza-Yates e Ribeiro-Neto (2013), descrevem a RI como:

A Recuperação de Informação (RI) trata da recuperação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens devem fornecer aos usuários facilidade de acesso às informações de seu interesse. (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 1).

Neste contexto, Ferneda (2003) explica que o processo de recuperação da informação compõe-se em identificar em um conjunto de documentos, quais informações atendem as necessidades do usuário. Ressalta que este processo busca a recuperação de informações e não apenas recuperar dados que satisfazem a expressão de busca, sintetizando que é neste sentido que a recuperação da informação se difere de um banco de dados. Desta forma, o autor aborda que os sistemas de recuperação da informação devem representar o conteúdo dos documentos, transmitindo de forma rápida as informações que satisfazem a necessidade do usuário através da expressão de busca. A figura 4 aborda a representação do processo de recuperação da informação proposta pelo autor.

FIGURA 4 – REPRESENTAÇÃO DO PROCESSO DE RECUPERAÇÃO DA INFORMAÇÃO (RI)



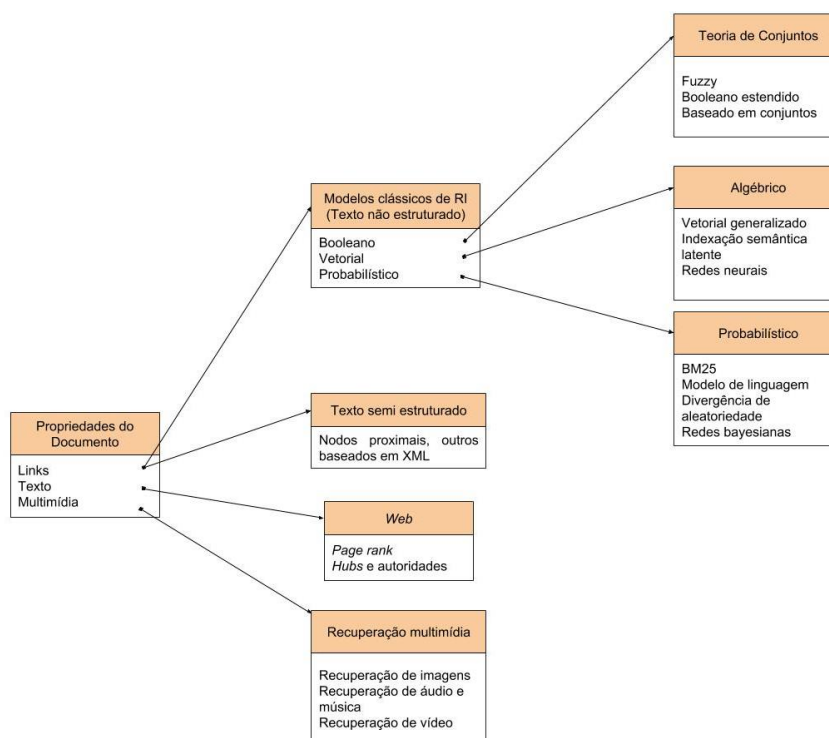
FONTE: Ferneda (2003, p.15).

Em sua publicação, Ferneda (2003) descreve todas as etapas do processo de recuperação da informação, ressaltando as mudanças de documentos provenientes da chegada dos ambientes digitais, onde na etapa de representação se busca identificar um documento através de seu conteúdo, sendo realizado por um processo de indexação. A fase função de busca se dá a partir da utilização de “filtros” para normalizar os termos, eliminando as palavras de significação e reduzindo o uso de seus radicais. Já a fase de expressão da busca, pode ocorrer tanto em linguagem natural quanto verbal, onde está deve resultar em um volume de documentos que possibilite a averiguação deles a fim de selecionar os úteis. E o usuário deve traduzir sua necessidade ao sistema informação, utilizando a linguagem adequada ao sistema (FERNEDA, 2003, p.14-17).

De acordo com Baeza-Yates e Ribeiro-Neto (2013), o usuário de um sistema de recuperação da informação precisa traduzir suas necessidades de acordo com a linguagem utilizada pelo sistema, além disso, os autores citam sobre os termos de pesquisa voltados a recuperação de documentos, pois a indexação é considerada um processo chave que exerce o papel de substantivo.

Para Baeza-Yates e Ribeiro-Neto (2013), a indexação torna o processo de recuperação da informação mais ágil e eficiente. Segundo eles, os sistemas de recuperação da informação, consiste em recuperar os elementos relevantes ao usuário, otimizando o tempo, minimizando a recuperação de documentos considerados irrelevantes (BAEZA-YATES; RIBEIRO-NETO, 2013). A figura 5 representa a taxonomia utilizada pelos autores para classificar os diferentes modelos de RI.

FIGURA 5 – TAXONOMIA DOS MODELOS DE RI



FONTE: Baeza-Yates e Ribeiro-Neto (2013, p. 24).

A partir da figura 5, os autores apresentam 3 (três) categorias de documentos, sendo eles: modelos organizados em textos, em *links* e em objetos multimídias.

Assim, os modelos organizados por texto se dividem em semiestruturados e não estruturados. Neste sentido, aos textos considerados semiestruturados, que levam a estrutura do texto e componentes como títulos, seções e parágrafos, além de utilizar técnicas de indexação baseados em XML. Enquanto para os textos considerados não estruturados se utiliza o modelo booleano, vetorial e probabilístico (BAEZA-YATES; RIBEIRO-NETO, 2013).

Já na *Web*, com a variedade de páginas existentes utilizar somente o ranqueamento a partir de textos não é viável, então se faz necessário utilizar o ranqueamento a partir de *links*, tal como a *PageRank* que simula o acesso de um usuário de forma aleatória na *Web*, bem como, o acesso à *hiperlinks*, entradas e saídas das páginas e a probabilidade de acesso (BAEZA-YATES; RIBEIRO-NETO, 2013).

A última categoria da taxonomia proposta por Baeza-Yates e Ribeiro-Neto (2013) está relacionada aos objetos em multimídia (imagens, vídeos e músicas),

considerando que esta necessita de uma estratégia diferenciada, visto que a informação não é totalmente relacionada a qualquer interpretação do usuário.

2.2 ESTATÍSTICA

Para Shiguti e Shiguti (2006), a estatística é aplicada a uma coleção de métodos para planejar experimentos, voltados para a obtenção de dados, organização, análise, interpretação e extração de conclusões significativas. Santos (2009) afirma que a estatística pode ser considerada a ciência de aprendizagem a partir de dados, considerando que é capaz de fornecer métodos que auxiliam no processo de tomada de decisão e está presente em todas as áreas da ciência que envolvam a coleta e análise de dados.

Granzotto (2002) define o método estatístico como:

Um método científico que diante da impossibilidade de manter as causas constantes (nas ciências sociais), admitem todas essas causas presentes variando-as, registrando essas variações e procurando determinar, no resultado final, que influências cabem a cada uma delas (GRANZOTTO, 2002, p.02).

São definidos como conceitos básicos da estatística segundo Shiguti e Shiguti (2006):

- **população:** conjunto de objetos ou indivíduos que apresentam no mínimo uma característica em comum;
- **censo:** coleção de dados relativos a todos elementos da população;
- **amostra:** parte representativa da população;
- **estatística:** medida numérica que expressa as características da amostra;
- **parâmetro:** medida numérica que expressa as características da população (SHIGUTI; SHIGUTI, 2006).

É possível complementar a lista de conceitos básicos com os princípios trazidos por Granzotto (2002):

- **dado estatístico:** dado numérico considerado a matéria prima para aplicação dos métodos estatísticos;
- **estimativa:** valor aproximado do parâmetro, calculado a partir da amostra;
- **atributo:** levantamento e tratamento de dados em aplicações qualitativas;
- **variável:** conjunto de resultados possíveis de um fenômeno (Granzotto, 2002).

As variáveis de interesse da estatística podem ser classificadas em qualitativas e quantitativas. As qualitativas são dadas a partir de uma classificação por tipos ou atributos; enquanto as quantitativas são expressas por números, sendo subdividas em discretas (obtidas a partir de alguma forma de contagem) e contínuas (assumem qualquer valor em um intervalo de variação) (BAILEIRO, 2008).

2.2.1 Estatística Descritiva

Para Hout (2002, p.60) a estatística descritiva é conceituada como “conjunto das técnicas e das regras que resumem a informação recolhida sobre uma amostra ou uma população, e isso sem distorção nem perda de informação”.

A estatística descritiva envolve a organização e sumarização de dados a partir de metodologias simples (SHIGUTI; SHIGUTI, 2006). A figura 6, demonstra a partir de um diagrama como ocorre o processo de aplicação da estatística descritiva.

FIGURA 6 – PROCESSO DA ESTATÍSTICA DESCRITIVA



FONTE: Shiguti e Shiguti (2006, p.06).

Dentre as etapas deste processo estão:

- **coleta de dados:** compilação dos dados das variáveis a ser estudada;
- **crítica dos dados:** revisão dos dados a fim de suprimir os valores desconhecidos ao levantamento, a fim de não prejudicar os resultados com enganos futuros;
- **apresentação dos dados:** organização do conjunto de dados de maneira prática e racional, expressa por meio de tabelas ou gráficos;
- **análise:** etapa de interpretação dos resultados obtidos (SHIGUTI; SHIGUTI, 2006).

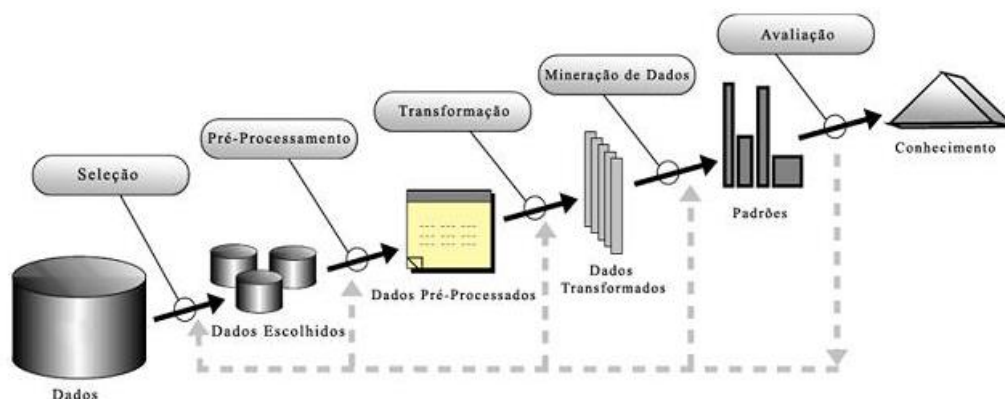
A aplicação da estatística descritiva tem como objetivo resumir as principais características de um conjunto de dados a partir da utilização de tabelas, gráficos e resumos numéricos (GUIMARÃES, 2007). É papel da estatística definir qual é a forma mais adequada de representar o conjunto de dados, sendo eles:

- **tabelas de frequência:** apropriado para resumir grande conjunto de dados, agrupando informações em categorias (pontuais ou intervalos);
- **gráficos:** possibilita uma visualização das principais características da amostra;
- **medidas descritivas:** uso de resumos numéricos para levantar informações importantes sobre o conjunto de dados (tendência, simetria, variabilidade, discrepância, etc) (GUIMARÃES, 2007).

2.3 KDD (*KNOWLEDGE DISCOVERY IN DATABASES*)

O KDD (*Knowledge Discovery in Databases*), significa em português a Descoberta de Conhecimento em Base de Dados. Neste contexto, o KDD é definido por Fayyad et al. (1996, p.11), como “o processo, não trivial de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados no banco de dados”. Consiste na descoberta de padrões de análise de grandes conjuntos de dados, cuja sua principal função é a etapa de mineração de dados, pois possibilita a execução prática de análise e permite a utilização de algoritmos específicos que sob limitações de eficiência, produzem uma relação específica de padrões de dados (FAYYAD et. al. 1996).

O processo de KDD é de natureza interdisciplinar, sendo capaz de agregar conceitos e técnicas de áreas como aprendizagem de máquinas, reconhecimento de padrões e banco de dados (FAYYAD et. al. 1996). Assim, Fayyad *et al.* (1996) definem 5 (cinco) etapas para a execução deste processo, sendo elas: seleção, pré-processamento, transformação, mineração de dados e avaliação dos resultados, como demonstra a figura 7.

FIGURA 7 – ETAPAS DO KDD (*Knowledge Discovery in Database*)

FONTE: Fayyad et. al. (1996, p.4).

As etapas do KDD podem ser descritas como:

1. **seleção de dados:** fase onde é determinado o conjunto de dados contendo todas as variáveis e registros que irão compor a análise;
2. **pré-processamento:** realiza-se o tratamento dos dados selecionados, além de efetuar a limpeza dos dados considerando a remoção de ruídos, garantindo a qualidade destes;
3. **transformação:** etapa onde ocorre a transformação dos dados de acordo com o propósito da aplicação, além disso, esta fase também dispõe dos processos de armazenamento e organização para que os algoritmos sejam aplicados adequadamente.
4. **mineração de dados:** são aplicadas as tarefas adequadas de acordo com os algoritmos que foram definidos, bem como, a parametrização dos mesmos considerando o conjunto de dados;
5. **avaliação dos resultados:** Interpretação dos padrões obtidos. (FAYYAD et. al. 1996).

Para Romão (2002), a viabilidade da aplicação de KDD se dá a partir de aspectos práticos e técnicos. Os aspectos práticos englobam reflexões sobre o impacto que a aplicação irá provocar a partir de critérios como, rendimento, redução de custos, qualidade de produtos e tempo. Já os aspectos técnicos competem à disponibilidade de dados, onde a complexidade do problema se utiliza mais atributos. A partir do domínio da aplicação, ambos os aspectos são capazes de contribuir para

a eficiência na busca da tarefa de mineração de dados, bem como, nas demais etapas do processo de KDD (ROMÃO, 2002).

2.3.1 Mineração de Dados

Segundo Castro e Ferrari (2016), a Mineração de Dados é considerada um elemento central à parte analítica do *Big Data*, visto que o processo de mineração explora dados utilizando algoritmos como ferramentas. Rokach e Maimon (2008) referem-se à mineração de dados como sendo a ciência, arte e tecnologia para explorar conjuntos de dados grandes e complexos a fim de descobrir padrões úteis. Rokach e Maimon (2008) ainda descrevem que a Mineração de Dados se deu a partir da necessidade em analisar o crescimento exponencial de dados e informações.

A Mineração de Dados para Fayyad *et al.* (1996), é uma etapa do KDD que permite aplicar algoritmos de análise e descoberta de dados a partir de um detalhamento específico de padrões. Para os autores, a mineração de dados possui duas categorias principais, a predição e a descrição. A predição engloba o uso de variáveis do banco de dados a fim de prever valores desconhecidos ou futuros de outras variáveis, enquanto a descrição busca a descoberta de padrões que descrevam os dados e que sejam possíveis de interpretação humana (FAYYAD *et. al.* 1996).

Neste contexto, Boente, Goldschmidt e Estrela (2008), definem tarefas de mineração de dados para suprir as necessidades de cada “tipo” de dados. Estas tarefas são denominadas: descoberta de associação; classificação; regressão; *clusterização*; sumarização; detecção de desvios e descoberta de sequências (BOENTE; GOLDSCHMIDT; ESTRELA, 2008).

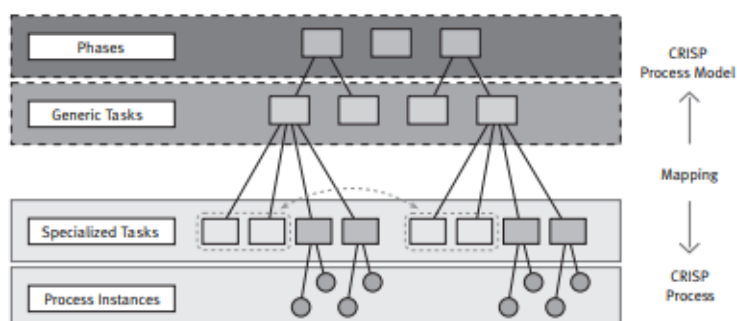
Na etapa de descoberta de associação, se busca por pontos que ocorram de forma simultânea em transações do banco de dados, podendo utilizar como ferramenta o algoritmo *Apriori*; já na etapa de classificação, é a fase para descobrir uma função que seja capaz de mapear o conjunto de dados em um conjunto de classes, a fim de prever em qual classe determinado registro se enquadra, podendo utilizar Redes Neurais e Algoritmos Genéticos; a fase regressão compreende na busca por uma função que mapeie estes registros em um valor real, ou seja, restrita

apenas a atributos numéricos, sendo possível utilizar ferramentas estatísticas para a realização desta etapa; a tarefa de *clusterização* separa os registros em subconjuntos, a fim de identificar propriedades comuns que distingam dos demais subconjuntos, maximizando a similaridade entre os dados, podendo utilizar o *K-means* para esta fase; a sumarização identifica semelhanças entre o conjunto de dados, podendo ser realizada a partir da lógica indutiva; a partir da detecção de desvios é possível identificar quais registros do banco de dados não atendem os padrões normais para a realização do processo de mineração, estes são conhecidos como *outliers* e a estatística fornece os recursos necessários para a execução desta tarefa; por fim, a descoberta de sequências é uma extensão da descoberta de associações, onde se busca os itens frequentes nas transações ao longo do período (BOENTE; GOLDSCHMIDT; ESTRELA, 2008).

Fayyad et al. (1996), incluem neste conjunto de tarefas a etapa de pós-processamento que engloba o tratamento do conhecimento conquistado a partir da aplicação da mineração de dados, cujo objetivo é facilitar a interpretação e avaliação do conhecimento descoberto.

A partir destas etapas Chapman et al. (2000), definem a metodologia CRISP-DM, sendo um modelo de processo hierárquico dividido em 4 (quatro) níveis que vão do geral ao mais específico, sendo eles: superior, genérico, especializado e instância do processo. O processo da metodologia CRISP-DM é apresentado na figura 8.

FIGURA 8 – DIVISÃO DE NÍVEIS DA METODOLOGIA CRISP-DM



FONTE: Chapman et al. (2000, p.6).

O nível superior envolve o processo de mineração de dados, sendo dividido em diversas etapas, onde cada etapa consiste em subáreas de tarefas genéricas; já o nível genérico tem como objetivo, cobrir todos os dados possíveis que ocorrem no

processo de mineração de dados de forma completa e estável; o terceiro nível é denominado especializado, pois nesta fase ocorre a descrição nas tarefas genéricas de ações que devem ser realizadas para compor determinada situação específica; e o quarto nível corresponde à instância do processo, onde ocorre o registro das ações, bem como, os resultados da mineração de dados e seu engajamento (CHAPMAN et. al. 2000).

Chapman et al. (2000), descrevem que a metodologia CRISP-DM envolve o modelo de referência que apresenta uma visão das fases, tarefas e saídas que compõe o processo de mineração de dados. Este modelo fomenta uma visão do ciclo de vida de uma aplicação de mineração de dados e suas respectivas tarefas, bem como, suas relações. O ciclo engloba 6 (seis) fases, porém a sequência não é inflexível, havendo a possibilidade de ir e voltar em diversas etapas conforme o necessário, onde as setas indicam as dependências mais frequentes entre as fases. Esta metodologia é exibida na figura 9.

FIGURA 9 – FASES DO MODELO DE REFERÊNCIA



FONTE: Chapman et. al. (2000, p.10).

O ciclo externo a figura, significa a natureza do ciclo de mineração de dados, onde não termina uma vez que a metodologia é implantada. A fase inicial é a de compreensão de negócios, onde se definem os objetivos e requisitos do projeto para em seguida converter este conhecimento em um problema de mineração de dados e planejar a melhor forma de se alcançar os objetivos; posteriormente, é necessário

realizar a compreensão de dados, realizando a coleta de dados iniciais e familiarizar-se a eles, a fim de identificar problemas, hipóteses e informações ocultas para detectar os primeiros *insights* sobre este conjunto de dado; já a fase de preparação dos dados, envolve as atividades para a construção do conjunto de dados final, ou seja, os dados que serão utilizados na modelagem visto que para a preparação é necessário realizar o registro, transformação e limpeza destes dados; na etapa de modelagem, são selecionados e aplicados parâmetros para este conjunto de dados, sendo estes definidos a partir do problema de mineração de dados, lembrando que geralmente será necessário voltar para a etapa de preparação de dados; a etapa de avaliação engloba o modelo em que foi construído, onde é necessário avaliá-lo para ter certeza que este atinge os objetivos traçados, assim o processo de mineração de dados será alcançado; por fim, chega-se a etapa de desenvolvimento, onde a criação do modelo não é considerada o momento final do projeto, pois geralmente ocorre o processo de aplicação, onde a mineração de dados é repetida para se fazer uso ao modelo gerado (CHAPMAN et. al. 2000).

2.4 ENOGASTRONOMIA

Segundo o Dicionário *Michaelis* (2018), o termo Enogastronomia significa “arte de harmonização entre cozinha e adega, isto é, a habilidade na escolha do vinho para acompanhar determinado prato”. Viganò (2015), escreve que o termo Enogastronomia possui 3 (três) vertentes, sendo elas: uma expressão da cultura local; forma de compor a identidade de uma região e a associação entre gastronomia, vinhos e turismo.

O conceito de Enogastronomia é complexo e amplo, pois é sustentado por um tripé entre o produtor rural, vinícolas/restaurantes e o território, onde são desenvolvidos e aplicados os diferentes conhecimentos, seja teórico, técnico e prático, oferecendo produtos únicos e de qualidade aos consumidores (SILVA, 2015). Beato (2006), descreve que o termo Enogastronomia é composto por entendimento e sentido, onde se deve alternar entre complementaridade e contraposição, inseridos em um processo dinâmico e construtivo que leva em conta as regras, costumes e culturas.

Dardeau (2004) apresenta a ideia de que este termo envolve conjugar vinho com comida, levando em conta a ligação entre estes elementos. Harmonizar não significa eleger vinhos de forma aleatória, exige observação das peculiaridades de cada elemento para que se transforme em uma combinação perfeita, de forma que não promova uma disputa entre os elementos. Para Beato (2006), o processo Enogastronômico envolve harmonizar vinhos e comidas, onde os dois devem estar alinhados considerando o mesmo nível de intensidade e qualidade, de forma que estes se unam e tornem o conjunto do todo.

De acordo com Larousse (2007), para a realização da harmonização é necessário seguir 3 (três) regras fundamentais: a cor entre prato e vinho; a densidade alcoólica do vinho e a concentração aromática do vinho deve acompanhar o sabor do prato; e, o aroma entre comida e vinho deve ser agradável, enfatizando a harmonia. Os processos de harmonização devem englobar os quesitos da criatividade, a fim de adquirir competência para desvendar o nuance de cada vinho, suas origens, aromas e sabores, bem como, os temperos e ingredientes que compõe o prato, para assim compor uma harmonização prazerosa (ROCHA, 2009).

Para César (2008), ao propor uma harmonização, pode-se utilizar 2 (dois) critérios distintos, sendo eles: a similaridade e o contraste. A similaridade, baseia-se na afinidade características encontradas no vinho e na comida, enquanto no contraste as características de vinho e comida se distinguem, com o intuito de suavizar algum elemento que se sobressai na composição. Deve existir um equilíbrio entre prato e vinho, onde a harmonização se baseie na análise dos componentes de paladar, aroma e texturas encontradas no vinho e na comida (CÉSAR, 2008).

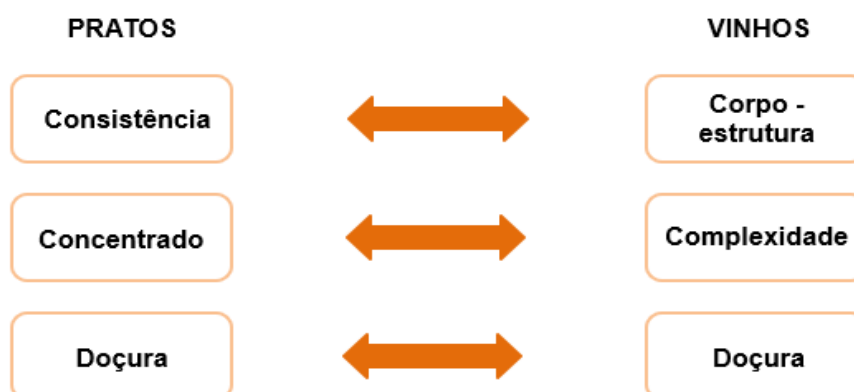
A Enogastronomia de acordo com a Confraria dos *Sommeliers* [2008?], define cinco variáveis que podem influenciar na harmonização, sendo elas: as variáveis das receitas preparadas pelos chefes de acordo com sua especialidade; a temperatura de serviço de vinho e do prato, onde se deve ter um conhecimento vasto para não ocorrer discordância de temperaturas; a tradição da região, onde ao preparar pratos com característica de determinada região vale utilizar vinhos provenientes da mesma área ou próximas; a safra do vinho, bem como o tipo e a conservação; e, o gosto pessoal.

Milazzo (2015, p.04, tradução nossa), cita que os fatores fundamentais para realizar a combinação entre vinho e alimentos são:

- **estações:** durante o verão, utiliza-se mais o vinho branco do que o vinho rosado; e o vinho rosado se utiliza mais do que o vinho tinto. Já no inverno, pode-se existir preferências opostas;
- **temperatura do alimento:** os vinhos mais adequados com aperitivos quentes são aqueles servidos em temperaturas não muito baixas. Para vinho branco (8°C) e vinho rosado (12° - 14°C);
- **gosto pessoal:** está fortemente relacionado ao hábito alimentar ou tradição, onde cada indivíduo possui uma escolha individual do vinho.

Segundo *Abbinamenti Cibo – Vino* ([2012?], tradução nossa), existem regras de concordância e contraste de sabores entre o vinho e comida, onde a correspondência por concordância ocorre quando o vinho acompanha as sensações do prato, enquanto a correspondência de contraste se dá quando o vinho expressa características opostas às dos alimentos. Para a Confraria dos *Sommeliers* [2008?], a Enogastronomia engloba a regra da justaposição entre características dos alimentos similares ou incompatíveis com as do vinho, por meio de concordância ou contraposição. A figura 10 representa como está relacionado a harmonização por concordância entre pratos e vinhos.

FIGURA 10 – HARMONIZAÇÃO POR CONCORDÂNCIA



FONTE: Confraria dos *Sommeliers* ([2008?], p. 03).

Bigonha (2016), aborda a harmonização por concordância como a busca por elementos semelhantes ao vinho, a fim de gerar uma agradável conjunção de aromas,

sabores e texturas. Beato (2006), enfatiza que no que tange a estrutura de harmonização, existem 2 (dois) modos de combinar, sendo eles: fusão e contraposição. A harmonização realizada por fusão é aquela que dá continuidade ao sabor do prato, existindo uma concordância entre prato e vinho. É apresentado na figura 11, a relação de harmonização por contraposição.

FIGURA 11 – HARMONIZAÇÃO POR CONTRAPOSIÇÃO



FONTE: Confraria dos *Sommeliers* ([2008?], p. 04).

Já a harmonização por contraposição se dá em dispor de um vinho com características opostas ao prato que está sendo combinado (BEATO, 2006). A harmonização por contraste ou contraposição, utiliza componentes que se opõe para criar um equilíbrio entre alimento e vinho (PUCKETTE; HAMMACK, 2016).

Além destas harmonizações, existem as harmonizações regionais, de estação e difíceis. As harmonizações regionais, que contemplam pratos e vinhos da mesma região, tendo um grande valor histórico, porém nem sempre seguem as regras corretas para uma harmonização; já as harmonizações da estação, estão relacionadas a apreciação em determinadas épocas do ano, com alimentos sazonais; e, as harmonizações difíceis está relacionada a alimentos que envolvem elevados níveis de acidez ou sabor, doces com base alcoólica, frutas e sorvetes (CONFRARIA dos *Sommeliers*, [2008?]).

Para Milazzo (2015, p.02, tradução nossa), a efervescência reforça a frescura da acidez do vinho, onde os alimentos com mais gordura são mais adaptáveis a vinhos ácidos e espumantes, enquanto o vinho tanino é capaz de neutralizar a gordura, limpando a boca e regenerando as papilas gustativas; Já os vinhos doces são recomendados para harmonizações com sobremesas, pois a acidez do vinho pode trazer uma sensação desagradável após ingerir o doce da sobremesa.

A orientação para realizar uma harmonização adequada entre alimentos e vinhos está em provocar as sensações do gosto de forma adequada, na qual os pratos leves devem ser acompanhados por vinhos frescos, secos, jovens e leves de álcool no corpo. Enquanto os pratos robustos, estruturados e concentrados, requerem vinhos com certo grau de envelhecimento, tânico, seco, alto teor alcoólico e encorpado (MILAZZO, 2015, p.01, tradução nossa).

Existem combinações e diversas possibilidades que podem fazer com que o processo Enogastronômico seja eficiente, visto que pode existir diferentes vinhos para determinado tipo de prato e diferentes pratos para certo tipo de vinho. Deve-se existir uma relação recíproca, para que assim surjam novos sabores para atuar no modo de como se sente a textura, consistência e gosto dos alimentos (BEATO, 2006).

Inseridos no *Ebook* “Harmonização: Aprendendo os conceitos” [2016?], estão situados os quesitos necessários para a avaliação e combinação entre comida e vinho, como expressa o quadro 8.

QUADRO 8 – CARACTERÍSTICAS DE HARMONIZAÇÃO ENTRE COMIDA E VINHO

(Continua)

	COMIDA	VINHO
CORPO	Característica mais importante na harmonização; Relacionada ao peso do prato, seja leve ou pesado;	Sensação de peso do vinho; Encorpado: Alto nível de álcool, densidade e concentração da fruta; Quanto mais líquido, mais leve é o vinho.
DOÇURA	Sacarose (açúcar de mesa); Frutose (frutas); Compostos sintéticos; Torna o sabor do prato mais amargo e ácido, além de reduzir a sensação de corpo e fruta do vinho.	Açúcar residual presente na fruta, no qual não se transformou no processo de fermentação.
SALGADO	Realça o sabor dos pratos; É facilmente percebido na boca; Torna o gosto do vinho mais encorpado, menos ácido e menos amargo.	Não se aplica.
AMARGOR	Valoriza o prato; Aumenta a sensação do amargor do vinho.	Não se aplica.
ACIDEZ	Gosto primário em uma harmonização; Podem ser naturais ou industriais (cítrico, láctico, tartárico, acético e málico); Sensação de "água na boca"; Torna o vinho menos ácido e com características mais encorpada e frutada.	Um dos principais fatores do vinho; Proporciona a sensação de salivação; Componente essencial para equilibrar vinho e comida.

	COMIDA	VINHO
FRUTADO	Trazem aromas e sabores marcantes ao prato; Alteram o gosto original do alimento.	Vinhos mais novos; Trazem sabores primários; Pouca ou nenhuma maturação em barricas; Com o passar do tempo o sabor fica mais sutil.
PICANTE	Traz calor ao paladar; Pode ser fraco, moderado ou forte; Uso de produtos que possuem Capsaicina e Piperina, onde são interpretadas pelo cérebro e respondem como endorfina, proporcionando a sensação de bem-estar.	Não se aplica.
GORDURA	Vegetal ou animal; Ofertam corpo e contraste harmônico com os vinhos.	Não se aplica.
SECURA	Pratos mais ou menos secos; Deve-se ter cuidado em harmonizar um prato muito seco com vinho tanino, pois a junção traz a sensação de secura na boca.	Aspereza ou maciez; Vinhos evoluídos trazem sensações amanteigadas e aveludadas na boca.
TEXTURA	Sensações táteis causadas por alimentos; Traz crocância, maciez e aspereza aos pratos.	Não se aplica.
ÁLCOOL	Não se aplica.	Fermentação dos açúcares; Sensação de calor ao paladar; Traz corpo ao vinho
TANINO	Não se aplica.	Presentes na casca da uva e nos barris de carvalho; Trazem sensação de adstringência; A maturação torna o vinho mais arredondado.

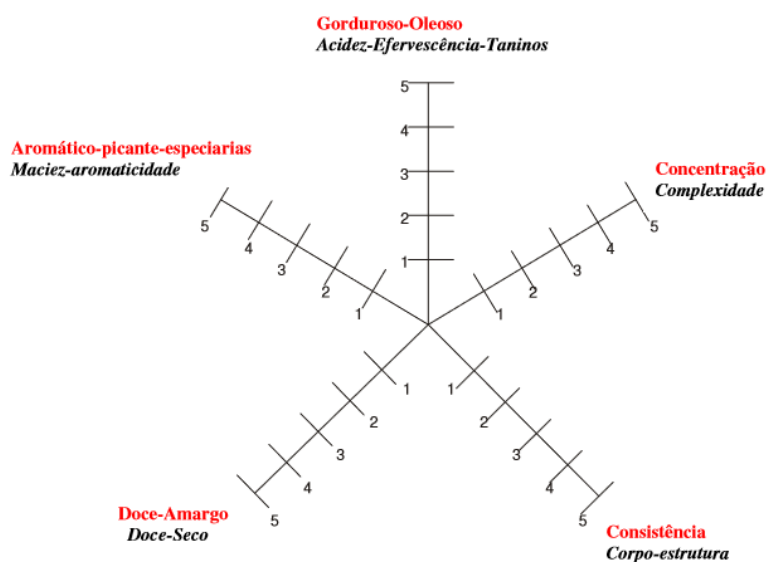
FONTE: Adaptado de Harmonização: Apreendendo os conceitos [2016?].

Existem alguns quesitos como proporção de álcool e o composto químico tanino não se enquadram na avaliação de comidas, enquanto os quesitos como o

salgado, amargor e picante não se enquadram na avaliação dos vinhos, para realizar uma harmonização Enogastronômica.

O ser humano tem a capacidade de identificar 6 (seis) sabores através do paladar, sendo eles, gorduroso, picante, salgado, ácido, doce e amargo. No momento de uma degustação, é possível sentir sensações, como efervescência, dormência e refrescância (PUCKETTE; HAMMACK, 2016). O Modelo Fisar, apresentado na figura 12, demonstra algumas combinações existentes que servem como base para a avaliação de um prato e que posteriormente é complementado com o Modelo Mercadini.

FIGURA 12 – MODELO FISAR DE HARMONIZAÇÃO



FONTE: Confraria dos *Sommeliers* ([2008?], p. 07).

Os elementos que compõe o Modelo Fisar de harmonização são descritos como:

- **Gorduroso/Oleoso:** gordura e óleos próprio do alimento ou adicionado ao mesmo;
- **Aromático/Picante:** adição de especiarias e substâncias aromáticas;
- **Doce/Amargo:** o gosto propriamente dito, representando uma sensação mais nítida;
- **Consistência:** compactação física do alimento;
- **Concentração:** grau de cozimento e maturação (VALDUGA, 2018).

2.4.1 Modelo Mercadini

A comida e o vinho devem se complementar, onde nenhum prevaleça no momento da degustação e sejam juntos mais saborosos do que consumidos individualmente (DONADINI, 2008). Desta forma, para a construção de uma harmonização completa, devem-se considerar componentes aromáticos e gustativos do prato, bem como, o doce ou salgado, suculência, untuosidade, aromaticidade, especiarias, sabor e acidez, enquanto dos vinhos envolve a maciez, acidez, aromaticidade, álcool e efervescência, pois a harmonia nasce a partir do contraste entre estes elementos (BORGES, 2016).

O método conhecido como Mercadini, baseia-se nos princípios de semelhança e contraste de sabores entre vinho e comida, onde se devem considerar as sensações gustativas provocadas com a degustação (*ABBINAMENTI CIBO – VINO*, [2012?], p.1, tradução nossa).

Segundo Vaccarini (2010, p.04, tradução nossa), os princípios fundamentais deste método pressupõem que as características do vinho estão em contraste com as do alimento, a fim de suavizar características agressivas desta harmonização ou de concordância, onde existe semelhança entre o alimento e vinho. Com a aplicação do método Mercadini, ao atribuir valores aos componentes do prato e do vinho, gera-se um gráfico, cujo estilo é radar, com o intuito de prever com objetividade a harmonização proposta. Além disso, é possível avaliar contrastes e semelhanças para texturas e aromas (BORGES, 2016).

Consideram-se como parte das regras básicas da Enogastronomia, os componentes básicos da comida e vinho. Para a comida, é necessário avaliar a suculência, aromas, condimentos, teores de gordura, acidez, tendência à doçura e traços de amargor. Para *Abbinamenti Cibo – Vino* ([2012?], tradução nossa), são considerados critérios de avaliação para a aplicação do Modelo Mercadini, sendo apresentado no quadro 9 os aspectos de avaliação para comidas.

QUADRO 9 – ASPECTOS DE AVALIAÇÃO DE COMIDAS

ASPECTOS DE AVALIAÇÃO - COMIDAS	
Acidez	Presença significativa de substâncias ácidas, preparações com algo azedo (vinagre balsâmico);
Amargor	Alimentos que possuem características intrínsecas (fígado, trufas e alcachofras);
Aromaticidade	Aroma natural dos alimentos, cozidos ou crus, acompanhados de ervas/temperos;
Doce	Alimentos que possuem uma tendência adocicada (pão, frutas, amidos, carnes e peixes);
Especiarias	Adição de temperos;
Gordura	Fornecem a percepção sólida entre a língua e o céu da boca (manteiga ou banha);
Gosto Olfativo	Quanto tempo após a mastigação, o gosto é mantido no paladar;
Oleosidade	Uso de gorduras líquidas na elaboração do prato (óleo vegetal);
Sabores	Doce, ácido, salgado e amargo;
Sensações	Gordura/oleosidade;
Tempero	Uso de especiarias ou ervas aromáticas.

FONTE: Adaptado de *Abbinmaneti Cibo – Vino* ([2012?], p.08, tradução nossa).

Enquanto para o vinho, deve-se identificar a maciez, efervescência, acidez, grau alcoólico, taninos e aromas (FERRÃO, 2011). O quadro 10 demonstra os aspectos de avaliação para os vinhos.

QUADRO 10 – ASPECTOS DE AVALIAÇÃO DE VINHOS

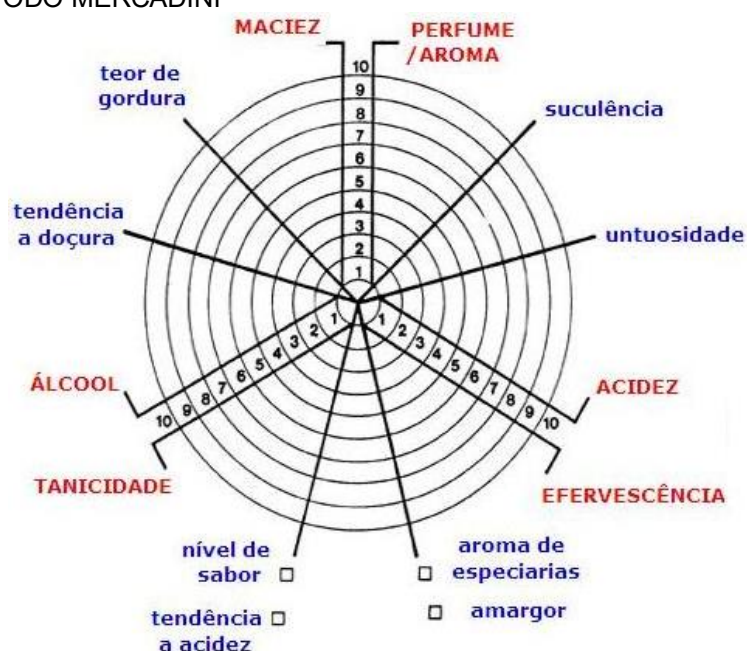
ASPECTOS DE AVALIAÇÃO - VINHOS	
Acidez	Efervescência, sabor de frescor e limpeza da boca;
Aroma	Junção de compostos orgânicos utilizados no processo de fabricação do vinho (uvas, fermentação e maturação);
Álcool/Tanicidade	Fornece a estrutura do vinho;
Maciez	Harmonia do vinho, dada pela combinação de álcool, açúcares e poliálcool;
Suavidade	Aromas que proporcionam amenização ao vinho.

FONTE: Adaptado de *Abbinmaneti Cibo – Vino* ([2012?], p.10, tradução nossa).

O Método Mercadini é apresentado de forma gráfica, de modo a avaliar que cada sensação percebida na comida e no vinho possa existir uma relação de oposição e concordância. É necessário reconhecer as sensações percebidas e cruzar o

quadrado que interessa na análise, com o intuito de avaliar o nível de percepção e posteriormente pontilhar uma intersecção da linha com o círculo. No final da análise, os pontos marcados são unidos com cores diferentes, para não confundir quais são os aspectos da comida e do vinho, formando assim dois polígonos que serão similares se houver emparelhamento harmonioso, mas com vértices opostas (*ABBINAMENTI CIBO – VINO*, [2012?], tradução nossa). A figura 13 demonstra como é a representação deste método, onde os aspectos de cor azul representam os critérios de comidas e os aspectos de cor vermelha, representam as características dos vinhos.

FIGURA 13 – MÉTODO MERCADINI



FONTE: Miranda (2009).

De acordo com *Abbinamenti Cibo - Vino* ([2012?], tradução nossa), para que a aplicação do Método Mercadini obtenha um valor significativo, é necessário que sejam avaliadas no mínimo 3 (três) sensações de alimentos e 3 (três) sensações de vinhos. Se os polígonos são semelhantes, mas pequenos, os componentes de comida e vinho possuem sensações olfativas de pouca importância. Mas se os polígonos são semelhantes, com vértices evidentes, a comida e vinho são estruturados e complexos com sensações sensíveis (*ABBINAMENTI CIBO – VINO*, [2012?], tradução nossa).

A composição gráfica do Método Mercadini possui 10 (dez) círculos concêntricos que acompanham números de 0 à 10, iniciando do núcleo para o exterior; além disso, são dispostas 6 (seis) linhas que destacam as características orgânicas do vinho e 6 (seis) linhas divergentes do centro, que expressam as sensações

percebidas para os alimentos (*ABBINAMENTI CIBO – VINO*, [2012?], tradução nossa). As sensações dos alimentos são avaliadas em uma escala de valores de 0 a 10, como demonstra o quadro 11.

QUADRO 11 – MODELO SE AVALIAÇÃO DAS SENSações DAS COMIDAS

AVALIAÇÃO DAS SENSações - COMIDAS		
LISTA	CONCEITO	DESCRIÇÃO
0 1 2	IMPRERCEPTÍVEL	Não avaliável, pois apenas está mencionado ou não está presente;
2 3 4	POUCO PERCEPTÍVEL	Sente-se a sensação em segundo plano;
4 5 6	BASTANTE PERCEPTÍVEL	A sensação é percebida o suficiente para ser identificada;
6 7 8	PERCEPTÍVEL	Sensação clara, bem identificada e forte, fácil de ser identificada;
8 9 10	MUITO PERCEPTÍVEL	Sensação predominante; quase caracteriza exclusivamente o prato.

FONTE: *Abbinamenti Cibo – Vino* ([2012?], p.12, tradução nossa).

Para realizar a avaliação das sensações relacionadas ao sabor do vinho, deve-se considerar o cartão de análise sensorial dos vinhos, onde é disponibilizado 5 (cinco) adjetivos para cada elemento do gosto olfativo e valores específicos para a efervescência do vinho, sendo necessário atribuir um valor numérico para cada adjetivo (*ABBINAMENTI CIBO – VINO*, [2012?], tradução nossa). A análise sensorial ocorre a partir da interpretação dos estímulos e interação que a bebida provoca no indivíduo, gerando uma interpretação das propriedades que compõe o vinho a partir de questionários respondidos no ato da prova da bebida, tendo como intuito verificar a aceitação gustativa das sensações identificadas no consumo da bebida (TECH, 2013). O modelo de avaliação das sensações é apresentado no quadro 12.

QUADRO 12 – MODELO DE AVALIAÇÃO DAS SENSAÇÕES DOS VINHOS

AVALIAÇÃO DAS SENSAÇÕES - VINHOS									
LISTA	INTENSIDADE OLFATIVA	PERSISTÊNCIA AROMÁTICA INTENSA (P.A.I)	EFERVESCÊNCIA	ACIDEZ	SABOR	ÁLCOOL	TANICIDADE	MACIEZ	DOÇURA
0 1 2	CARENTE	CURTO	SEM EFERVESCÊNCIA	FRACO	SEM SABOR	POUCO ÁLCOOLICO	SUAVE	FINO	SECO
2 3 4	POUCO INTENSO	POUCO PERSISTENTE	POUCO EFERVESCENTE	POUCO ÁCIDO	POUCO SABOR	POUCO ÁLCOOLICO	POUCO TÂNICO	POUCO MACIO	ADOCICADO
4 5 6	BASTANTE INTENSO	BASTANTE PERSISTENTE	BASTANTE EFERVESCENTE	BASTANTE ÁCIDO	BASTANTE SABOR	BASTANTE ÁLCOOLICO	BASTANTE TÂNICO	BASTANTE MACIO	AGRADÁVEL
6 7 8	INTENSO	PERSISTENTE	EFERVESCENTE	ÁCIDO	SABOROSO	ÁLCOOLICO	TÂNICO	MACIO	DOCE
8 9 10	MUITO INTENSO	MUITO PERSISTENTE	MUITO EFERVESCENTE	MUITO ÁCIDO	MUITO SABOR	MUITO ÁLCOOLICO	MUITO TÂNICO	MUITO MACIO	EXTREMAMENTE DOCE

FONTE: *Abbinmaneti Cibo – Vino* ([2012?], p.13, tradução nossa).

Para *Abbinmaneti Cibo – Vino* ([2012?], tradução nossa), não existe uma sequência pré-estabelecida, porém geralmente inicia-se com a sensação predominante na comida que prossegue gradualmente para as demais sensações. O valor estabelecido pelas sensações é anotado no cartão, onde posteriormente desenha-se a intersecção entre os círculos e linhas. Quanto maior o valor, mais fora da cruz à sensação está. Deve-se realizar a avaliação de todos os quesitos, tanto para comidas quanto para vinhos para que o polígono apareça. Com os polígonos desenhados no Modelo Mercadini, é necessário observar se existe uma proporção das dimensões entre a comida e o vinho, a fim de definir a estrutura dos alimentos, onde se deve avaliar o tamanho do polígono entre pequeno, médio e grande e aplicar os seguintes quesitos:

- 1) Pouco estruturado
- 2) Muito estruturado
- 3) Estruturado

Enquanto para o corpo do vinho:

- 1) Fraco
- 2) Do corpo
- 3) Robusto

Avaliação da combinação entre comida e vinho:

- 1) Não harmonioso
- 2) Muito harmonioso
- 3) Harmônico

Assim é possível observar o grau de harmonia entre prato e vinho, mas também se deve confiar no gosto, visto que alguma avaliação errada pode criar um sabor devastador no momento da degustação (*ABBINAMENTI CIBO – VINO*, [2012?], tradução nossa).

2.4.2 Combinações Comuns Entre Vinhos e Comidas

Segundo Puckette e Hammack (2016), “a harmonização com alimentos é a arte de criar uniões adequadas ao considerar fatores como gosto, textura, aroma e intensidade”. Para Pereira (2014), as combinações entre comidas e vinhos vêm sendo testadas ao longo do tempo e várias regras foram criadas, agradando à maioria das pessoas que experimenta desta harmonização. Qualquer harmonização leva em conta os sabores principais (doce, salgado, amargo e ácido), pois:

- o salgado reforça o amargo;
- o amargo reforça a acidez;
- o doce suaviza os sabores ácidos, amargos e salgados;
- o ácido deve ser igualado.

Inspirada nas obras de Simon (2000) e Larousse (2004), Queiroz (2009) afirma que o importante para combinar pratos e vinhos é casar pesos semelhantes, para que um não ocupe o lugar do outro. Desta forma, é necessário conhecer a estrutura do prato, bem como, acidez, doçura, intensidade, temperos e aromas, além de conhecer a intensidade do vinho. O quadro 13, demonstra algumas harmonizações comuns baseadas em contraste e similaridade.

QUADRO 13 – COMBINAÇÕES COMUNS ENTRE PRATOS E VINHOS

COMBINAÇÕES COMUNS ENTRE VINHOS E PRATOS	
ACIDEZ + ACIDEZ	- Em um prato ácido como peixes e frutos do mar, o ideal é combinar com um vinho igualmente ácido ou mais ácido que o prato; - Deve-se atentar com a harmonização em saladas, visto que a acidez dos temperos pode inviabilizar qualquer harmonização.
DOCE + DOCE	- Uma sobremesa combina com um vinho doce, porém se existir acidez no prato é necessário que acompanhe um vinho com características ácidas; - É importante que a doçura do vinho seja maior do que a da sobremesa.
ACIDEZ + GORDURA	- Pratos ricos em gorduras envolvem toda a boca, onde para existir um contraste e quebrar a gordura é necessário utilizar um vinho ácido e permitir a sensação de limpeza da boca.
DOCE + SALGADO	- Pratos muito salgados devem compatibilizar com vinhos doces, considerando que a quantidade de sal e açúcar deve ser equivalente.
TANINO + PEIXE	- Está é uma exceção na harmonização, visto que geralmente vinhos que apresentam taninos não combinam com peixes; - Para peixes untuosos e encorpados como o bacalhau e salmão, os vinhos tintos taninos harmonizam facilmente com o prato.
TANINO + CARNE BEM PASSADA	- Carnes bem passadas, harmonizam com vinhos de média tanicidade; - Outro tipo de harmonização pode proporcionar a sensação de amargor.

FONTE: Adaptado de Queiroz (2009, p.41).

Para Borges (2006), os princípios de equilíbrio entre comidas e vinhos são definidos em:

- os frutos do mar, crustáceos, mariscos, ostras e moluscos harmonizam com vinhos brancos e espumantes que possuam corpo leve ou médio;
- as aves e carnes (bovina, suína, ovina e caprina), combinam com vinhos brancos, rosados ou tintos e, de acordo com o peso, envolvem vinhos leves de médio corpo ou encorpados;
- as caças permitem a utilização de vinhos tintos entre bom corpo para encorpados;
- os cogumelos e trufas, necessitam de vinhos rosados ou tintos de bom corpo.

A partir da apresentação dos conceitos relacionados a este estudo, apresentam-se na próxima seção os encaminhamentos metodológicos adotados.

3 ENCAMINHAMENTOS METODOLÓGICOS

Esta seção apresenta os procedimentos metodológicos, bem como, a caracterização da pesquisa e seus materiais e métodos necessários para realizar a extração das informações que compuseram a base de dados de estudo e a mineração de dados.

3.1 CARACTERIZAÇÃO DA PESQUISA

Esta pesquisa é de natureza **qualitativa** tendo como propósito a coleta e análise de dados relacionados às características provenientes de comidas e vinhos, a fim de se atingir o objetivo geral estabelecido para este projeto, visto que não há fórmulas delineadas para orientar o pesquisador, dependendo de sua capacidade e estilo (GIL,2008).

Com relação a sua finalidade, trata-se de uma pesquisa **aplicada**, na qual se busca uma aplicação prática e dirigida à solução do problema levantado neste estudo (GIL, 2008).

De acordo com os objetivos propostos, classifica-se esta pesquisa de caráter **exploratório**, visto que este tipo de pesquisa tem como finalidade desenvolver, esclarecer e investigar conceitos e ideias, com o propósito de formular problemas mais precisos e hipóteses para trabalhos futuros, na qual este projeto investiga uma temática ainda não explorada na área da Gestão da Informação (GIL, 2008).

Quanto aos procedimentos, trata-se de um estudo **experimental**, no qual a partir de um objeto de estudo, pretende-se descobrir e selecionar as variáveis cabíveis a este estudo, além de observar os efeitos e resultados destas variáveis na aplicação (GIL,2008).

A figura 14 mostra de forma sintetizada a caracterização final da pesquisa, de acordo com as classificações apresentadas por GIL (2008).

FIGURA 14 – CARACTERIZAÇÃO DA PESQUISA



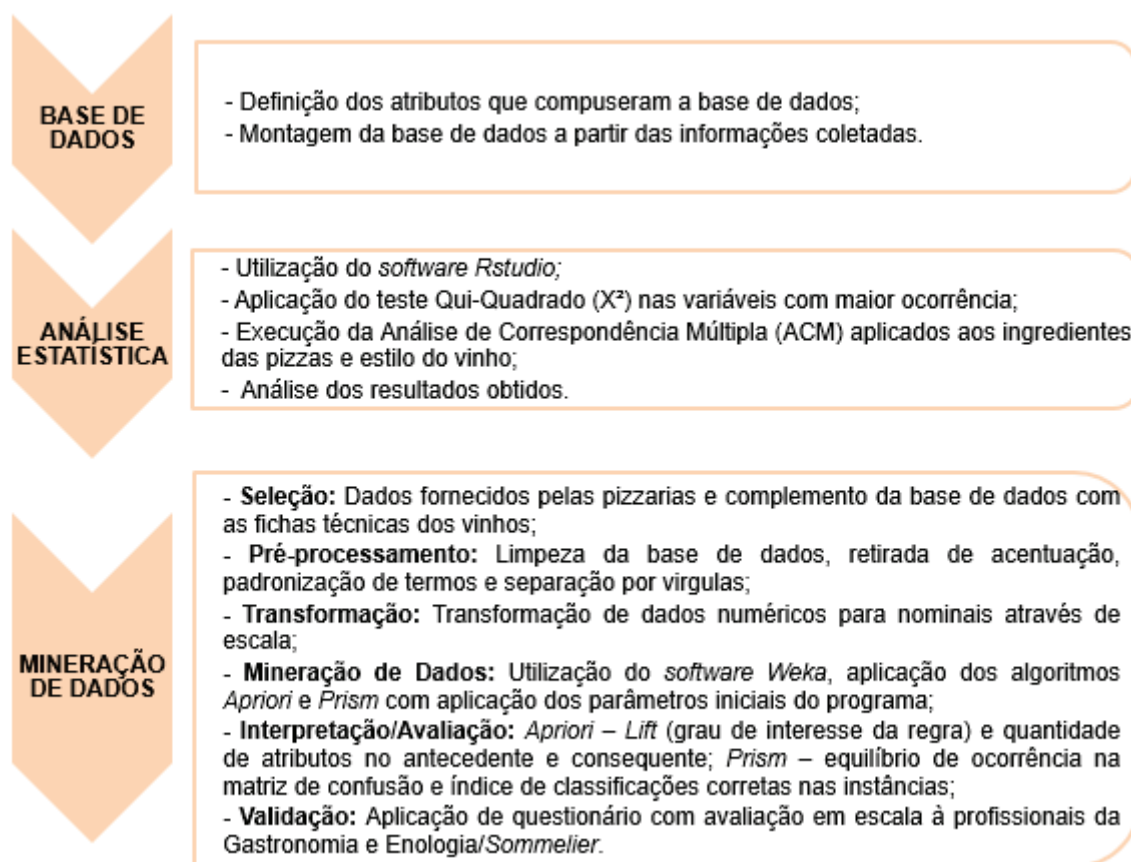
FONTE: A autora (2018).

A seguir são apresentados os materiais e métodos utilizados para alcançar os objetivos deste estudo.

3.2 MATERIAIS E MÉTODOS

Esta seção apresenta a base de dados deste trabalho, bem como, seus atributos e formatos, o tratamento dos dados, escolha dos testes estatísticos e algoritmos de mineração de dados. Definiram-se os procedimentos operacionais utilizados para alcançar os objetivos propostos. A figura 15 demonstra como ocorreu este processo.

FIGURA 15 – PROCEDIMENTOS METODOLÓGICOS



FONTE: A autora (2018).

A seguir, serão detalhados como ocorreu cada um destes procedimentos.

3.2.1 Base de Dados

Para a construção de uma base de dados Enogastronômica, foram necessários definir quais atributos seriam relevantes para realizar a comparação entre os pratos e vinhos, então, através da literatura pertinente e as informações colhidas com um *Sommelier* buscou-se quais atributos seriam importantes para realizar a aplicação da mineração de dados, totalizando 21 (vinte e um) atributos, sendo 16 (dezesesseis) voltados para os vinhos e 5 (cinco) voltados ao prato.

O quadro 14 representa os metadados para cada um dos atributos definidos para a categoria denominada vinhos, que representam as características do vinho, bem como, as uvas utilizadas em cada rótulo, tempo de amadurecimento e tipo de barrica utilizada neste processo, o valor de comercialização da garrafa e a sua origem de fabricação.

QUADRO 14 – DESCRIÇÃO DOS ATRIBUTOS DO VINHO INCLUINDO SEUS TIPOS E VARIAÇÕES

(Continua)

PARA VINHOS			
ATRIBUTO	TIPO DO ATRIBUTO	VALORES DO ATRIBUTO	DESCRIÇÃO
ID	Código numérico	Valor numérico de até três algarismos. Exemplo: 1,2,3,4,5,6	Identificação numérica do rótulo do vinho
Nome	<i>String</i> de caracteres	Valores categóricos em texto. Exemplo: Vinho Victorio Malbec	Identificação nominal do rótulo do vinho
Região	<i>String</i> de caracteres	Valores categóricos em texto. Exemplo: Mendoza	Região de produção do vinho
Origem	<i>String</i> de caracteres	Valores categóricos em texto. Exemplo: Argentina	País onde o vinho foi produzido
Preço	Moeda	*Valor numérico de até seis algarismos para representar o valor monetário do vinho. Exemplo: R\$100,00	Preço de comercialização do vinho na casa
Estilo	Código numérico	*Valores categóricos em texto: 1 - Espumante; 2 - Vinho Branco Leve; 3 - Vinho Branco Encorpado; 4 - Vinho Branco Aromático; 5 - Vinho Rosado; 6 - Vinho Tinto Leve; 7 - Vinho Tinto Médio; 8 - Vinho Tinto Encorpado; 9 - Vinho Sobremesa	Estilo de como o vinho é caracterizado
Uva	<i>String</i> de caracteres	Valores categóricos em texto separados por vírgula. Exemplo: Malbec, Sirah, Carmenere	Variedade da fruta para fabricação do vinho
Corpo	Código numérico	*Valores categóricos em texto: 1 - Leve; 2 - Entre Leve e Médio; 3 - Médio; 4 - Entre Médio e Encorpado; 5 - Encorpado	Nível de sabor da junção de todos os elementos do vinho

(Conclusão)

Doçura	Código numérico	*Valores categóricos em texto: 1 - Extraseco; 2 - Seco; 3 - Meio Seco; 4 - Doce; 5 - Extra Doce	Nível de açúcar residual do vinho (frutose)
Tanino	Código numérico	*Valores categóricos em texto: 1 - Baixo ; 2 - Entre Baixo e Médio ; 3 - Médio ; 4 - Adstringente ; 5 - Muito Adstringente	Nível de amargor do vinho
Acidez	Código numérico	*Valores categóricos em texto: 1 - Baixa; 2 - Entre Baixa e Média; 3 - Média; 4 - Entre Média e Alta; 5 - Alta	Nível de ácido disponível nas uvas e a adquiridos no tempo de maturação do vinho
Álcool	Porcentagem	*Valores numéricos de até três algarismos. Exemplo: 14,2%	Nível de teor alcoólico disposto em cada garrafa de vinho
Amadurecimento	Número inteiro	Valores numéricos inteiros para representar os meses de maturação. Exemplo: 10 - Dez Meses	Representação do tempo (em meses) necessário para realizar a maturação do vinho
Barrica	<i>String</i> de caracteres	Valores categóricos em texto. Exemplo: Carvalho	Tipo de barrica utilizada para a maturação do vinho
Temperatura	Número inteiro	Valores numéricos inteiros com limite de 0 à 20. Exemplo: 16	Representação da temperatura de serviço em Graus Celsius
Aromas possíveis	<i>String</i> de caracteres	Valores categóricos em texto separados por vírgula. Exemplo: Amora, Café, Damasco	Aromas que podem ser sentidos no momento da degustação do vinho

* De acordo com Puckette e Hammack (2016).

FONTE: A autora (2018).

O quadro 15 apresenta os metadados para cada um dos atributos que compõem a categoria dos pratos, considerando os sabores de pizzas disponíveis nas pizzarias que participaram do estudo. Como a produção de pizza é algo simplório para a gastronomia, escolheu-se como metadados apenas os ingredientes de cada sabor e as especiarias utilizadas.

QUADRO 15 – DESCRIÇÃO DOS ATRIBUTOS DOS PRATOS INCLUINDO SEUS TIPOS E VARIAÇÕES

PARA PRATOS			
ATRIBUTO	TIPO DO ATRIBUTO	VALORES DO ATRIBUTO	DESCRIÇÃO
ID	Código numérico	Valor numérico de até três algarismos. Exemplo: 1,2,3,4,5,6	Identificação numérica do sabor da pizza
Nome	<i>String</i> de caracteres	Valores categóricos em texto. Exemplo: Pizza Marguerita	Identificação nominal da pizza
Ingredientes essenciais de preparo	String de caracteres	Valores categóricos em texto separados por vírgula. Exemplo: Muçarela, Calabresa, Azeitona	Tipos de ingredientes utilizados no preparo da pizza
Especiarias	<i>String</i> de caracteres	Valores categóricos em texto separados por vírgula. Exemplo: Orégano, Manjerição	Tipos de especiarias/temperos utilizados no preparo da pizza

FONTE: A autora (2018).

A partir da lista de atributos dos vinhos e dos pratos, todos os dados coletados que eram numéricos, como o preço de comercialização da garrafa de vinho e a temperatura de serviço do vinho foram transformados em nominais; e, os atributos em texto foram separados por vírgula, sendo alocados em colunas distintas e preenchidos com “sim” na instância de ocorrência, considerando que os algoritmos utilizados para o processo de mineração de dados para a descoberta de padrões se restringem apenas ao uso de atributos em texto.

O quadro 16 demonstra o resultado desta transformação para os atributos voltados aos vinhos, onde os metadados que possuem um asterisco (*) são baseados no modelo de classificação de Puckette e Hammack (2016).

QUADRO 16 – TRANSFORMAÇÃO DOS TIPOS DE ATRIBUTOS DOS VINHOS

(Continua)

PARA VINHOS			
ATRIBUTO	TIPO DO ATRIBUTO	VALORES DO ATRIBUTO	DESCRIÇÃO
Nome	String de caracteres	Valores categóricos em texto. Exemplo: Vinho Victorio Malbec	Identificação nominal do rótulo do vinho
Região	String de caracteres	Valores categóricos em texto. Exemplo: Mendoza	Região de produção do vinho
Origem	String de caracteres	Valores categóricos em texto. Exemplo: Argentina	País onde o vinho foi produzido
Preço	String de caracteres	*Valores categóricos em texto a partir da faixa de preços expressa a seguir: A - Até R\$60,00; B - Entre R\$60,01 e R\$100,00; C - Entre R\$100,01 e R\$170,00; D - Entre R\$170,01 e R\$250,00; E - Mais que R\$250,01	Faixa de preço comercialização do vinho na casa
Estilo	String de caracteres	*Valores categóricos em texto: 1 - Espumante ; 2 - Vinho Branco Leve ; 3 - Vinho Branco Encorpado ; 4 - Vinho Branco Aromático ; 5 - Vinho Rosado ; 6 - Vinho Tinto Leve ; 7 - Vinho Tinto Médio ; 8 - Vinho Tinto Encorpado ; 9 - Vinho Sobremesa	Estilo de como o vinho é caracterizado
Uva	String de caracteres	Valores categóricos em texto alocados em colunas distintas. Exemplo: Malbec	Variedade da fruta para fabricação do vinho
Corpo	String de caracteres	*Valores categóricos em texto: 1 - Leve ; 2 - Entre Leve e Médio ; 3 - Médio ; 4 - Entre Médio e Encorpado ; 5 - Encorpado	Nível de sabor da junção de todos os elementos do vinho
Doçura	String de caracteres	*Valores categóricos em texto: 1 - Extraseco ; 2 - Seco ; 3 - Meio Seco ; 4 - Doce ; 5 - Extra Doce	Nível de açúcar residual do vinho (frutose)

(Conclusão)

Tanino	String de caracteres	*Valores categóricos em texto: 1 - Baixo ; 2 - Entre Baixo e Médio ; 3 - Médio ; 4 - Adstringente ; 5 - Muito Adstringente	Nível de amargor do vinho
Acidez	String de caracteres	*Valores categóricos em texto: 1 - Baixa ; 2 - Entre Baixa e Média ; 3 - Média ; 4 - Entre Média e Alta ; 5 - Alta	Nível de ácido disponível nas uvas e a adquiridos no tempo de maturação do vinho
Álcool	String de caracteres	Valores categóricos em texto, utilizando o arredondamento para a porcentagem de álcool. Exemplo: Quatorze	Nível de teor alcoólico disposto em cada garrafa de vinho
Amadurecimento	String de caracteres	Valores categóricos em texto. Exemplo: Seis	Representação do tempo (em meses) necessário para realizar a maturação do vinho
Barrica	String de caracteres	Valores categóricos em texto. Exemplo: Carvalho	Tipo de barrica utilizada para a maturação do vinho
Temperatura	String de caracteres	*Valores categóricos em texto, expressos a partir da escala de temperatura a seguir: Gelado (3° a 7°C); Frio (7° a 13°C); Temperatura de Adega (13° a 16°C); Fresco (16° a 20°C)	Representação da temperatura de serviço em Graus Celsius
Aromas possíveis	String de caracteres	Valores categóricos em texto alocados em colunas distintas. Exemplo: Amora	Aromas que podem ser sentidos no momento da degustação do vinho

* De acordo com Puckette e Hammack (2016).

FONTE: A autora (2018).

Os atributos voltados para as pizzas não sofreram modificações, pois já se tratavam de dados nominais. Neste caso, houve a exclusão do atributo denominado "ID", visto que para a identificação das pizzas optou-se em utilizar suas denominações. O quadro 17 representa esta exclusão.

QUADRO 17 – TRANSFORMAÇÃO DOS TIPOS DE ATRIBUTOS DOS PRATOS

PARA PRATOS			
ATRIBUTO	TIPO DO ATRIBUTO	VALORES DO ATRIBUTO	DESCRIÇÃO
Nome	<i>String</i> de caracteres	Valores categóricos em texto. Exemplo: Pizza Marguerita	Identificação nominal da pizza
Ingredientes essenciais de preparo	<i>String</i> de caracteres	Valores categóricos em texto alocados em colunas distintas. Exemplo: Muçarela	Tipos de ingredientes utilizados no preparo da pizza
Especiarias	<i>String</i> de caracteres	Valores categóricos em texto alocados em colunas distintas. Exemplo: Manjericão	Tipos de especiarias/temperos utilizados no preparo da pizza

FONTE: A autora (2018).

Para dar início ao processo de coleta de dados voltados para a concepção da base de dados, delimitou-se o estilo dos estabelecimentos que iriam compor o estudo, limitando-se nas Pizzarias de Curitiba premiadas nos concursos: Comer&Beber ofertado pela revista Veja no ano 2017/2018 e Bom Gourmet ofertado pelo Jornal Gazeta do Povo no ano de 2018.

No concurso Comer&Beber edição 2017/2018, foram avaliados 310 (trezentos e dez) estabelecimentos, em 20 (vinte) categorias diferentes, sendo uma delas a de “Melhor Pizza”. Para esta categoria, foram premiadas 7 (sete) pizzarias, sendo elas descritas em ordem decrescente no quadro 18.

QUADRO 18 – PIZZARIAS PREMIADAS NO CONCURSO VEJA BOM GOURMET 2017/2018

VEJA COMER & BEBER EM CURITIBA 2017/2018		
PIZZARIA	PIZZA PREMIADA	ENDEREÇO
Avenida Paulista	Amalfitana (Berinjela, burrata, presunto e manjericão) e Margherita D.O.C (Molho de tomate italiano, muçarela, burrata, tomate cereja e manjericão)	Rua Emiliano Pernetta, 680 - Centro
Mercearia Bresser	Bresser (Tomate, muçarela de búfala, manjericão gigante e oesto de azeitona) e Concórdia (Berinjela, queijo de búfala e parmesão)	Avenida Sete de Setembro, 5831 - Batel
Baggio Pizzeria e Focacceria	Escarola com bacon e Primadona (muçarela, requeijão, cebola e azeitona preta)	Rua Salustiano Cordeiro, 14 - Água Verde
Abaré Pizzaria	Marguerita (Muçarela, tomate picado, parmesão e manjericão fresco) e Alcachofra (Alcachofra, presunto de parma e muçarela de búfala)	Avenida Vereador Toaldo Tulio, 1506 - Santa Felicidade
Armazém Dom Carmino	La Vera (Molho de tomate marinado, lascas de alho, muçarela de búfala e manjericão) e Alcachofra com presunto	Rua Jacarezinho, 21 - Mercês
Baviera	Marguerita e Quatro queijos (Muçarela, parmesão, provolone e catupiry)	Alameda Augusto Stellfeld, 18 - Centro
Carolla Pizza DOC	Carlotta (Muçarela, copa, alcachofra, cogumelo seco e ovo) e Boscaiola (Linguíça artesanal moída e shitake)	Alameda Dom Pedro II, 24 - Batel

FONTE: A autora (2018).

Já o concurso Bom Gourmet 2018, avaliou 181 (cento e oitenta e um) estabelecimentos de Curitiba, voltados a 36 (trinta e seis) categorias gastronômicas, onde na categoria “Melhor Pizza” premiou 6 (seis) pizzarias. O quadro 19 demonstra quais foram às pizzarias premiadas nesta edição do concurso, alocadas em ordem decrescente.

QUADRO 19 – PIZZARIAS PREMIADAS NO CONCURSO BOM GOURMET GAZETA DO POVO 2018

PRÊMIO BOM GOURMET GAZETA DO POVO 2018		
PIZZARIA	PIZZA PREMIADA	ENDEREÇO
Baggio Pizzaria & Focacceria	Prima de Parma (Molho pesto, presunto de parma, muçarela e lascas de parmesão) e Atum (Atum e muçarela)	R. Salustiano Cordeiro, 14 - Água Verde
Piola Curitiba	Curitiba (Cogumelos Shitake, tomates frescos, muçarela, muçarela de búfala e salsinha)	Alameda Dom Pedro II, 105 - Batel
Alessandro e Frederico	Burrata (Queijo burrata, tomates cereja fresco, presunto parma, aspargos laminados e queijo grana padano)	Shopping Pátio Batel - Avenida do Batel, 1.868 - Piso L3
Mercearia Bresser	Zola & Funghi (Cogumelo Paris, muçarela e queijo gorgonzola)	Avenida Sete de Setembro, 5.831 - Batel
Carolla	Margherita Gourmet (Muçarela de búfala, fatias de tomate, lascas de parmesão e brasílico)	Alameda Dom Pedro II, 24 - Batel
RM Cine Pizza	Kit Kat com Morango (Creme Ovomaltine, morango, kit kat e chantilly)	Avenida Nossa Senhora Aparecida, 1.061 - Seminário

FONTE: A autora (2018).

Entre os 2 (dois) concursos 10 (dez) pizzarias foram premiadas, sendo que 3 (três) delas foram premiadas em ambos os concursos, que é o caso das pizzarias Baggio, Carola e Mercearia Bresser. Após verificar as pizzarias premiadas nestes concursos, buscou-se em seus *sites* e páginas de *Facebook* e foi realizado contato telefônico com cada um destes estabelecimentos, a fim de verificar se a pizzaria contava com adega e se possuía profissional(is) responsável(is) pelas harmonizações (*Maitre* e *Sommelier*). Então, realizou-se uma visita nas pizzarias que dispunham de adega e de profissionais responsáveis pela harmonização, com o propósito de apresentar os objetivos deste estudo, considerando que os sabores das pizzas e os vinhos escolhidos para a harmonização ficariam a critério dos mesmos. Junto a estes profissionais, estabeleceu-se também o uso de 3 (três) rótulos de vinhos para cada sabor de pizza.

Posteriormente, verificou a disponibilidade das 10 (dez) pizzarias premiadas nos dois concursos quanto a possibilidade da participação destes estabelecimentos

no processo de concepção da base de dados, onde foram obtidas as seguintes respostas:

- 2 (duas) pizzarias aceitaram participar do estudo;
- 2 (duas) destas pizzarias possuem adegas, mas não dispõem de um profissional responsável pelas harmonizações, no qual, os garçons recebem o treinamento dos fornecedores de vinhos para realizar as recomendações;
- 2 (duas) pizzarias estavam com os profissionais responsáveis pela realização das harmonizações em período de férias nas datas estabelecidas no cronograma voltado para a construção da base de dados;
- 2 (duas) pizzarias não dispunham de adega;
- 1 (uma) pizzeria estava em reforma no período estabelecido para a montagem da base de dados; e,
- 1 (uma) pizzeria aceitou participar do estudo, mas ao decorrer do cronograma não demonstrou interesse em fornecer as informações necessárias para a composição da base de dados.

Definidos os estabelecimentos participantes do estudo, foram realizadas às visitas para o processo de coleta de dados. As informações coletadas foram retiradas dos cardápios da pizzeria e as informações do vinho, foram retiradas do rótulo e avaliação do *Sommelier*. Para o preenchimento dos atributos de tempo de amadurecimento, tipo de barrica e aromas dominantes do vinho, buscou-se as fichas técnicas de cada rótulo. Ao final desta coleta, a base de dados dispôs de 96 (noventa e seis) rótulos de vinhos e 32 (trinta e dois) sabores de pizzas.

Por se tratar de uma pesquisa qualitativa e o número de instâncias ser inferior a 100 (cem), se fez necessário a reunião de alguns atributos em conjunto de dados, especialmente os atributos voltados aos ingredientes da pizza e aos aromas possíveis do vinho, considerando que o baixo índice de determinado atributo não seria relevante para a pesquisa, visto que não seria possível encontrar os padrões com esta aplicação. A partir do quadro 20, é possível verificar a existência de 10 (dez) conjuntos de atributos dos ingredientes das pizzas, onde a junção foi realizada a partir de ingredientes que possuem similaridade. Neste agrupamento, destaca-se o atributo denominado embutidos, pois foi o atributo com maior número de ingredientes inseridos no conjunto.

QUADRO 20 – CONJUNTO DE ATRIBUTOS SIMILARES DE INGREDIENTES DAS PIZZAS

ATRIBUTO	COMPONENTES
Atum	Atum e Atum Branco
Azeite	Óleo e Azeite de Oliva
Azeitonas	Azeitona Verde e Azeitona Preta
Cebola	Cebola Roxa e Cebola Caramelizada
Embutidos	Calabresa, Calabresa Picante, Capocollo, Linguiça Toscana, Lombo Defumado, Panceta, Peito de Peru, Presunto, Presunto Defumado e Presunto de Parma
Manjericão	Basílico, Pesto de Basílico e Manjericão
Molho de Tomate	Molho de Tomate, Molho de Tomate Fresco e Molho de Tomate Italiano
Muçarela	Muçarela, Muçarela de Búfala, Muçarela de Burrata, Muçarela Especial e Muçarela Tradição
Pimentão	Pimentão e Pimentão Assado
Outros Queijos	Queijo de Cabra, Catupiry, Queijo Emmental, Queijo Gorgonzola e Queijo Parmesão
Tomates	Tomate, Tomate Cereja, Tomate Fresco e Tomate Seco

FONTE: A autora (2018).

Na formação do conjunto por similaridade dos aromas possíveis do vinho, alocou-se estes aromas em 19 (dezenove) atributos, onde o atributo denominado especiarias obteve-se destaque, pois foi o atributo formado com o maior número de componentes. Este agrupamento é representado pelo quadro 21.

QUADRO 21 – CONJUNTO DE ATRIBUTOS SIMILARES AOS AROMAS POSSIVEIS DO VINHO

ATRIBUTO	COMPONENTES
Ameixa	Ameixa e Ameixa Preta
Amora	Amora, Amora Preta e Amora Silvestre
Caramelo	Caramelo e Toffee
Cerâmica	Cerâmica e Porcelana
Cereja	Cereja, Cereja Negra e Xarope de Cereja
Especiarias	Alecrim, Canela, Cravo, Ervas Finas, Erva Doce, Especiarias, Folhas Secas, Hortelã, Orégano e Tomilho
Florais	Flor de Lima, Florais, Flores Brancas, Rosas e Violetas
Frutados	Frutado Frutas Maduras e Frutos Negros
Frutos	Frutos Negros e Frutos Silvestres
Geléia	Compotas, Geléia de Ameixa e Geléia de Frutas Vermelhas
Groselha	Groselha, Groselha Preta e Groselheira
Manteiga	Amanteigado e Manteiga
Melão	Melão e Melão Verde
Pêssego	Pêssego, Pêssego Branco e Pêssego em Calda
Pimenta	Pimenta, Pimenta Branca, Pimenta Preta, Pimentão e Tabasco
Pão	Pão e Pão Tostado
Tabaco	Tabaco, Tabaco Doce e Fumo de Corda
Tomate	Tomate e Tomate Assado
Toranja	Toranja e Toranja Branca

FONTE: A autora (2018).

Após agrupar os termos similares, foram excluídos aqueles atributos que não foram classificados em nenhum conjunto de atributos e que não seriam pertinentes para a análise devido ao baixo volume de ocorrências, como o ID de identificação das instâncias e os atributos relacionados aos aromas possíveis do vinho, ingredientes da pizza e variedade da uva e que obtiveram até 3 (três) ocorrências na base de dados, como demonstra o quadro 22.

QUADRO 22 – ATRIBUTOS DESCONSIDERADOS DA BASE DE DADOS

ATRIBUTOS	COMPONENTES
AROMAS POSSÍVEIS	Cerâmica, Cedro, Cranberry, Damasco, Eucalipto, Figo, Folhas Secas, Frutos, Madeira, Manteiga, Melão, Pão, Pedra, Tomate e Tostado
INGREDIENTES	Alecrim, Alho Poró, Azeite Balsâmico, Beringela, Erva Doce, Escarola, Palmito, Rúcula, Salsinha e Vinho Branco
UVA	Alicant Bouschet, Aragonez, Arinto, Barbera, Bonarda, <i>Brut Royal</i> , Castelão, Croatina, Gouveia, Graciano, <i>Gewutztraminer</i> , Loureiro, Malvasia Fina, Mazuelo, Montepulciano, Moscatel Galego, Nebbiolo, Pinotage, Primitivo, Viognier, Viosinho, Tannat, Tinta Barroca, Trajadura, Trincadeira e Zinfanel

FONTE: A autora (2018).

Os ingredientes comuns das pizzas também foram desconsiderados, visto que todas utilizam de massa, molho de tomate base e orégano. Além disso, foram excluídos os atributos que possuem o mesmo valor em todas as instâncias, como a doçura do vinho que em todos os rótulos de vinhos foram classificados como “seco”. Optou-se por utilizar apenas um atributo para representar o local de produção do vinho, escolhendo para a região de fabricação dos vinhos. Além disso, optou-se pela utilização de um único atributo para descrever a variedade do vinho que pode ser descrita através dos atributos: estilo e corpo, onde foi utilizado somente do atributo denominado estilo. Se deu esta escolha, pois os atributos (estilo e corpo do vinho) estão diretamente ligados por associação e quando aplicados ao processo de mineração de dados poderiam gerar regras evidentes. Já os atributos de preço e temperatura de serviço, foram transformados em escala baseado nos critérios de Puckette e Hammack (2016) disponíveis no livro “*Wine Folly: O guia essencial do vinho*” como demonstrado anteriormente no quadro 16. Porém, o atributo temperatura de serviço foi desconsiderado do estudo, visto que os registros com a temperatura cuja categoria é “fresco” estavam presentes 77,08% das instâncias e poderia induzir o estudo a conclusões irrelevantes.

3.2.2 Testes Estatísticos da Base de Dados

Para a realização da análise estatística da base de dados, utilizou o *software* RStudio na versão 1.1463, pois se trata de um *software* livre com ambiente de desenvolvimento integrado ao R, cuja finalidade é a execução da linguagem de programação para a realização de cálculos estatísticos e conseqüentemente, a geração de gráficos. Optou-se em utilizar a mesma base de dados para todas as aplicações, então estabeleceu a aplicação de testes voltados aos atributos nominais. Foram definidos a aplicação dos os seguintes testes: Qui-quadrado (χ^2) e a Análise de Correspondência Múltipla (ACM).

O teste Qui-quadrado (χ^2) tem como propósito realizar comparações entre as proporções, a fim de verificar possíveis divergências entre frequências. Destina-se em comparar a distribuição de uma amostragem com variáveis categóricas nominais e avaliar qualitativamente a existência da associação entre elas, não dependendo de parâmetros populacionais. Sua aplicação permite avaliar se as proporções observadas possuem diferenças significativas ou não possuem diferenças significativas quanto a proporção dos acontecimentos (CORREA; QUEIROZ; TREVISAN, 2015). Foram executados neste teste, os atributos de estilo de vinho distintos com maior índice de ocorrência na base de dados e os ingredientes da pizza mais comuns entre os 32 (trinta e dois) sabores de pizzas que compõe a base de dados. Como modelo de avaliação, utilizou-se o grau de confiança da aplicação de 95%, onde a probabilidade de significância (P-Valor) é dividida em duas formas de avaliação: Se P-Valor < 0,05 significa que a aplicação apresenta diferenças estatisticamente significativas; e, se P-Valor > 0,05 significa que a aplicação não apresenta diferenças estatisticamente significativas (NETTO, 2007).

Já no teste de Análise de Correspondência Múltipla (ACM), os dados são disponibilizados a partir de 2 (duas) ou mais variáveis categóricas. Estes dados são organizados em forma de matriz retangular, cujo propósito é representar graficamente as linhas e colunas, a fim de expressar geometricamente as relações entre objetos e variáveis. Aplicou-se neste teste, todos os atributos voltados aos ingredientes das pizzas em relação ao estilo dos vinhos. Neste caso, foram realizadas 3 (três) etapas com retirada dos atributos de menor ocorrência na base de dados, a fim de melhorar a visualização gráfica.

3.2.3 Mineração de Dados

Para a aplicação da mineração de dados foram determinadas as tarefas de mineração e os algoritmos que são capazes de satisfazer as necessidades da base de dados, de acordo com os atributos coletados. Considerando que a mineração de dados é voltada para a descoberta de padrões e verificar as propensões de regras, este estudo baseou-se no processo *Knowledge Discovery in Databases* (KDD), considerando o fato de que este permite a utilização de algoritmos para a realização de análise e descoberta de dados. As fases do processo KDD foram aplicadas da seguinte forma:

- a partir da literatura pertinente e de conversas com *Sommeliers* foram definidos os atributos que compuseram a base de dados;
- as etapas de seleção e dados-alvo abrangem a construção da base de dados Enogastronômica. Esta base de dados é composta de dados de pizzarias que possuem adegas e profissionais responsáveis pelas harmonizações, vencedoras em um dos Concursos analisados, sendo eles: Prêmio Veja Comer&Beber edição 2017/2018 ofertado pela revista Veja e Prêmio Bom Gourmet edição 2018 ofertado pelo Jornal Gazeta do Povo, e que se dispuseram a participar do estudo fornecendo as informações necessárias;
- os dados foram obtidos a partir de informações colhidas em entrevistas com o *Maitre* e *Sommelier* das pizzarias, sendo definido em conjunto a utilização de 3 (três) rótulos de vinho para cada sabor de pizza definido pelo entrevistado para realizar a harmonização. Estes dados foram coletados em planilhas eletrônicas (*Excel*) separadas em cada uma das pizzarias.
- por se tratar de uma pesquisa qualitativa, este estudo obteve o total de 96 (noventa e seis) instâncias, onde a aplicação da fase de pré-processamento e limpeza dos dados foi realizada a partir das ferramentas obtidas pelo *Excel*, como a função substituir. Os dados coletados foram reunidos em uma única planilha, onde se aplicou o uso de fórmulas para determinar as escalas de avaliação de determinados atributos e posteriormente, voltou-se para eliminar os dados redundantes e discrepantes ao conjunto de dados, isto é, os ruídos, como a acentuação de palavras e espaços desnecessários, além de transformar os dados numéricos coletados nas entrevistas em nominais

para poder realizar a aplicação dos algoritmos de mineração de dados voltados para a descoberta de padrões.

- para a execução dos procedimentos da mineração de dados foi utilizado a ferramenta denominada *Waikato Environment for Knowledge Analysis* (WEKA), visto que é um *software* de código aberto, programado em Linguagem Java que engloba o pré-processamento dos dados, classificação, regressão, agrupamento, regras de associação e visualização; além de ser o *software* utilizado na disciplina de mineração de dados do curso de Gestão da Informação.
- a partir das características do problema no qual este estudo pretende responder, avaliou-se quais os métodos seriam mais adequados para a base de dados estudada e que estão disponíveis no *software* WEKA. Optou-se por utilizar os algoritmos *Apriori* e *Prism*.
 - o algoritmo *Apriori* se enquadra na categoria descritiva, pois busca encontrar padrões que descrevem os dados de forma interpretável de modo não realiza uma busca aprofundada e gera conjuntos de itens (padrões) (FAYYAD et. al. 1996). Este algoritmo realiza a mineração de dados a partir de regras de associação, possibilitando observar os conjuntos frequentes de ocorrência (VASCONCELOS; CARVALHO, 2014). O *Apriori* é capaz de encontrar todos os conjuntos frequentes de ocorrência, permitindo resultados de implicações que consistem em mais de um item (ROMÃO et. al. 1999).
 - o algoritmo *Prism* é considerado de modelo preditivo, pois utiliza de técnicas matemáticas para encontrar relações entre as variáveis e os preditores, a fim de prever valores futuros da variável de destino (HALILI; RUSTEMI, 2016, tradução nossa). Este algoritmo é baseado no algoritmo de Árvore de Decisão (ID3), porém considera todos os atributos da base de dados como categóricos, até mesmo os contínuos, tendo como objetivo induzir regras de classificação a partir do conjunto de dados (BRANNER, 2007). O *Prism* possibilita fazer com que as regras sejam exibidas e interpretadas na ordem com que são cobertas, permitindo a eliminação de exemplos cobertos por regras conformadas (ARANDA; SOTOLONGO, 2013, tradução nossa);

- a etapa de conhecimento engloba a interpretação e avaliação dos resultados obtidos com a aplicação dos métodos escolhidos. Considerando a aplicação como forma de gerar regras de associação, se faz necessário abordar os critérios aplicados ao conjunto de dados, a fim de explicar os resultados obtidos, expressos a seguir:
 - para a aplicação do algoritmo *Apriori*, considerou a função “SE → ENTÃO” para gerar as regras de associação, optou-se por realizar testes nas configurações do método, como o suporte (percentual das transações entre atributos, indicando sua relevância), confiança (representa dentre as aplicações que possui determinado tipo de atributo, a porcentagem de ocorrência junto a outro atributo, indicando assim a validade da regra), o número de regras (valor máximo de regras gerado para cada aplicação) e o número de atributos antecedentes e consequentes que compõe a regra, a fim de observar quais seriam os melhores resultados. Considerou-se para a avaliação dos resultados, os seguintes valores: suporte = 0.1; confiança = 0.9 (90%); e, número de regras = 100, visto que a aplicação para valores maiores que 100 aplicados a estas configurações traziam o mesmo número de resultados, que neste caso foi de 88 (oitenta e oito) regras. Após as regras serem geradas, observaram-se quais foram capazes de comparar os ingredientes das pizzas com os atributos dos vinhos e as regras separadas exclusivamente por características dos vinhos ou das pizzas, observando o grau de interesse da regra, o denominado *lift*, onde quanto maior o valor considera-se mais provável a combinação da regra gerada.
 - para a execução do algoritmo *Prism*, se fez necessário preencher todas as células vazias com a ocorrência “não” visto que o algoritmo não desconsidera as instâncias que não possuem ocorrência. A aplicação do método visa a maximização de informação, aumentando a probabilidade de classificação, onde neste caso, foi aplicado aos componentes principais das pizzas e dos vinhos. Neste caso, aplicou-se a validação cruzada de 10 (dez) partições, sendo que a partir destes cortes é aplicada a fórmula de treinamento $X-1$, sendo X o valor da validação cruzada para a execução dos testes, onde o processo é

repetido X vezes para utilizar todos os conjuntos de testes e ao final calcular a média dos resultados de cada ocorrência. Nesta aplicação, os resultados voltados aos ingredientes das pizzas e aos tipos de uvas utilizados na produção de vinhos obtiveram o mínimo de 63,54% da taxa de acerto na classificação das instâncias, desconsiderando a ocorrência de equilíbrio entre as opções disponíveis; enquanto os atributos considerados para o questionário de validação da proposta, possuem no mínimo 85,42% de taxa de acerto e equilíbrio entre as variáveis; e, para os atributos voltados aos aromas possíveis do vinho, a menor taxa de acerto foi de 34,38%, porém a execução com estes atributos como atributo-meta foram desconsiderados, devido a discrepância entre as ocorrências.

- a aplicação dos 2 (dois) algoritmos de mineração de dados, gerou um conjunto de resultados. Dentre estes resultados separou-se 20 (vinte) regras, sendo estas escolhidas a partir dos antecedentes e consequentes e grau de interesse da regra do *Apriori* e do equilíbrio da taxa de acerto na matriz de confusão e índice de acerto das classificações em instâncias para o *Prism*. Estas regras compuseram o questionário de validação da proposta, onde considerou o seguinte modelo de avaliação:
 - **não se aplica** (o avaliador não possui nenhum conhecimento sobre a situação descrita);
 - **não concordo com esta situação** (o avaliador possui conhecimento sobre a situação, mas não concorda com o resultado);
 - **esta situação é comum** (o avaliador possui conhecimento sobre a situação e a trata como uma combinação comum no ato de uma harmonização);
 - **esta situação é surpreendente** (o avaliador não possui conhecimento sobre a situação, mas acredita que a combinação pode ocorrer, gerando um novo conhecimento).

A aplicação deste questionário foi destinada aos profissionais das áreas de Gastronomia e Enologia/*Sommelier*, com o propósito verificar se houve a comprovação de regras já existentes na Enogastronomia ou se a aplicação destes padrões gerou um novo conhecimento. Desta forma, o avaliador demarca com um único “X” dentre as opções disponíveis na escala de

avaliação, em qual modelo de avaliação a situação descrita se enquadra de acordo com sua experiência e conhecimento. O modelo do questionário aplicado pode ser observado no Apêndice 1.

Na próxima seção, apresentam-se os resultados referentes a base de dados e a aplicação dos testes estatísticos e de mineração de dados.

4 RESULTADOS E ANÁLISES

Nesta seção, estão descritos os resultados obtidos a partir da análise estatística da base de dados (Qui-quadrado (χ^2) e Análise de Correspondência Múltipla (ACM)) e da aplicação dos algoritmos *Apriori* e *Prism* de mineração de dados.

4.1 BASE DE DADOS

A base de dados Enogastronômica conta com 66 (sessenta e seis) atributos, considerando a desmembração realizada entre os ingredientes da pizza, tipos de uvas que compõem o vinho e os aromas possíveis no ato da degustação. Destes atributos, 15 (quinze) são voltados aos ingredientes que compõem as pizzas e 51 (cinquenta e um) atributos estão relacionados às características do vinho, tais como, valor, região, uvas, estilo, acidez, doçura, tanino, nível alcóolico, tempo e tipo de barrica, temperatura de serviço do vinho e os aromas possíveis.

Dentre os sabores de pizzas disponíveis entre as pizzarias que participaram deste estudo, alocou-se a harmonização de vinhos 32 (trinta e dois) sabores de pizzas, sendo 14 (quatorze) sabores da pizzaria denominada “A” e 18 (dezoito) sabores da pizzaria denominada “B”, como demonstra o quadro 23.

QUADRO 23 – SABORES DAS PIZZAS HARMONIZADAS

PIZZARIA	NOME DA PIZZA	PIZZARIA	NOME DA PIZZA
A	Alcachofras	B	Albina
A	Amalfitana	B	Aliche
A	Atum	B	Atum
A	Carciofil Parma	B	Balila
A	Emmental e Capocollo	B	Caetano Pinto
A	Fiorella	B	Catupiry
A	La Madonna Medievale	B	Classe A
A	Marguerita DOC	B	Diavola
A	Palmito Concasse	B	Finocchio
A	Parma	B	Florita
A	Parma I Capri	B	Germينو
A	Pastorela	B	Gorgonzola
A	Portena	B	Juventus
A	Sinfonia Di Saperi	B	Marcolina
		B	Peixão
		B	Piratininga
		B	Vila Judite
		B	Zola e Funghi

FONTE: A autora (2018).

Entre os sabores de pizzas, destacou-se as pizzas que contém como ingrediente o atributo denominado muçarela (muçarela, muçarela de búfala, muçarela de burrata, muçarela especial e muçarela tradição), que contou com 60 (sessenta) ocorrências e a categoria embutidos (calabresa, calabresa picante, capocollo, linguiça toscana, lombo defumado, panceta, peito de peru, presunto, presunto defumado e presunto de parma) que apresentou 48 (quarenta e oito) ocorrências, como demonstra o quadro 24.

QUADRO 24 – OCORRÊNCIA NAS INSTÂNCIAS QUANTO AOS INGREDIENTES NA COMPOSIÇÃO DAS PIZZAS

INGREDIENTE	OCORRÊNCIA
Alcachofra	9
Aliche	6
Atum	24
Azeite	6
Azeitona	21
Cebola	6
Embutidos	48
Manjericão	27
Molho de Tomate	30
Muçarela	60
Outros Queijos	36

FONTE: A autora (2018).

Para cada sabor de pizza, foram selecionados 3 (três) rótulos de vinho. Esta base de dados é composta por 86 (oitenta e seis) rótulos de vinhos distintos, porém houve repetição de rótulos como o caso do Vinho Tinto Manz Contador de Estória e o Vinho Branco Saint Clair, que foram recomendados para 3 (três) sabores distintos de pizza; e, os Vinhos Tintos Porta 6, Ramirana Reserva e Santa Ema Gran Reserva e os Vinhos Brancos Muros de Vinha e Trevo que foram sugeridos para 2 (dois) sabores de pizza. A lista completa dos rótulos que compõe a base de dados, estão disponíveis no Apêndice B.

Estes rótulos são diversificados entre diferentes países que produzem vinhos, havendo maior ocorrência de vinhos Argentinos da região de Mendoza, Chile da região do Valle Del Maipo e Portugal com a região de Douro. A lista completa de países e regiões de produção e fabricação dos vinhos está disponível no Apêndice C.

Quanto ao valor de comercialização das garrafas de vinho entre as pizzarias, o maior índice de ocorrência foi voltado aos vinhos com valores pertencentes a faixa de preço de “B” (Entre R\$60,01 e R\$100,00) que representou 30,21% dos rótulos

disponíveis e “C” (Entre R\$100,01 e R\$170,00) contemplando 51,04% dos rótulos disponíveis na base de dados. A faixa de preço completa dos rótulos analisados é apresentada no quadro 25.

QUADRO 25 – FAIXA DE PREÇO DE COMERCIALIZAÇÃO DOS VINHOS

FAIXA DE PREÇO	OCORRÊNCIA
A - Até R\$60,00;	0
B - Entre R\$60,01 e R\$100,00	29
C - Entre R\$100,01 e R\$170,00	49
D - Entre R\$170,01 e R\$250,00	12
E - Mais que R\$250,01	6

FONTE: A autora (2018).

Considerando o aspecto estilo do vinho, dentre os 9 (nove) estilos apresentados de acordo com Puckette e Hammack (2016), a base de dados contemplou 8 (oito), visto que as pizzarias participantes do estudo são tradicionais e não trabalham com pizzas doces para realizar a harmonização com vinhos de sobremesa disponíveis na carta da casa. Assim, destacou-se o estilo de vinho tinto encorpado contemplado por 37,50% dos rótulos disponíveis na base de dados; o estilo de vinho tinto de médio corpo representado por 33,33% dos rótulos; e, o estilo de vinho branco leve com 10,41% de representatividade dentre os rótulos disponíveis. A ocorrência de todos os estilos de vinhos que constituem a base de dados é demonstrada no quadro 26.

QUADRO 26 – ESTILO DO VINHO

ESTILO	OCORRÊNCIA
Espumante	3
Vinho Branco Aromático	2
Vinho Branco Encorpado	2
Vinho Branco Leve	10
Vinho Rosé	2
Vinho Tinto de Médio Corpo	32
Vinho Tinto Encorpado	36
Vinho Tinto Leve	9

FONTE: A autora (2018).

Dentre os diferentes estilos de vinhos disponíveis, contemplaram a composição destes com mais de 3 (três) ocorrências 15 (quinze) tipos de uvas distintas, destacando-se as uvas Cabernet Sauvignon (15,62%) e Malbec (14,06%), visto que

são uvas que empregadas na fabricação de vinhos tintos, que neste caso houve maior ocorrência em relação a variedade dos vinhos espumantes, brancos e rosé. O quadro 26 demonstra a ocorrência dos demais resultados.

QUADRO 27 – VARIEDADE DE UVAS UTILIZADAS NA FABRICAÇÃO DOS VINHOS

UVAS	OCORRÊNCIA
Cabernet Franc	5
Cabernet Sauvignon	20
Carmenere	8
Chardonnay	5
Malbec	18
Merlot	7
Petit Verdot	5
Pinot Noir	8
Sangiovese	6
Sauvignon Blanc	5
Sirah	11
Tempranillo	6
Tinta Roriz	8
Touriga Franca	5
Touriga Nacional	11

FONTE: A autora (2018).

Analisando as características de composição do vinho, deve-se levar em consideração a doçura, corpo, acidez do vinho. Todos os vinhos da base de dados foram classificados com doçura equivalente a categoria “seco”, isto é, o açúcar presente na bebida é residual, contemplado após o processo de fermentação e amadurecimento. Já o corpo do vinho, houve uma distribuição variada entre os diferentes rótulos, visto que leva em consideração a junção das características de composição do vinho. Quanto ao atributo tanino, deve-se levar em conta que este é comum nos vinhos tinto e rosé, pois esta característica é dada a partir das cascas das uvas e em barricas de madeira; geralmente os vinhos brancos e espumantes não se enquadram nesta categoria, porque são fermentados sem as cascas da uva. Destacaram-se os taninos médio e adstringente, pois o grau de tanicidade do vinho é capaz de limpar o paladar para pratos com carnes gordurosas, queijos e massas (PUCKETTE; HAMMACK, 2016). O álcool é o responsável por transportar o aroma do vinho até o olfato, sendo assim, dentre os rótulos avaliados a quantidade de teor alcoólico se enquadra entre as categorias: baixo e médio (10,0% até 11,5%); médio (11,5% até 13,5%); e, entre médio e alto (13,5% até 15,0%) por garrafa. No caso desta

aplicação, o índice de maior ocorrência de álcool esteve presente na categoria entre alto e médio, representando 69,79% de ocorrência dentre os rótulos analisados. O quadro a seguir, representa a avaliação das características de composição do vinho.

QUADRO 28 – CARACTERÍSTICAS DE COMPOSIÇÃO DO VINHO

CORPO	OCORRÊNCIA
Leve	21
Médio	35
Entre Médio e Encorpado	16
Encorpado	24

ACIDEZ	OCORRÊNCIA
Baixa	6
Entre Baixa e Média	3
Média	42
Alta	42
Muito Alta	3

TANINO	OCORRÊNCIA
Não Possui	17
Baixo	10
Entre Baixo e Médio	6
Médio	27
Adstringente	24
Muito Adstringente	12

TEOR ALCÓOLICO	OCORRÊNCIA
Onze	2
Doze	2
Treze	25
Quatorze	58
Quinze	9

FONTE: A autora (2018).

O tempo de amadurecimento e a barrica utilizada, é o momento onde ocorre a transformação de cor, aroma e sabor de cada rótulo. Dentre as opções de meses listadas para o tempo de amadurecimento dos vinhos, o maior índice foi de 15 (quinze) ocorrências ao tempo de amadurecimento de 12 (doze) meses, seguidos por 11 (onze) e 9 (nove) meses de amadurecimento, nos quais foram listados 6 (seis) vezes cada. Para o atributo denominado “tipo de barrica”, 51,4% dos rótulos passam pelo processo de amadurecimento em barricas de carvalho, pois este tipo de barrica tem a capacidade de aumentar a complexidade do vinho (COPELLO, 2016). Além disso, 28 (vinte e oito) instâncias deste atributo não foram identificados a partir das buscas realizadas nas fichas técnicas de cada rótulo. A lista completa de amadurecimento e tipo de barrica utilizada é apresentada no Apêndice D.

Em questão de temperatura de serviço do vinho, 74 (setenta e quatro) rótulos devem ser servidos em temperatura fresco (16° a 20°C); seguidos da temperatura frio (7° a 13°C) com 16 (dezesesseis) ocorrências; temperatura de adega (13° a 16°C) referentes à 4 (quatro) rótulos; e, gelado (3° a 7°C) com apenas 2 (duas) ocorrências. Cada estilo de vinho possui uma temperatura de serviço, geralmente os vinhos com temperatura fria e gelada estão associados aos espumantes, vinhos brancos e rosé, a fim de trazer mais frescor à degustação, enquanto os vinhos fresco e em temperatura de adega são destinados aos vinhos tintos, pois são mais estruturados devido a presença do tanino (CONFRARIA DOS SOMMELIERS, 2008?).

No ato da degustação do vinho é possível sentir determinados aromas provenientes das características do vinho, como tempo de amadurecimento e tipo de barrica utilizada no processo de armazenamento da bebida. Ao final do tratamento da base de dados e eliminação dos atributos que continham valor igual ou menor que 3 (três) ocorrências, totalizaram para a categoria de aromas possíveis do vinho 26 (vinte e seis) odores diferentes entre especiarias, florais, frutas/frutados, frutas secas, geleias, minerais e odores específicos como baunilha, café, caramelo, chocolate, couro e pimenta. São considerados odores mais comuns entre os rótulos, os aromas de especiarias e florais com 43 (quarenta e três) ocorrências cada; seguidos das frutas ameixa, cereja e amora que contaram com 39 (trinta e nove), 38 (trinta e oito) e 33 (trinta e três) ocorrências respectivamente. No Apêndice E, são listados todos os aromas possíveis identificados a partir das fichas técnicas de cada rótulo.

4.1.1 Análise Estatística da Base de Dados

A fim de realizar uma análise descritiva dos dados e verificar a ocorrência de associação entre as variáveis, optou-se por utilizar a aplicação de 2 (dois) testes estatísticos, sendo eles: Qui-quadrado (χ^2), com o propósito de verificar alguma semelhança entre as proporções; e, a Análise de Correspondência Múltipla (ACM), pois é capaz de explorar relações com mais de 2 (duas) variáveis e expressa os resultados de forma gráfica, facilitando a compreensão.

4.1.1.1 Qui-quadrado (χ^2)

Na execução do teste Qui-quadrado (χ^2), optou-se em utilizar os estilos de vinho que possuíam maior ocorrência e apresentassem características distintas, sendo eles: o estilo “vinho branco leve” com 10 (dez) instâncias e o “vinho tinto encorpado” presente em 36 (trinta e seis) instâncias; enquanto para os ingredientes das pizzas, optou-se por utilizar os atributos com maior presença na base de dados, sendo eles: os embutidos, molho de tomate, muçarela e outros queijos.

Para a aplicação considerando o atributo embutidos, neste conjunto de dados 26 (vinte e seis) instâncias possuem como ingredientes os embutidos, enquanto 20 (vinte) instâncias não possuem. Os resultados são apresentados no quadro 29.

QUADRO 29 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE EMBUTIDOS

ESTILO DO VINHO	EMBUTIDOS	
	SIM	NÃO
Vinho Branco Leve	10,00%	90,00%
Vinho Tinto Encorpado	69,44%	30,56%

FONTE: A autora (2018).

De acordo com o quadro acima, apenas 10,00% das pizzas que possuem embutidos como ingredientes são sugeridos vinho com estilo branco leve; enquanto 69,44% das sugestões são voltadas ao vinho tinto encorpado. Deste modo, é possível identificar diferenças estaticamente significativas nas proporções, considerando Qui-quadrado ($\chi^2(1) = 8,9643$, P-Valor < 0,0027).

Já a aplicação com o atributo denominado molho de tomate, apresenta 16 (dezesesseis) instâncias que possuem o ingrediente molho de tomate nas pizzas e 30 (trinta) instâncias não possuem este ingrediente. É possível verificar os resultados no quadro 30.

QUADRO 30 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE MOLHO DE TOMATE

ESTILO VINHO	MOLHO DE TOMATE	
	SIM	NÃO
Vinho Branco Leve	50,00%	50,00%
Vinho Tinto Encorpado	30,56%	69,44%

FONTE: A autora (2018).

Conforme o quadro anterior nota-se que não é possível identificar proporções estatísticas, pois para o estilo do vinho branco leve houve 50,00% de ocorrência para cada uma das opções disponíveis no atributo. Os resultados obtidos desta aplicação não apresentam diferenças significativas, considerando os seguintes resultados: Qui-quadrado ($\chi^2(1) = 0,58804$, P-Valor < 0,4432).

A aplicação do conjunto de dados muçarela, contou com 27 (vinte e sete) ocorrências para a opção “esta pizza contém este ingrediente” e 19 (dezenove) ocorrências para a opção “esta pizza não contém este ingrediente”. Esta execução é apresentada no quadro 31.

QUADRO 31 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE MUÇARELA

ESTILO VINHO	MUÇARELA	
	SIM	NÃO
Vinho Branco Leve	50,00%	50,00%
Vinho Tinto Encorpado	61,11%	38,89%

FONTE: A autora (2018).

Assim, como o caso do atributo molho de tomate, o atributo muçarela também obteve 50,00% em cada uma das opções disponíveis para o estilo de vinho branco leve. Já para o estilo de vinho tinto encorpado é recomendado em 61,11% dos casos. Verificando os valores de Qui-quadrado ($\chi^2(1) = 0,071984$, P-Valor < 0,7885, é possível confirmar que não existem diferenças estatisticamente significativas nesta aplicação.

Já a execução com o atributo denominado outros queijos, apresentou 20 (vinte) ocorrências para a opção “contém este ingrediente na pizza” e 26 (vinte e seis) para “não contém este ingrediente na pizza”. Os resultados desta aplicação são demonstrados no quadro 32.

QUADRO 32 – COMPARAÇÃO DE PROPORÇÕES ENTRE VINHOS BRANCO LEVE E TINTO ENCORPADO COM O INGREDIENTE OUTROS QUEIJOS

ESTILO VINHO	OUTROS QUEIJOS	
	SIM	NÃO
Vinho Branco Leve	20,00%	80,00%
Vinho Tinto Encorpado	50,00%	50,00%

FONTE: A autora (2018).

Para esta execução, é possível visualizar que dentre as recomendações de vinhos com pizzas que possuem o conjunto de dados “outros queijos” apenas 20,00% são voltadas para o estilo de vinho branco leve; enquanto o estilo de vinho encorpado, fica dividido entre as opções disponíveis. A partir dos resultados obtidos: Qui-quadrado ($\chi^2(1) = 1,7754$, P-Valor < 0,1827, comprova-se a não existência de diferenças estatisticamente significativas entre este atributo, considerando ainda que a sugestão do estilo de vinho branco é 21,74% menor do que a ocorrência de sugestão do vinho tinto encorpado.

4.1.1.2 Análise de Correspondência Múltipla (ACM)

Para a aplicação do teste ACM, foram selecionados os atributos da base de dados que pertencem aos ingredientes das pizzas e o estilo do vinho, visando a geração de uma relação harmônica entre estes elementos. O teste foi realizado em 3 (três) etapas: a primeira etapa, foram utilizados todos os atributos de ingredientes (alcachofra, aliche, atum, azeite, azeitona, cebola, embutidos, manjeriço, muçarela, molho de tomate, outros queijos, pimentão, tomates), porém não foi possível identificar agrupamentos, devido a um acúmulo de atributos existente; optou-se então em retirar os atributos próximos ao centro do gráfico na etapa seguinte, sendo eles: azeite, cebola e pimentão, pois eram os atributos com menor índice de ocorrência na base de dados; nesta aplicação a visualização melhorou, porém não foi possível identificar nenhum agrupamento; por fim, realizou uma nova eliminação dos ingredientes situados no centro do gráfico (aliche, atum, azeitona, molho de tomate e cogumelo paris), onde foi possível identificar a existência de agrupamentos no gráfico.

Neste teste, aplicou-se a regra de “autovalor” (variância padronizada associada a um fator particular) a fim de verificar a quantidade de dimensões geradas para explicar o método utilizado nesta base de dados, onde foram geradas 13 (treze) dimensões. A figura 15 demonstra o percentual visível da aplicação em cada uma das dimensões.

FIGURA 16 – PERCENTUAL ACUMULADO ENTRE AS DIMENSÕES

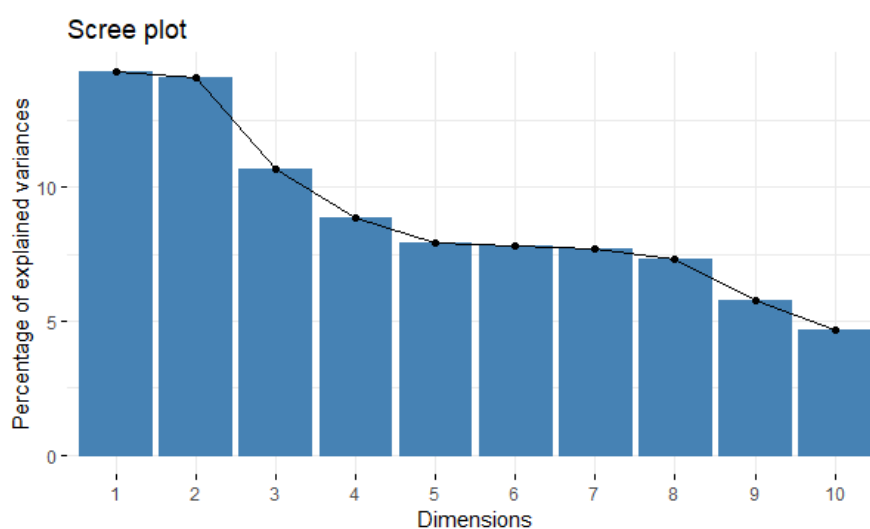
	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	0.26529152	14.284928	14.28493
Dim.2	0.26083529	14.044977	28.32991
Dim.3	0.19833175	10.679402	39.00931
Dim.4	0.16460204	8.863187	47.87249
Dim.5	0.14667147	7.897694	55.77019
Dim.6	0.14474829	7.794139	63.56433
Dim.7	0.14285714	7.692308	71.25663
Dim.8	0.13566305	7.304933	78.56157
Dim.9	0.10743090	5.784741	84.34631
Dim.10	0.08724854	4.697998	89.04431
Dim.11	0.07709620	4.151334	93.19564
Dim.12	0.07456413	4.014992	97.21063
Dim.13	0.05180255	2.789368	100.00000

FONTE: Dados da pesquisa (2018).

No gráfico 1 é possível verificar graficamente o resultado desta aplicação, onde nota-se que das 13 (treze) dimensões, as duas primeiras colunas estão muito

próximas. Analisando a ocorrência das dimensões é verificar que na segunda dimensão é possível observar 28,33% do método, facilitando a explicação do usuário do material visualizado. A partir da terceira dimensão, os valores percentuais variam de 3,00% a 11,00% da observação da aplicação, não sendo vantajoso escolher mais uma dimensão para representar a explicação da visualização, pois o torna mais complexo.

GRÁFICO 1 – DIMENSÕES DA APLICAÇÃO ACM



FONTE: Dados da pesquisa (2018).

O gráfico 2, é uma aplicação de nuvem de pontos, demonstra a relação trazida entre os ingredientes das pizzas e o estilo dos vinhos. Os círculos do gráfico demarcam cada uma das relações.

GRÁFICO 2 – AGRUPAMENTO DAS VARIÁVEIS GERADAS



FONTE: Dados da pesquisa (2018).

É possível observar a existência de 4 (quatro) agrupamentos distintos de harmonização entre os ingredientes das pizzas e o estilo do vinho, sendo eles:

- **vinho rosé:** não harmoniza com tomates e manjeriço;
- **vinho tinto leve:** harmoniza com tomates e manjeriço;
- **vinho tinto de médio corpo:** harmoniza com muçarela, mas não harmoniza com alcachofra, outros queijos e embutidos;
- **vinho tinto encorpado:** harmoniza com embutidos e outros queijos, mas não harmoniza com muçarela.

As pizzas que contêm como ingrediente a alcachofra e os vinhos que pertencem ao estilo de vinho branco (leve, aromático e encorpado) e estilo espumante, não foram inseridas em nenhum dos outros agrupamentos identificados.

4.2 MINERAÇÃO DE DADOS

A fim de encontrar padrões existentes na base dados, aplicou-se 2 (dois) algoritmos de mineração de dados: O *Apriori*, que é um algoritmo de associação e o *Prism* que é um algoritmo de classificação. Os resultados obtidos a partir destas duas aplicações são apresentados a seguir.

4.2.1 Algoritmo *Apriori*

Para a aplicação do algoritmo *Apriori* foram definidos como parâmetros iniciais: suporte de 0,1; confiança de 0,9 (90%); e, número de regras até 100 (cem). Este número de regras foi mantido ao longo de todos os experimentos, visto que, o número máximo de regras descobertas foi de 88 (oitenta e oito). Para este método, não foi necessário preenchimento das células vazias, uma vez que o *Apriori* tem a opção (por *default*) de ignorá-las.

Dentre as regras geradas, analisaram-se os valores de *lift*, uma vez que representa a medida de grau de interesse da regra e avalia a importância da mesma. Todos os valores de *lift* entre as regras descobertas foram superiores à 1 (um), demonstrando uma correlação positiva entre os atributos. O valor inicial de *lift* gerado durante esta execução foi de *lift* igual a 1,45. Utilizou-se o modelo de avaliação de regra de interesse proposto por Freitas (1998, tradução nossa), conhecido como modelo subjetivo ou orientado pelo usuário, onde o grau de interesse das regras é dado a partir do conhecimento prévio e das crenças do usuário, isto é, se ele encontra fatores interessantes na regra gerada que podem agregar às suas crenças e até mesmo gerar um novo conhecimento.

Assim, foram escolhidas 8 (oito) regras para compor o questionário de validação da proposta. Priorizaram-se as regras que contemplassem a combinação de ingredientes das pizzas, ingredientes das pizzas com atributos relacionados às características dos vinhos e atributos que envolvem apenas as características dos vinhos. Dentre as regras selecionadas, uma delas apresenta um atributo no antecedente e outro no conseqüente da regra:

- **SE** Vinho Possui Uva = Touriga Nacional **ENTÃO** Aroma do Vinho = Florais 10 → <conf:(0,91)> *lift*: (2,03)

Outras 6 (seis) regras contemplam duas condições no antecedente da regra:

- **SE Pizza** Possui Ingrediente = Outros Queijos **E** Pizza Possui Tempero = Manjericão **ENTÃO** Pizza Possui Ingrediente = Embutidos 12 → <conf:(1)> *lift*: (2)

- **SE** Vinho Possui Uva = Malbec **E** Barrica = Carvalho **ENTÃO** Pizza Possui Ingrediente = Embutidos 11 → <conf:(1)> *lift*: (2)
- **SE** Vinho Possui Tanino = Muito Adstringente **E** Teor Alcoólico do Vinho = **Quatorze ENTÃO** Estilo do Vinho = Vinho Tinto Encorpado 11 → <conf: (1)> *lift*: (2,67)
- **SE** Aroma do Vinho = Ameixa **E** Aroma do Vinho = Baunilha **ENTÃO** Pizza Possui **Ingrediente** = Muçarela 11 → <conf:(0,92)> *lift*: (1,47)
- **SE** Estilo do Vinho = Vinho Tinto Encorpado **E** Acidez do Vinho = Média **ENTÃO** Pizza Possui Ingrediente = Embutidos 11 → <conf:(0,92)> *lift*: (1,83)
- **SE** Pizza **Possui** Ingrediente = Embutidos **E** Região de Produção do Vinho = Mendoza **ENTÃO** Vinho Possui Uva = Malbec 10 → <conf:(0,91)> *lift*: (4,85)

Finalmente, uma regra selecionada apresentou 3 (três) atributos no antecedente e 2 (dois) no conseqüente. Desta forma, relacionaram em uma mesma regra, cinco atributos, sendo um da pizza e os demais do vinho.

- **SE** Pizza Possui Ingrediente = Muçarela **E** Estilo de Vinho = Vinho Tinto Encorpado **E** Acidez do Vinho = Alta **ENTÃO** Faixa de Preço do Vinho = C (Entre R\$100,01 à R\$170,00) **E** Barrica = Carvalho 12 → <conf:(0,92)> *lift*: (2,95)

Em regras que relacionam mais que 2 (dois) atributos (um no antecedente e outro no conseqüente) existem vantagens e desvantagens que devem ser analisadas. Como vantagem pode se levar em consideração que quanto mais atributos a regra possui, mais específica ela se torna, onde se tem uma previsão com maior certeza, devido ao volume de atributos dispostos para prever algo. Porém, por se tratar de uma regra “muito específica” pode ocorrer o processo de *overfitting*, isto é, quando o conjunto de atributos observados na regra é ajustada a ela, mas se mostra ineficaz para prever novos resultados, visto que por ser tão específica ela pode ocorrer somente uma vez.

4.2.2 Algoritmo *Prism*

Para a aplicação do algoritmo *Prism*, foi necessário preencher todas as instâncias vazias da base de dados com a descrição “não”, uma vez que este método não possui a opção de tratar os valores vazios (desconhecidos). Após este procedimento, foram realizados testes de validação cruzada com 10 (dez) partições, sendo eles: os ingredientes das pizzas, características do vinho (preço, região, estilo, acidez, corpo e tanino), tipos de uvas que compõe os rótulos e aromas possíveis do vinho.

A partir da validação cruzada com 10 (dez) partições, analisou-se a matriz de confusão de cada execução, a fim de verificar qual continha a maior cobertura de atributos e encontrava-se em equilíbrio de ocorrências entre “sim” e “não”. Por exemplo, o atributo denominado embutido que contempla 48 (quarenta e oito) ocorrências para cada uma das opções e o atributo muçarela que apresenta 60 (sessenta) ocorrências para a opção “contém muçarela” e 36 (trinta e seis) ocorrências para a opção “não contém muçarela”.

De início, foram desconsiderados os atributos que não apresentaram nenhuma regra. Dentre eles, está o atributo denominado faixa de preço dos vinhos, que obteve 40,63% de taxa de acerto; o atributo região de produção do vinho com apenas 14,58% de taxa de acerto; e, o estilo do vinho com 40,63% de taxa de acerto. Todos os atributos relacionados aos aromas possíveis do vinho foram eliminados, pois traziam aproximadamente 50% de taxa de acerto e apresentavam grande discrepância quanto as ocorrências. A aplicação da tarefa de classificação, considerando os diferentes tipos de uva do vinho, retornaram resultados a partir de a 63,54%, referentes a taxa de acerto. No entanto, verificou-se que isto ocorria porque exista a classificação pela maioria. Por exemplo, em um atributo com 92 (95,83%) ocorrências de um determinado tipo e 4 (4,17%) de outro, a taxa de acerto era de 91,67% pois o método retornava à maioria como classe prevista.

Considerou como atributo meta a aplicação dos diferentes ingredientes que compõem as pizzas, que apresentavam uma proporção equilibrada de ocorrências e que continham taxa de acerto superior a 85%. Por exemplo, o atributo denominado “outros queijos” apresenta 36 (37,50%) ocorrências para “contém outros queijos” e 60 (62,50%) para “não contém outros queijos”, onde a taxa de acerto foi de 86,46%. Foi desconsiderada a aplicação como atributo meta de tais ingredientes das pizzas: atum,

alcachofra, aliche, cebola, pimentão e azeitona, pois atendiam aos requisitos estabelecidos.

Dentre as aplicações dos seguintes atributos: embutidos, molho de tomate, muçarela, outros queijos e tomate, definiu-se 12 (doze) regras para o processo de validação de resultados junto aos especialistas. A escolha das regras envolveu a harmonização ou não do conjunto de ingredientes, ingredientes das pizzas e vinhos e características dos vinhos. Não foram escolhidas as regras geradas envolvendo o nome das pizzas e os nomes dos vinhos, visto que o propósito é buscar padrões entre os componentes e a nomenclatura das pizzas e dos rótulos de vinhos servem para identifica-los.

A partir da aplicação do algoritmo *Prism*, com o atributo meta ingrediente sendo o molho de tomate, 93 instâncias foram classificadas, sendo 27 (vinte e sete) ocorrências para “possui o ingrediente”, contra 66 (sessenta e seis) para “não possui o ingrediente”, apresentando 96,88% de classificações corretas; e, 3 (três) instâncias não classificadas, totalizando 3,12%. Esta aplicação gerou 30 (trinta) regras, sendo que as regras escolhidas para a aplicação do questionário de validação da proposta foram:

- **SE Pizza Possui Ingrediente = Pimentão ENTÃO Pizza Possui Ingrediente= Molho de Tomate**
- **SE Estilo do Vinho = Vinho Branco Encorpado ENTÃO Pizza Possui Ingrediente= Molho de Tomate**
- **SE Pizza Possui Ingrediente = Alcachofra ENTÃO Pizza Não Possui Ingrediente = Molho de Tomate**
- **SE Vinho Possui Uva = Touriga Franca ENTÃO Pizza Não Possui Ingrediente = Molho de Tomate**

Quanto aos resultados apresentados, quando o atributo meta foi o conjunto de dados denominado muçarela, contou com 90 (noventa) instâncias classificadas corretamente, sendo 36 (trinta e seis) para a opção “contém muçarela” e 54 (cinquenta e quatro) para a opção “não contém muçarela”, totalizando uma taxa de acerto de 93,75%; 1 (uma) instância foi classificada incorretamente, correspondendo a 1,04%; e 5 (cinco) instâncias não foram classificadas, representando 5,21%. Ao total foram geradas 24 (vinte e quatro) regras, sendo que (vinte e três) delas está associada ao

nome da pizza. Então, escolheu-se para demonstração, a regra que associa as características do vinho e a ocorrência do ingrediente muçarela.

- **SE** Acidez do Vinho = Baixa **ENTÃO** Pizza Possui Ingrediente = Muçarela

Já para o atributo que representa o conjunto embutidos foram geradas 31 (trinta e uma) regras, sendo 24 (vinte e quatro) regras está relacionada ao nome da pizza. Esta aplicação resultou em 83 (oitenta e três) instâncias avaliadas corretamente, representando 86,46%; 2 (duas) instâncias foram avaliadas incorretamente, correspondendo à 2,08%; e, 11 (onze) instâncias não foram classificadas, sendo equivalente a 11,46% da aplicação. Entre as regras escolhidas, foi selecionada uma regra que possui ocorrência de harmonização e outra que afirma a não ocorrência da harmonização, como demonstrado a seguir.

- **SE** Tempo de Amadurecimento = Treze Meses **ENTÃO** Pizza Possui **Ingrediente** = Embutidos
- **SE** Pizza Possui **Ingrediente** = Atum **ENTÃO** Pizza Não Possui Ingrediente = Embutidos

A utilização do conjunto de dados denominado outros queijos como atributo meta, resultou em 28 (vinte e oito) regras sendo que 20 (vinte) delas correspondem à denominação da pizza. Totalizou dentre as classificações corretas 83 (oitenta e três) instâncias correspondendo a 86,46% da aplicação, sendo 58 (cinquenta e oito) para a opção “não possui este ingrediente” e 25 para a opção “possui” este ingrediente; 3 (três) instâncias foram classificadas erradas, totalizando 3,12%; e, 10 (dez) instâncias não foram classificadas, sendo representadas por 10,42% da aplicação. Dentre as regras selecionadas, estão relacionados a ocorrência ou não do atributo meta, combinação com demais ingredientes da pizza, características do vinho e a relação com faixa de preço do vinho, visto que esta categoria possui queijos são mais nobres do que os queijos muçarelas, por exemplo.

- **SE** Pizza Possui Ingrediente = Aliche **ENTÃO** Pizza Não Possui Ingrediente = Outros Queijos

- **SE Acidez** do Vinho = Baixa **ENTÃO** Pizza Não Possui Ingrediente = Outros Queijos
- **SE Faixa de Preço** do Vinho = E (Mais que R\$250,01) **ENTÃO** Pizza Possui Ingrediente = Outros Queijos

Considerando como atributo meta o ingrediente tomate, observa-se que sua aplicação foi a que gerou mais regras dentro do contexto analisado (32 regras), porém foi a aplicação que resultou no menor índice de classificações corretas representado por 85,42%, com 82 (oitenta e duas) instâncias, sendo 24 (vinte e quatro) para a ocorrência “possui tomate” e 58 (cinquenta e oito) para a opção “não possui tomate”; 1 (uma) instância foi classificada incorretamente, correspondendo à 1,04% da aplicação; e, 13 (treze) instâncias não foram classificadas, totalizando 13,54%. Entre as regras geradas, destacou-se a não harmonização com outro ingrediente e a regra que envolve as características do vinho.

- **SE Pizza** Possui Ingrediente = Cogumelo Paris **ENTÃO** Pizza Não Possui Ingrediente = Tomate
- **SE Tempo de Amadurecimento** do Vinho = 13 (Treze) Meses **ENTÃO** Pizza **Não** Possui Ingrediente = Tomate

É importante ressaltar a aplicação do algoritmo *Prism*, resultou em um grande volume de regras geradas a partir da denominação das pizzas, porém este atributo não foi desconsiderado da base de dados, pois sua exclusão reduzia o índice de classificação corretas das instâncias, e conseqüentemente aumentava o índice das classificações incorretas.

4.2.3 Validação da Proposta

A fim de verificar se as regras geradas na aplicação dos algoritmos de mineração de dados acontecem em uma harmonização entre pizzas e vinhos, aplicou-se um questionário aos profissionais da Gastronomia e Enologia. Para este questionário, foram selecionadas 20 (vinte) regras de acordo com os maiores valores de grau de interesse das regras do *Apriori* e o índice (%) de taxa de acerto para as regras do *Prism*. A aplicação do questionário aconteceu no período de 27 de outubro

à 10 de novembro de 2018. Ao total foram coletadas 10 (dez) respostas, sendo que 3 (três) dos respondentes são profissionais da Gastronomia e outros 7 (sete) são profissionais da Enologia. Dentre os respondentes, estão profissionais que atuam em estabelecimentos como adegas e restaurantes; e os profissionais que atuam na área acadêmica.

A partir das situações apresentadas no questionário, 3 (três) obtiveram 100% das respostas voltadas para a alternativa “esta opção é comum”, cuja descrição está atrelada sobre o conhecimento da situação e percebe que ela se trata de uma situação habitual no ato de uma harmonização. São apresentadas a seguir, as regras que obtiveram unanimidade nos resultados.

- Se o vinho possui tanino **muito adstringente** e teor alcóolico igual à **14% (entre médio e alto)**, então, este vinho é considerado um **vinho tinto encorpado**.
- Se a pizza contém **atum**, então, ela não contém **embutidos**.
- Se o vinho possui acidez **baixa**, então, a pizza não possui **outros queijos**.

Quanto as demais regras avaliadas, houve 4 (quatro) regras onde o atributo consequente são os embutidos. A combinação que leva os atributos outros queijos, manjerição e tomate, 60,00% dos respondentes a classificaram como comum e 20,00% como uma regra surpreendente, podendo verificar que existe uma relação entre os ingredientes avaliados. Já regra que envolve os tipos de uva (Malbec) e barrica de amadurecimento (Carvalho), 80,00% dos respondentes classificam esta situação como comum e os demais 30,00% definem estas categorias divididas, destinando 10,00% para cada uma das opções existentes. A regra que envolve os embutidos como atributo consequente, abrange o estilo do vinho (Encorpado) e sua acidez (Média), é uma situação onde 70,00% à avaliaram comum e os demais 30,00% avaliam a regra como surpreendente. Já a situação em que realiza a harmonização entre embutidos com vinhos que possuem amadurecimento de 13 (treze) meses, resulta em 80,00% dos respondentes classificam esta situação como comum e os outros 20,00% se dividem entre surpreendente e que não se aplica.

As regras que possuem como o consequente o atributo denominado muçarela, envolvem a acidez e os aromas possíveis do vinho. A regra que envolve a combinação da muçarela e vinhos com aroma de ameixa e baunilha é avaliada em 60,00% com a

opção onde os profissionais possuem conhecimento da situação, mas não concordam com os resultados; porém, 30,00% dos respondentes a classificam como uma combinação comum e 10,00% como surpreendente. Enquanto a regra classifica a relação em que a acidez baixa do vinho harmoniza com muçarela, 70% dos respondentes a clássica como uma situação comum e 20% dos respondentes não sabem opinar nesta regra.

Dentre as regras geradas, onde o atributo consequente é o molho do tomate associados à alcachofra, estilo da uva (touriga franca), pimentão, vinho branco encorpado, os respondentes ficam divididos na questão de avaliação, tendo o maior índice de respostas de classificação da situação como comum.

Já o atributo outros queijos como consequente, envolvem o ingrediente aliche e a faixa de preço de comercialização do vinho. Entre a regra que associa o ingrediente aliche, os respondentes ficam divididos entre as opções de resposta, 20,00% não souberam opinar, 30,00% não concordam com a regra gerada, 30,00% associam a regra como uma harmonização comum e 20,00% acredita que esta regra é surpreendente.

Considerando o atributo tomate como consequente, obteve-se duas regras, uma associada ao ingrediente a não combinação com cogumelo paris e outra ao tempo de amadurecimento do vinho. Com relação a regra que avalia que o cogumelo paris não combina com tomate, cerca de 60,00% dos respondentes não concordam com esta situação e 40,00% acreditam que esta combinação é comum. Houve uma situação em que 90,00% dos respondentes não concordam com a regra gerada e 10,00% não possui conhecimento sobre a ocorrência da situação, isto significa que não existem restrições entre uma harmonização entre estes elementos. A regra gerada é:

- Se o vinho passa por **13 meses** de amadurecimento, então, ele não harmoniza com pizzas que contém **tomate**.

O quadro completo dos resultados obtidos na aplicação do questionário, estão alocados no Apêndice F.

4.3 SÍNTESE DOS RESULTADOS

A partir da análise estatística e aplicação do teste Qui-quadrado (χ^2) conforme apresentado na seção 4.1.1, foi possível verificar que dos atributos avaliados, apenas o grupo de embutidos apresentou diferenças estatisticamente significativas. Já a execução do teste de ACM, gerou 4 (quatro) grupos distintos de harmonização. Analisando os resultados obtidos na análise descritiva e na aplicação dos algoritmos de mineração de dados, é possível verificar que um grupo do gráfico de nuvem de pontos do teste de ACM, cujo estilo do vinho é encorpado se complementa com as regras geradas no algoritmo *Prism*, como é demonstrado a seguir. Não foi possível fazer uma relação entre o teste ACM e o algoritmo *Apriori*, pois a aplicação do algoritmo não é binária.

O teste Qui-quadrado (χ^2), para os conjuntos de queijos gerou as seguintes regras:

- **SE** o vinho tinto é encorpado, **ENTÃO** ele não harmoniza com o ingrediente muçarela
- **SE** o **vinho** tinto é encorpado, **ENTÃO** ele harmoniza com o ingrediente outros queijos

Enquanto a aplicação do algoritmo *Prism*, cujo atributo é o conjunto de queijos muçarela, retratou a seguinte regra:

- **SE** a pizza contém queijo muçarela, **ENTÃO** a acidez do vinho é baixa

Já a execução do algoritmo *Prism* com atributo-meta denominado outros queijos, identificou a seguinte regra:

- **SE** a **pizza** contém outros queijos, **ENTÃO** a acidez do vinho é alta

Considerando que os vinhos cujo estilo é encorpado, dispõe de propriedades densas e concentradas, quanto mais ácido o vinho mais encorpado ele se torna e apresenta elevado nível de adstringência no paladar. Para que haja harmonia entre os elementos do vinho é necessário o máximo equilíbrio entre estas propriedades,

pois se o vinho é encorpado, mas possui pouca acidez, ele se torna difícil de beber; por outro lado, se o rótulo contém uma acidez “muito alta” perante as demais características, a degustação se torna agressiva e desagradável (MIWA, 2016). Em uma harmonização de queijos é necessário agrupa-los por afinidade, os queijos suaves ou maturados como a muçarela, combinam com vinhos brancos e tintos de corpo leve ou médio (MARINI,2017). Enquanto, os queijos que possuem sabores mais fortes harmonizam com vinhos encorpados, considerando que os queijos fortes possuem uma acidez maior devido ao tempo de preparo, se equilibrando com acidez e tanino do vinho (CORDEIRO,2016).

Outra relação existente entre a aplicação da ACM e do *Prism* está relacionada a combinação do vinho tinto encorpado com o atributo denominado embutidos. A regra gerada pela aplicação foi:

- **SE** o **vinho** é tinto encorpado, **ENTÃO** ele harmoniza com o ingrediente embutidos

Já a regra gerada na execução do algoritmo *Prism* com atributo-meta o conjunto de embutidos, foi a seguinte:

- **SE** a pizza contém embutidos, **ENTÃO** o amadurecimento do vinho é de 13 (treze) meses

Existem vinhos, como os que possuem tanino (substância química presente na casca e semente da uva) que depois da produção necessitam amadurecer, com o propósito de suavizar o tanino e liberar o aroma para que o vinho esteja pronto para ser consumido. Este processo pode ser realizado em até 3 (três) anos. à primeira vista, os embutidos e curados combinam com vinhos tintos potentes, podendo ser aprimorada a partir dos diferentes tipos de gordura, sal e especiarias utilizadas na produção dos embutidos (REDAÇÃO PALADAR, 2018). Para realizar a harmonização com este produto, o vinho deverá possuir um bom grau de acidez, por exemplo, os salames combinam com vinhos frutados e com odor de violetas; enquanto para a harmonização com presuntos existe vasta variedade de combinações entre vinhos tintos e espumantes, desde que o vinho possua uma acidez elevada (MARINI, 2017).

Considerando que a aplicação dos algoritmos de mineração de dados teve como propósito extrair algum conhecimento desta base de dados, foi possível observar que o *Apriori* retorna algumas regras que consideram dois ou mais atributos no conseqüente enquanto que o *Prism* apresenta apenas o atributo meta (no conseqüente). Além disso, observa-se que a aplicação do *Apriori* rendeu relações melhores entre os ingredientes das pizzas e as características do vinho, apresentando resultados notáveis quanto ao grau de interesse das regras, por exemplo, a regra com menor *lift* foi a regra: “Se o teor alcóolico do vinho for 14% e o amadurecimento do rótulo for de 12 (doze) meses, então a pizza contém o ingrediente muçarela” com confiança de 91% e *lift* igual à 1,45 ; enquanto a regra com maior *lift* foi a seguinte: “ Se a pizza contém o ingrediente embutidos e a região de produção do vinho é Mendoza, então o vinho possui uva Malbec” com confiança de 91% *lift* igual a 4,85. Estes resultados são capazes de proporcionar a geração de “resultados surpreendentes” para a mineração de dados.

5 CONSIDERAÇÕES FINAIS

Inicialmente, pretendia-se utilizar uma base de dados Enogastronômica disponível em repositórios ou outros estudos anteriormente realizados que englobasse qualquer tipo de gastronomia, porém não houve sucesso nas buscas. Posteriormente, foram exploradas as vinícolas de Curitiba e Região Metropolitana, a fim de verificar se alguma delas possuía alguma base de dados ou algum documento que descrevesse a sugestão de harmonização em que estes locais sugerem aos seus clientes, sendo também descartada esta possibilidade após a verificação de que nenhuma vinícola dispõe deste conhecimento digitalmente organizado, além de não possuírem especialidade em harmonizações, mas sim nas especificações dos vinhos ali produzidos.

Após diversas pesquisas, foram definidas as pizzarias como objeto de estudo, pois se trata de um modelo de estabelecimento voltado à gastronomia Italiana e algumas delas, dispõem de adega própria, uma vez que a combinação destes dois elementos (pizzas e vinhos) é tradicional na cozinha italiana.

Destaca o fato de que, mais uma vez, nenhuma das pizzarias selecionadas dispõe de banco de dados de harmonizações. Desta forma, houve a necessidade de concepção de uma base de dados Enogastronômica que contemplasse esta combinação.

5.1 ALCANCE DOS OBJETIVOS

Para atingir o objetivo geral deste estudo: realizar uma análise de dados a partir da aplicação de técnicas de estatística e mineração de dados, a fim de identificar relações e padrões de ocorrência em uma base de dados Enogastronômica, foi necessário alcançar 3 (três) objetivos específicos.

O primeiro objetivo específico consiste em construir uma base de dados que reúna elementos da Gastronomia e Enologia. Este objetivo foi alcançado a partir de um levantamento bibliográfico e consulta com especialistas voltados aos elementos importantes para a realização de uma harmonização, sendo definidos para a coleta dos dados 16 (dezesesseis) atributos voltados às propriedades do vinho e 5 (cinco) voltados aos componentes das pizzas. A base de dados após o processo desmembramento de elementos, como os aromas possíveis, contou com 66 (sessenta

e seis) atributos, sendo 51 (cinquenta e um) voltados aos vinhos e 15 (quinze) voltados aos ingredientes das pizzas. Compôs esta base de dados, 32 (trinta e dois) sabores de pizzas sendo recomendado 3 (três) rótulos de vinhos para sabor, totalizando 96 (noventa e seis) instâncias e o uso de 86 (oitenta e seis) rótulos de vinhos distintos.

O segundo objetivo está relacionado em comparar os resultados obtidos a partir de dois métodos de mineração de dados, utilizou-se o *software* WEKA, aplicado a algoritmos voltados a descoberta de padrões, como o *Apriori* que apresenta regras por associação e o *Prism* que expressa regras geradas a partir da classificação. Os parâmetros utilizados em todas as aplicações estão comumente abordados nas literaturas utilizadas. A aplicação do algoritmo *Apriori* rendeu 88 (oitenta e oito) regras, cujo grau de interesse foi superior a 1,45 demonstrando uma correlação positiva entre os atributos, além de considerar dois ou mais atributos no consequente, sendo capaz de aprofundar a regra; e, o algoritmo *Prism* classificou os atributos que contém proporção equilibrada de ocorrência com 85,42% de taxa de acerto de classificação das regras, trazendo regras que envolvem apenas o atributo meta no consequente.

O terceiro objetivo está atrelado em realizar uma análise descritiva da base de dados e analisar os resultados obtidos. Utilizou-se o *software* R (com o RStudio) para realizar a análise, aplicando os testes Qui-quadrado (χ^2) para realizar uma análise entre as variáveis com maior ocorrência na base de dados e a Análise de Correspondência Múltipla (ACM) para expressar graficamente as relações entre os objetos e variáveis. Aplicou-se o teste Qui-quadrado (χ^2) nos atributos dos ingredientes das pizzas (embutidos, molho de tomate, muçarela e outros queijos) e dos dois estilos de vinho com características distintas (vinho branco leve e vinho tinto encorpado) com maior ocorrência na base de dados, verificando que apenas a aplicação com o atributo embutidos apresentou diferenças estatisticamente significativas entre as proporções. Já a aplicação do teste de Análise de Correspondência Múltipla (ACM), demonstrou-se que é possível explicar 28,33% do método com apenas duas dimensões, facilitando a demonstração da aplicação; e, o gráfico de nuvem de pontos que apresentou a existência de 4 (quatro) grupos distintos de harmonização.

5.2 CONTRIBUIÇÕES

Ao buscar por pesquisas no campo da Gestão da Informação, foi possível verificar que não existem estudos relacionados à Enogastronomia na área. Este estudo foi capaz de afirmar que é possível aplicar a mineração de dados e a estatística neste ramo do conhecimento.

Elaborou-se uma base de dados voltada a harmonização de pizzas e vinhos que pode ser utilizada como modelo para demais projetos voltados à Enogastronomia.

Outra forma de contribuição é a de disponibilizar a base de dados individual de cada um dos estabelecimentos participantes, de forma a auxiliar as demais pessoas que não possuem o conhecimento no ato de uma harmonização, quando o *Sommelier* estiver ausente, pois atualmente o conhecimento de harmonização está centralizado apenas neste profissional.

5.3 TRABALHOS FUTUROS

Para os trabalhos futuros, sugere-se a ampliação da base de dados buscando abranger um maior número de pizzarias e outros estabelecimentos tais como *trattorias*, por exemplo, e realizar uma pesquisa quantitativa dos dados.

Recomenda-se ainda, estender a ideia da criação de bases Enogastrônômica empregando a outros tipos de gastronomias, tais como, massas em geral, frutos do mar, aperitivos e sobremesas.

Como forma de contribuição às pizzarias participantes, os resultados alcançados serão repassados ao profissional que auxiliou no processo de construção da base de dados, visando ampliar o conhecimento com toda a equipe das pizzarias. Com as informações colhidas e os resultados obtidos, sugere-se a criação de um guia de harmonização de acordo com os sabores de pizza e rótulos de vinhos disponíveis na casa, a fim simplificar a escolha do cliente, seja pessoalmente ou no atendimento via *delivery*.

REFERÊNCIAS

ABBINAMENTI CIBO – VINO: Método Mercadini. **Spazioprever Lezioni**: a. Prever Pinerolo. p.1-19, [2012?]. Tradução nossa. Disponível em:

<<https://www.spazioprever.it/salabar/abbinamenti/mercadini.php>>. Acesso em: 18 maio 2018.

ARANDA, Yadura Robles; SOTOLONGO, Anthony R. Integración de los algoritmos de minería de datos 1R, PRISM, ID3 e a PostgreSQL. **Journal of Information Systems and Technology Management**, v. 10, n.2, p. 389-406, 2013. Disponível em: <<http://www.jistem.fea.usp.br/index.php/jistem/article/view/10.4301%252FS1807-17752013000200012>>. Acesso em: 04 jun. 2018.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação da informação: conceitos e tecnologia das máquinas de busca**. 2. ed. Porto Alegre: Bookman, 2013. 590 p.

BAILEIRO, Julio César. **Introdução à Estatística**. São Paulo, 2008. 71 slides, color, 10 X 5,625 cm. Disponível em: <http://www.usp.br/gmab/discip/zab5711/aula1_slides.pdf>. Acesso em: 30 nov. 2018.

BEATO, Manuel. **Guia de Vinhos Larousse**. São Paulo: Larousse do Brasil, 2006. 223 p.

BIGONHA, Roberto. **Princípios da Harmonização de Vinhos e Comidas**. Belo Horizonte., 2016. 44 p. Disponível em: <<http://homepages.dcc.ufmg.br/~bigonha/Livros/livro-harmonizacao-v1.pdf>>. Acesso em: 16 maio 2018.

BOENTE, Alfredo Nazareno P.; GOLDSCHMIDT, Ronaldo R.; ESTRELA, Vânia Vieira. Uma metodologia de suporte ao processo de descoberta do conhecimento em bases de dados. **Seget - Simpósio de Excelência em Gestão e Tecnologia**, Rio de Janeiro, p.1-14, 2008.

BORGES, Armando Coelho. O vinho na Bíblia. **Gula**. São Paulo:Peixes.v.13, n. 159, p.72-77, jan. 2006.

BORGES, José Luiz. **A arquitetura da refeição**. São Paulo: Winestile, 2016. 6 slides, color, 19,5X24,5. Disponível em: <<http://www.artwine.com.br/edicoes/wine-style-11-arquitetura-da-refeicao.pdf>>. Acesso em: 17 maio 2018.

BRANNER, Max. **Principles of Data Mining**. Londres: Springer-verlag London Limited, 2007. 342 p.

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à Mineração de Dados: Conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016. 559 p.

CÉSAR, Nelson Barroso. **Curso de Tecnologia em Gastronomia**. São Paulo: Universidade Anhembi Morumbi, 2008. Ebook. Disponível em: <<https://slidex.tips/download/curso-de-tecnologia-em-gastronomia>>. Acesso em: 09 mai. 2018.

CHAPMAN, Pete; CLINTON, Julian; KERBER, Randy; et al. Crisp-Dm 1.0. **CRISP-DM Consortium**, p. 76, 2000.

CHOO, Chun Wei. **A organização do conhecimento**: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. Tradução Eliana Rocha. São Paulo: Editora Senac São Paulo, 2003.

CONFRARIA DOS SOMMELIERS. São Paulo, [2008?]. Ebook. Disponível em: <<https://sommelier.weebly.com/moacutedulos--textos.html>>. Acesso em: 08 mai. 2018.

COPELLO, Marcelo. **Carvalho e sua influência sobre o vinho**. 2016. Disponível em: <https://revistaadega.uol.com.br/artigo/carvalho-o-tempero-do-vinho_8022.html>. Acesso em: 23 out. 2018.

CORDEIRO, Carlos. **Vinhos e queijos, a combinação perfeita**: Um guia para harmonizar um dos mais clássicos encontros na mesa. Colocamos em prova mais de 20 tipos de queijos. 2016. Disponível em: <https://revistaadega.uol.com.br/artigo/queijos-e-vinhos-pares-perfeitos_3311.html>. Acesso em: 09 nov. 2018.

CORREA, Ana Paula Araujo; QUEIROZ, Eder; TREVISAN, Newton. **TESTE DO QUI-QUADRADO**. Curitiba, 2015. 21 slides, P&B, 10 X 5,625 cm. Disponível em: <http://www.leg.ufpr.br/lib/exe/fetch.php/disciplinas:ce001:teste_do_qui-quadrado.pdf>. Acesso em: 05 nov. 2018.

CURRY, Andrew. Nosso romance de 9.000 anos com bebidas. **National Geographic**, Washington. fev. 2017.

DARDEAU, Rogério. **Vinhos, uma festa dos sentidos**. Rio de Janeiro: Mauad, 3^o Edição revista e ampliada, 2004. 168 p.

DAVENPORT, T. H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998. 312 p.

DE ALBUQUERQUE BARRETO, Aldo. Os destinos da ciência da informação: entre o cristal e a chama. **Informação & Sociedade**, v. 9, n. 2, 1999.

DICIONÁRIO MICHAELIS. Disponível em: <<http://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/enogastronomia/>>. Acesso em: 10 abr. 2018.

DONADINI, Gianluca et al. Evaluation of ideal Everyday Italian Food and Beer Pairings with Regular Consumers and Food and Beverage Experts. **Journal Of The Institute Of Brewing**. Piacenza, p. 329-342. 2008. Disponível em:

<<https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2050-0416.2008.tb00777.x>>.
Acesso em: 17 maio 2018.

DUARTE, Renato Luis da Rocha. **O sommelier de vinhos no mercado Enogastronômico de São Paulo: hospitalidade e prática**.2016. 105 f. Dissertação (Mestrado) - Curso de Hospitalidade, Planejamento e Gestão Estratégica em

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**,v. 3, n. 17, p.1-18,1996.

FERNEDA, Edberto. Recuperação de Informação. **Análise sobre a contribuição da**, 2003.

FERRÃO, Tháys. APRECIÇÃO DO VINHO: Arte, Cultura e Conhecimento. **Instituto Cultural do Vinho**, Curitiba, p.1-22, 2011. Disponível em: <http://www.mppr.mp.br/arquivos/File/imprensa/2014/DescomplicandoVinho_DiaDoServidor.pdf>. Acesso em: 19 maio 2018.

FIGUEREIDO, Dayana Ester Andrade. **Recuperação da informação: uma análise sobre os sistemas de busca da web**.2006. 64 f. TCC (Graduação) - Curso de Biblioteconomia, Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2006.

FREITAS, Alex A. On objective measures of rule surprisingness. In: **European Symposium on Principles of Data Mining and Knowledge Discovery**. Springer, Berlin, Heidelberg, 1998. p. 1-9.

GAZETA DO POVO (Paraná) (Org.). **Prêmio Bom Gourmet: Sabor com identidade**. Curitiba: Gazeta do Povo, 2018. 262 p.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GRANZOTTO, Alexandre José. **Estatística Básica**. 2002. 33 p. Disponível em: <<http://www.etepiracicaba.org.br/cursos/exercicios/em/ResumaoEstatisticaBasica.pdf>>. Acesso em: 30 nov. 2018.

GUIMARÃES, Paulo Ricardo Bittencourt. **Métodos Quantitativos Estatísticos**. Curitiba: IESDE Brasil S.A., 2007. 64 p. Disponível em: <<http://www.inf.ufsc.br/~vera.carmo/LIVROS/LIVROS/Metodos%20Quantitativos%20%20Estatisticos%20Paulo%20Ricardo%20BittencourtGuimar%e3es.pdf>>. Acesso em: 30 nov. 2018.

HALILI, Festim; RUSTEMI, Avni. Predictive Modeling: Data Mining Regression Technique Applied in a Prototype. **International Journal Of Computer Science And Mobile Computing**. Tetovo, p. 207-215. ago. 2016. Disponível em: <<https://www.ijcsmc.com/docs/papers/August2016/V5I8201642.pdf>>. Acesso em: 04 jun. 2018.

HARMONIZAÇÕES: aprendendo os conceitos. Caxias do Sul: VinumDay, [2016?]. Disponível em: <<http://blog.vinumday.com.br/baixar-e-book/Vinumday-Harmonizaçã>> Acesso em: 07 mai. 2018.

HOSPITALIDADE, Universidade Anhembi Morumbi, São Paulo, 2016. Disponível em: <http://portal.anhembi.br/wp-content/uploads/2017/06/Dissertacao_RENATO-LUIS-DA-ROCHA-DUARTE.pdf>. Acesso em: 29 maio 2018.

Huot, Réjean (2002). Métodos quantitativos para as ciências humanas (tradução de Maria Luísa Figueiredo). Lisboa: Instituto Piaget.

ISTITUTO EUROPEO DI DESIGN (São Paulo). **ESCOLA DE SOMMELIERS**. 2011. Disponível em: <https://ied.edu.br/biblioteca/sao-paulo/01_01_6558.pdf>. Acesso em: 14 nov. 2018.

LAROUSSE (São Paulo). **Laousse do Vinho**: 2. ed. São Paulo: Larousse do Brasil, 2007. 385 p. Tradução de Larousse des Vins.

MARCHIORI, Patrícia Zeni. A ciência e a gestão da informação: compatibilidades no espaço profissional. **Ciência da informação**, v. 31, n. 2, 2002,

MARINI, Jhonatan. **Quais são os vinhos que mais bem acompanham frios e embutidos?** 2017. Disponível em: <<http://blog.famigliavalduga.com.br/quais-sao-os-inhos-que-mais-bem-acompanham-frios-e-embutidos/>>. Acesso em: 09 nov. 2018.

MARINI, Jhonatan. **Queijos e vinhos: dicas deliciosas para não errar na harmonização**. 2017. Disponível em: <<http://blog.famigliavalduga.com.br/queijos-e-inhos-dicas-deliciosas-para-nao-errar-na-harmonizacao/>>. Acesso em: 09 nov. 2018

MCGEE, James; PRUSAK, Laurence. Gerenciamento estratégico da informação: aumente a competitividade e a eficiência de sua empresa utilizando a informação como uma ferramenta estratégica. **Rio de Janeiro: Campus**, 1994.

MILAZZO, Giuseppe. TECNICA DI ABBINAMENTO CIBO VINI. **Dispensa di Abbinamento Cibo-vino**, p.1-24, dez. 2015. Disponível em: <<http://www.giuseppemilazzo.it/wp-content/uploads/2015/12/ABVIN.pdf>>. Acesso em: 16 maio 2018.

MIRANDA, Fernando. **Eneventos: A arte de harmonizar**. 2009. Modelo do Método Mercadini. Disponível em: <<http://www.eneventos.com.br/colunistas/fernando/fernando003.htm>>. Acesso em: 24 maio 2018.

MIRANDA, Silvânia Vieira de; STREIT, Rosalvo Ermes. O processo de gestão da informação em organizações públicas. **ENCONTRO DE ADMINISTRAÇÃO DA INFORMAÇÃO**, v. 1, p. 1-17, 2007.

MIWA, Marcel. **Quem tem medo da acidez?** 2016. Disponível em: <https://revistaadega.uol.com.br/artigo/quem-tem-medo-da-acidez_691.html>. Acesso em: 09 nov. 2018.

MONTECINOS, Antonio et al. **Turismo em pauta: Quem pensa e faz o turismo acontecer.** Rio de Janeiro: Senac, 2012. 52 p. Disponível em: <http://cnc.org.br/sites/default/files/arquivos/turismo_pauta_10_web.pdf>. Acesso em: 20 maio 2018.

MOREIRA, Cleverson Bayer. **Gestão da Informação.** Guarapuava: Unicentro, 2015. 42 p.

NETTO, Augusto Pimazoni. **Como Interpretar a Significância Estatística.** 2007. Disponível em: <https://www.diabetes.org.br/publico/colunistas/20-dr-augusto-pimazoni-netto/125-como-interpretar-a-significancia-estatistica?fbclid=IwAR0Uxuth3G-IWQpnopo1RgDOGvxAm3InXbrt4BwnZ_9vOKsNaJghQrqspeE>. Acesso em: 08 nov. 2018.

PEREIRA, Sandra. Enologia e Harmonização. **Secretaria de Educação e Esportes,** Pernambuco, p.1-91, 2014.

PINTO, Daniel. **Manual Didático do Vinho – Iniciação à enologia.** São Paulo: Anhembi Morumbi, 2009.

PONJUÁN DANTE, Gloria. **Gestión de información en las organizaciones: principios, conceptos y aplicaciones.** Santiago, CL: Universidad de Chile, Centro de Información en Capacitación, 1998.

PORTAL DE PERIÓDICOS CAPES/MEC. Disponível em: <<http://www.periodicos.capes.gov.br/>>. Acesso em: 09 abr. 2018.

PUCKETTE, Madeline; HAMMACK, Justin. **O guia essencial do vinho: Wine Folly.** Rio de Janeiro: Intrínseca, 2016. 239 p. Tradução de Lucas Cordeiro e Renato Ferreira Pires.

QUEIROZ, Selma Jamar de. **Mitos e Controvérsias no Cosmo da Harmonização de Vinhos.** 2009. 67 f. Monografia (Especialização) - Curso de Magistério Superior, Pró - Reitoria de Pesquisa, Pós-graduação, Extensão e Cultura - Proppec, Universidade do Vale do Itajaí - Univali, Balneário Camboriú, 2009. Disponível em: <[http://siaibib01.univali.br/pdf/Selma Jamar de Queiroz.pdf](http://siaibib01.univali.br/pdf/Selma%20Jamar%20de%20Queiroz.pdf)>. Acesso em: 21 maio 2018.

REDAÇÃO PALADAR (Ed.). **Guia de Harmonização de Queijos e Embutidos com Vinhos e Cervejas.** 2018. Disponível em: <<https://paladar.estadao.com.br/noticias/comida,guia-de-harmonizacao-de-queijos-e-embutidos-com-vinhos-e-cervejas,70002480844>>. Acesso em: 09 nov. 2018.

ROCHA, Adriano Alves. **Elaboração de um curso de iniciação ao conhecimento do vinho e à degustação para o Serviço Nacional de Aprendizagem**

Comercial. 2006. 51 f. TCC (Pós-Graduação) - Curso de Especialização em Gastronomia e Segurança Alimentar, Centro de Excelência em Turismo, Universidade de Brasília, Brasília, 2006.

ROMÃO, Wesley et. al. Extração de regras de associação em C&T: O algoritmo Apriori. **XIX Encontro Nacional em Engenharia da Produção**, v.34, p.01-20, 1999.

ROMÃO, Wesley. **Descoberta de conhecimento relevante em banco de dados sobre ciência e tecnologia.** 2002. 253 f. Tese (Doutorado) - Curso de Engenharia de Produção, Universidade Federal de Santa Catarina, Santa Catarina, 2002.

SALLES, Daniel et al. **Onde comer boas pizzas em Curitiba.** 2017/2018. Disponível em: <<https://veja.abril.com.br/entretenimento/onde-comer-boas-pizzas-em-curitiba/>>. Acesso em: 15 ago. 2018.

SANTOS, Alcione Miranda dos. **Introdução à Estatística.** Maranhão: N/a, 2009. 105 slides, color, 10 X 5,625 cm. Disponível em: <<https://docente.ifsc.edu.br/gianpaulo.medeiros/MaterialDidatico/M%C3%A9todos%20Est%C3%A1tisticos/estatistica%20aula%201.pdf>>. Acesso em: 30 nov. 2018.

SCIENTIFIC ELETRONIC LIBRARY ONLINE (SciELO). Disponível em: <<http://www.scielo.org/php/index.php>>. Acesso em: 09 abr. 2018.

SCOPUS. Disponível em: <<https://www-scopus.ez22.periodicos.capes.gov.br/>>. Acesso em: 09 abr. 2018.

SETZER, Valdemar W. Dado, informação, conhecimento e competência. **Datagramazero: revista de ciência da informação**, São Paulo, dez. 1999.

SHIGUTI, Wanderley Akira; SHIGUTI, Valéria da Silva Cruz. **Apostila de Estatística.** Brasília: N/a, 2006. 70 p. Disponível em: <http://www.inf.ufsc.br/~paulo.s.borges/Download/Apostila5_INE5102_Quimica.pdf>. Acesso em: 30 nov. 2018.

SILVA, Roberto do Nascimento e. **A Enogastronomia no processo de formação de identidade territorial do Vale dos Vinhedos/RS.** 2015. 178 f. Dissertação (Mestrado) - Curso de Desenvolvimento Regional, Programa de Pós-graduação em Desenvolvimento Regional - Mestrado e Doutorado, Universidade de Santa Cruz do Sul - UNISC, Santa Cruz do Sul - Rs, 2015.

SISTEMA DE BIBLIOTECAS - UFPR. Disponível em: <<http://www.portal.ufpr.br/>>. Acesso em: 09 abr. 2018.

VACCARINI, Giuseppe. **Manuale degli abbinamenti.** Giunti editore, 2010.

VALDUGA, Vander. **ENOGASTRONOMIA.** Curitiba: Universidade Federal do Paraná., 2018. 25 slides, color, 10 X 5,625 cm.

VALENTIM, Marta Lúcia Pomim. Informação e conhecimento em organizações complexas. **Gestão da informação e do conhecimento. São Paulo: Polis**, p. 11-

25, 2008.

VASCONCELOS, Livia Maria Rocha de; DE CARVALHO, Cedric Luiz. Aplicação de regras de associação para mineração de dados na *web*. **Instituto de Informática da Universidade Federal de Goiás**, 2004.

VIGANÒ, Giovanni. **Enogastronomy: a path towards the discovery of the local culture**. Milão, p.1-4, 2015.

TARAPANOFF, Kira. **Inteligência, informação e conhecimento em corporações**. Brasília: IBICIT e UNESCO, 2006. p.453.

TECH, Ellen Cristina Moronte. **Uma abordagem metodológica para quantificar os efeitos cognitivos na análise sensorial de alimentos**. 2013. 85 p. Dissertação (Mestrado em Engenharia de Alimentos) – Faculdade de Zootecnia e Engenharia de Alimentos – Universidade de São Paulo, Pirassununga, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/74/74132/tde-05042013-082843/en.php>>. Acesso em: 30 maio 2018.

WAIKATO, Machine Learning Group At The University Of (Org.). **Weka 3: Data Mining Software in Java**. 2006. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 01 jun. 2018.

WEB OF SCIENCE. Disponível em: <<http://appswebofknowledge.ez22.periodicos.capes.gov.br>>. Acesso em: 09 abr. 2018.

APÊNCICE A – QUESTIONÁRIO DE VALIDAÇÃO DA PESQUISA**QUESTIONÁRIO DE VALIDAÇÃO DE PROPOSTA****TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

Pesquisa: Questionário voltado a validação de regras geradas a partir da aplicação de algoritmos de mineração de dados aplicados à uma Base de Dados Enogastronômica de pizzas e vinhos.

Pesquisadora: Andréia Karolina Lapkouski - Graduanda em Gestão da Informação - UFPR.

Orientadora: Prof. Dra. Denise Fukumi Tsunoda.

1. **Natureza da pesquisa:** Você é convidado a participar desta pesquisa, que tem como objetivo validar as regras de classificação geradas a partir da aplicação de algoritmos de mineração de dados em uma Base de Dados Enogastronômica de pizzas e vinhos.
2. **Público Alvo da Pesquisa:** Profissionais Graduados em Gastronomia e/ou Enologia residentes em Curitiba ou Região Metropolitana de Curitiba.
3. **Envolvimento da Pesquisa:** Ao participar deste estudo você deve **permitir** que a pesquisadora Andréia Karolina Lapkouski utilize os dados contidos nas respostas deste questionário. A participação ao responder todas as perguntas leva aproximadamente 10 minutos. Você tem a liberdade de se recusar a participar desta pesquisa, sem qualquer prejuízo para você. No entanto, pedimos sua colaboração para completar a etapa de validação da proposta junto à especialistas, garantindo assim a verificação da ocorrência dos resultados listados no ato de uma harmonização. Em caso de dúvidas, faça contato com a pesquisadora Andréia através do *E-mail*: andreiaklap@gmail.com.
4. **Riscos e Desconforto:** Os procedimentos utilizados nesta pesquisa obedecem aos Critérios da Ética na Pesquisa com Seres Humanos **conforme** a Resolução No. 466/2012 do Conselho Nacional de Saúde. Nenhum dos procedimentos utilizados oferece riscos à sua dignidade. Sendo que estas livre para desistir da pesquisa em qualquer momento sem nenhum prejuízo.
5. **Benefícios:** Esperamos que este estudo traga informações importantes sobre o tema. No futuro essas informações poderão ser usadas em **benefícios** de outros.
6. **Confidencialidade:** Todas as informações coletadas neste estudo são estritamente confidenciais. Apenas os membros do grupo de pesquisa terão conhecimento dos dados e mesmo os membros do grupo não serão capazes de identificá-lo individualmente.
7. **Pagamento:** Você não terá nenhum tipo de despesa por participar desta pesquisa, bem como nada será pago por sua participação.

Aceita participar livremente desta pesquisa? () Sim () Não

INSTRUÇÕES

Os dados analisados foram retirados de uma aplicação que contém 32 sabores de pizza e 86 rótulos de vinho. Por se tratar de uma análise qualitativa de dados, foram realizados testes em busca de melhores resultados, onde definiu-se a necessidade de alocar determinados ingredientes em conjuntos de dados, sendo eles:

Embutidos: Calabresa, Calabresa Picante, Capocollo, Linguiça Toscana, Lombo Defumado, Panceta, Peito de Peru, Presunto, Presunto Defumado e Presunto de Parma.

Florais: Flor de Lima, Florais, Flores Brancas, Rosas e Violetas.

Molho de Tomate: Molho de Tomate, Molho de Tomate Fresco e Molho de Tomate Italiano.

Muçarela: Muçarela, Muçarela de Búfala, Muçarela de Burrata, Muçarela Especial e Muçarela Tradição.

Outros Queijos: Queijo de Cabra, Catupiry, Queijo Emmental, Queijo Gorgonzola e Queijo Parmesão.

Tomate: Tomate, Tomate Cereja, Tomate Fresco e Tomate Seco.

Este questionário possui 20 situações. Para a realização do preenchimento do mesmo, é necessário assinalar um "X" na alternativa em que você acredita que a situação melhor se enquadre. Para cada situação deve ser escolhida apenas **uma** alternativa. Dentre as alternativas, se encontram as seguintes opções:

Não se aplica: Você não possui nenhum conhecimento sobre esta situação.

Não concordo com esta situação: Você possui conhecimento sobre a situação, mas não concorda com o resultado.

Esta situação é comum: Você possui conhecimento sobre esta situação e percebe que se trata de uma combinação comum.

Esta situação é surpreendente: Você não possui conhecimento sobre a situação, mas acredita que a combinação pode ocorrer, gerando assim um novo conhecimento.

Sou um profissional de: () Gastronomia

() Enologia/Sommelier

SITUAÇÕES	Não se aplica	Não concordo com esta situação	Esta situação é comum	Esta situação é surpreendente
Se a pizza contém como outros queijos e tempero de manjeriço , então, estes ingredientes harmonizam com embutidos .				
Se o vinho contém uva malbec e é armazenado em barrica de carvalho , então este vinho combina com pizzas que contém embutidos .				
Se o vinho possui tanino muito adstringente e teor alcóolico igual à 14% (entre médio e alto) , então, este vinho é considerado um vinho tinto encorpado .				
Se a pizza contém muçarela e o vinho é encorpado com acidez alta , então, recomenda-se um vinho com faixa de preço C (Entre R\$100,01 à R\$170,00) que seja amadurecido em barrica de carvalho .				
Dentre os possíveis aromas do vinho é possível identificar os aromas de ameixa e baunilha , então, recomenda-se uma pizza que contenha muçarela .				
Se o vinho é tinto encorpado e possui acidez média , então, harmoniza com pizzas que contém embutidos .				
Se o vinho contém a uva touriga nacional , então, dentre os aromas possíveis do vinho se encontra os florais .				

SITUAÇÕES	Não se aplica	Não concordo com esta situação	Esta situação é comum	Esta situação é surpreendente
Se a pizza contém embutidos como ingrediente e a região de produção do vinho é Mendoza – Argentina , então, o vinho contém uva malbec .				
Se a pizza contém pimentão , então, ela também contém molho de tomate .				
Se o vinho é branco encorpado , então, a pizza contém molho de tomate .				
Se a pizza possui alcachofra , então, ela <u>não</u> possui molho de tomate .				
Se o vinho contém uva touriga franca , então, não recomenda-se uma pizza que contenha molho de tomate .				
Se o vinho possui acidez baixa , então, a pizza contém muçarela .				
Se o vinho passa por 13 meses de amadurecimento, então, ele harmoniza com pizzas que contém embutidos .				
Se a pizza contém atum , então, ela <u>não</u> contém embutidos .				
Se a pizza contém aliche , então, ela <u>não</u> possui outros queijos como ingrediente.				
Se o vinho possui acidez baixa , então, a pizza <u>não</u> possui outros queijos .				

SITUAÇÕES	Não se aplica	Não concordo com esta situação	Esta situação é comum	Esta situação é surpreendente
Se o vinho possui a faixa de preço E (Mais que R\$250,01) , então, este vinho harmoniza com pizzas que contém outros queijos .				
Se a pizza contém cogumelo paris , então, ela não possui tomate .				
Se o vinho passa por 13 meses de amadurecimento, então, ele <u>não</u> harmoniza com pizzas que contém tomate .				

Comentários (opcional):

APÊNDICE B – VARIEDADE DE VINHOS

(Continua)

OCORRÊNCIA	VINHO	ESTILO DO VINHO
1	Almaviva	Vinho Tinto Encorpado
1	Amalaya	Vinho Tinto De Médio Corpo
1	Amalia Dos Fincas	Vinho Branco Leve
1	Antano	Vinho Tinto De Médio Corpo
1	Ars In Vitro Tempranillo Organico	Vinho Tinto Leve
1	Ars Nova	Vinho Tinto De Médio Corpo
1	Barbera Montebello	Vinho Tinto Encorpado
1	BaroloTerlo	Vinho Tinto Encorpado
1	Barone Nebbiolo	Vinho Tinto De Médio Corpo
1	Benmarco	Vinho Tinto De Médio Corpo
1	Calathus Gran Cortel	Vinho Tinto Encorpado
1	Calvet Merlot	Vinho Tinto De Médio Corpo
1	Casanova Gran Reserva	Vinho Tinto Encorpado
1	Casarena 505 Esencia	Vinho Tinto De Médio Corpo
1	Casa Rivas	Vinho Tinto De Médio Corpo
1	Casa Rivas Cabernet Sauvignon	Vinho Tinto De Médio Corpo
1	Casa Silva Reserva Couvée Merlot	Vinho Tinto Encorpado
1	Chianti DOCCG Il Palazzo	Vinho Tinto De Médio Corpo
1	Chianti DOCG Castelo Di Bossi	Vinho Tinto Encorpado
1	Chianti Reserva Il Paladazzo DOCG	Vinho Tinto Encorpado
2	Chocalam Reserva	Vinho Tinto Encorpado
1	Crios da Susana Balbo	Vinho Rosé
1	Crios Malbec	Vinho Tinto De Médio Corpo
1	Croatina	Vinho Tinto Encorpado
1	Don Roman	Vinho Tinto Encorpado
1	Dos Fincas	Vinho Branco Leve
1	Esteva Douro	Vinho Tinto De Médio Corpo
1	Flor de Maio	Vinho Tinto De Médio Corpo
1	Gaggiarone	Vinho Tinto De Médio Corpo
1	Garda Merlot DOC	Vinho Tinto De Médio Corpo
1	Grey Ventisquero	Vinho Tinto Encorpado

(Conclusão)

1	Heartland	Vinho Tinto Encorpado
1	Humberto Canale	Vinho Tinto De Médio Corpo
1	Klein Kloof	Vinho Tinto De Médio Corpo
1	La Joven Gewuztraminer	VinhoBrancoDeMedioCorpo
1	La Joven Viognier	Vinho Branco Leve
1	Lan Crianza	Vinho Tinto Encorpado
1	Lan Crianza Rioja	Vinho Tinto Encorpado
1	Leyda Classic	Vinho Tinto Leve
1	Il Palazzo Chianti	Vinho Tinto De Médio Corpo
1	Luigi Bosca Malbec	Vinho Tinto Encorpado
3	Manz Contador de Estória	Vinho Tinto Encorpado
1	Manz Douro	Vinho Tinto Encorpado
1	Manz Penedo	Vinho Tinto Leve
1	Monte da Gloria	Vinho Tinto De Médio Corpo
1	Mural Douro DOC	Vinho Branco Encorpado
1	Mural Tinto	Vinho Tinto De Médio Corpo
2	Muros de Vinha	Vinho Branco Leve
1	Muros de Vinha	Vinho Rosé
1	Muros de Vinha	Vinho Tinto De Médio Corpo
1	Music D´Aria	Vinho Tinto Leve
1	Norton Malbec	Vinho Tinto Encorpado
1	Perdriel	Vinho Tinto De Médio Corpo
1	Pettit Bobo	Vinho Branco Encorpado
1	Pilandro Merlot 70 Anni	Vinho Tinto Encorpado
1	Pommery	Espumante
1	Pommery Brut Royal	Espumante
2	Porta 6	Vinho Tinto De Médio Corpo
1	Portal Colheita	Vinho Tinto Encorpado
1	Potty Lazzarotto	Espumante
1	PrimitivoDeSalento	Vinho Tinto Leve
1	Punto Maximo	Vinho Tinto De Médio Corpo
2	Ramirana Reserva	Vinho Tinto De Médio Corpo
3	Saint Clair	Vinho Branco Leve
2	Santa Ema Gran Reserva	Vinho Tinto Encorpado
1	Saurus Pinot	Vinho Tinto Leve
1	Sena	Vinho Tinto Encorpado
1	Splendore	Vinho Tinto Encorpado
1	Stagnari	Vinho Tinto Encorpado
1	Steenberg Sauvignon Blanc	Vinho Branco de Médio Corpo
1	Sutter Home	Vinho Tinto Leve
2	Trevo	Vinho Branco Leve
1	Trumpeter Malbec	Vinho Tinto Encorpado
1	Ventisquero Reserva	Vinho Tinto Leve
1	Villa Medoro Montepulciano DOC	Vinho Tinto De Médio Corpo
1	Villa Pani	Vinho Tinto De Médio Corpo
1	Villa Pani Toscana Rosso IGT	Vinho Tinto De Médio Corpo
1	Vina Amalia Gran Cabernet	Vinho Tinto Encorpado
1	Vina Amalia Gran Malbec	Vinho Tinto Encorpado
1	Vina Amalia Gran Reserva Cabernet Sauvignon	Vinho Tinto Encorpado
1	Vina Amalia Reserva	Vinho Tinto Encorpado
1	Vistalago Cabernet Sauvignon	Vinho Tinto De Médio Corpo
1	Vistalago	Vinho Tinto Encorpado
1	Von Siebenthal Parcela 7 Gran Reserva	Vinho Tinto Encorpado
1	Yellow Tail	Vinho Tinto Leve
1	Zorzau Terroir Único	Vinho Tinto De Médio Corpo

APÊNDICE C – PAÍSES E REGIÕES DE PRODUÇÃO DOS RÓTULOS

QUANTIDADE DE RÓTULOS POR PAÍS	PAÍS	QUANTIDADE DE RÓTULOS POR REGIÃO	REGIÃO
3	África do Sul	1	Constantia Valley
		1	Tygerbeg
		1	Western Cape
20	Argentina	17	Mendoza
		1	Patagônia
		1	San Patricio
		1	Valles Calchaquies
2	Austrália	1	Langhorne Creek
		1	South Eastern
4	Brasil	2	Altos Montes
		1	Rodeio
		1	São José dos Pinhais
19	Chile	1	Puente Alto
		1	Valle Central
		6	Valle Del Conchagua
		1	Valle Del Leyda
		8	Valle Del Maipo
		2	Valle Del Maule
6	Espanha	2	Navarra
		1	Ribera Del Duero
		3	Rioja
2	Estados Unidos	2	Lod
3	França	1	Bordeaux
		2	Champagne
14	Itália	1	Abruzzo
		1	Lago Di Guarda
		4	Lombardia
		1	Piemonte
		1	Puglia
		6	Toscana
3	Nova Zelândia	3	Mallborough
19	Portugal	1	Alentejano
		1	Dao
		9	Douro
		3	Lisboa
		3	Península de Setúbal
		2	Vinho Verde
1	Uruguai	1	La Caballada

APÊNDICE D – TEMPO DE AMADURECIMENTO (EM MESES) E TIPO DE BARRICA PARA MATURAÇÃO DO VINHO

AMADURECIMENTO	OCORRÊNCIA
0 meses	2
2 meses	4
4 meses	6
6 meses	5
7 meses	1
8 meses	2
9 meses	6
10 meses	1
11 meses	7
12 meses	15
13 meses	4
14 meses	3
20 meses	3
22 meses	1
24 meses	5
36 meses	3
Não Informado	28

BARRICA	OCORRÊNCIA
Não Identificado	18
Carvalho	49
Concreto	1
Inox	16
Inox e Carvalho	12

APÊNDICE E – AROMAS POSSÍVEIS NO ATO DA DEGUSTAÇÃO DO VINHO

AROMAS POSSÍVEIS	OCORRÊNCIA
Abacaxi	11
Alcacuz	5
Ameixa	39
Amora	33
Baunilha	24
Café	4
Caramelo	7
Cassis	8
Cereja	38
Chocolate	13
Couro	4
Especiarias	43
Florais	43
Framboesa	24
Frutado	4
Frutas Secas	5
Geléia	5
Groselha	11
Maracujá	5
Minerais	11
Mirtilo	13
Morango	23
Pêssego	10
Pimenta	12
Tabaco	6
Toranja	5

APÊNDICE F – RESULTADO DA APLICAÇÃO DO QUESTIONÁRIO DE VALIDAÇÃO DA PROPOSTA

SITUAÇÃO	AVALIAÇÃO			
	Não se aplica	Não concordo com esta situação	Esta situação é comum	Esta situação é surpreendente
Se a pizza contém como outros queijos e tempero de manjeriço , então, estes ingredientes harmonizam com embutidos .	1	1	6	2
Se o vinho contém uva malbec e é armazenado em barrica de carvalho , então este vinho combina com pizzas que contém embutidos .		1	8	1
Se o vinho possui tanino muito adstringente e teor alcóolico igual à 14% (entre médio e alto) , então, este vinho é considerado um vinho tinto encorpado .			10	
Se a pizza contém muçarela e o vinho é encorpado com acidez alta , então, recomenda-se um vinho com faixa de preço C (Entre R\$100,01 à R\$170,00) que seja amadurecido em barrica de carvalho .	1	5	4	
Dentre os possíveis aromas do vinho é possível identificar os aromas de ameixa e baunilha , então, recomenda-se uma pizza que contenha muçarela .		6	3	1
Se o vinho é tinto encorpado e possui acidez média , então, harmoniza com pizzas que contém embutidos .			7	3
Se o vinho contém a uva touriga nacional , então, dentre os aromas possíveis do vinho se encontra os florais .		1	9	
Se a pizza contém embutidos como ingrediente e a região de produção do vinho é Mendoza – Argentina , então, o vinho contém uva malbec .	3	2	5	
Se a pizza contém pimentão , então, ela também contém molho de tomate .		3	6	1
Se o vinho é branco encorpado , então, a pizza contém molho de tomate .	3	2	5	
Se a pizza possui alcachofra , então, ela não possui molho de tomate .		4	6	
Se o vinho contém uva touriga franca , então, não recomenda-se uma pizza que contenha molho de tomate .	3	3	3	1
Se o vinho possui acidez baixa , então, a pizza contém muçarela .	2	1	7	
Se o vinho passa por 13 meses de amadurecimento, então, ele harmoniza com pizzas que contém embutidos .	1		8	1
Se a pizza contém atum , então, ela não contém embutidos .			10	
Se a pizza contém aliche , então, ela não possui outros queijos como ingrediente.	2	3	3	2
Se o vinho possui acidez baixa , então, a pizza não possui outros queijos .			10	
Se o vinho possui a faixa de preço E (Mais que R\$250,01) , então, este vinho harmoniza com pizzas que contém outros queijos .		3	6	1
Se a pizza contém cogumelo paris , então, ela não possui tomate		6	4	
Se o vinho passa por 13 meses de amadurecimento, então, ele não harmoniza com pizzas que contém tomate .	1	9		