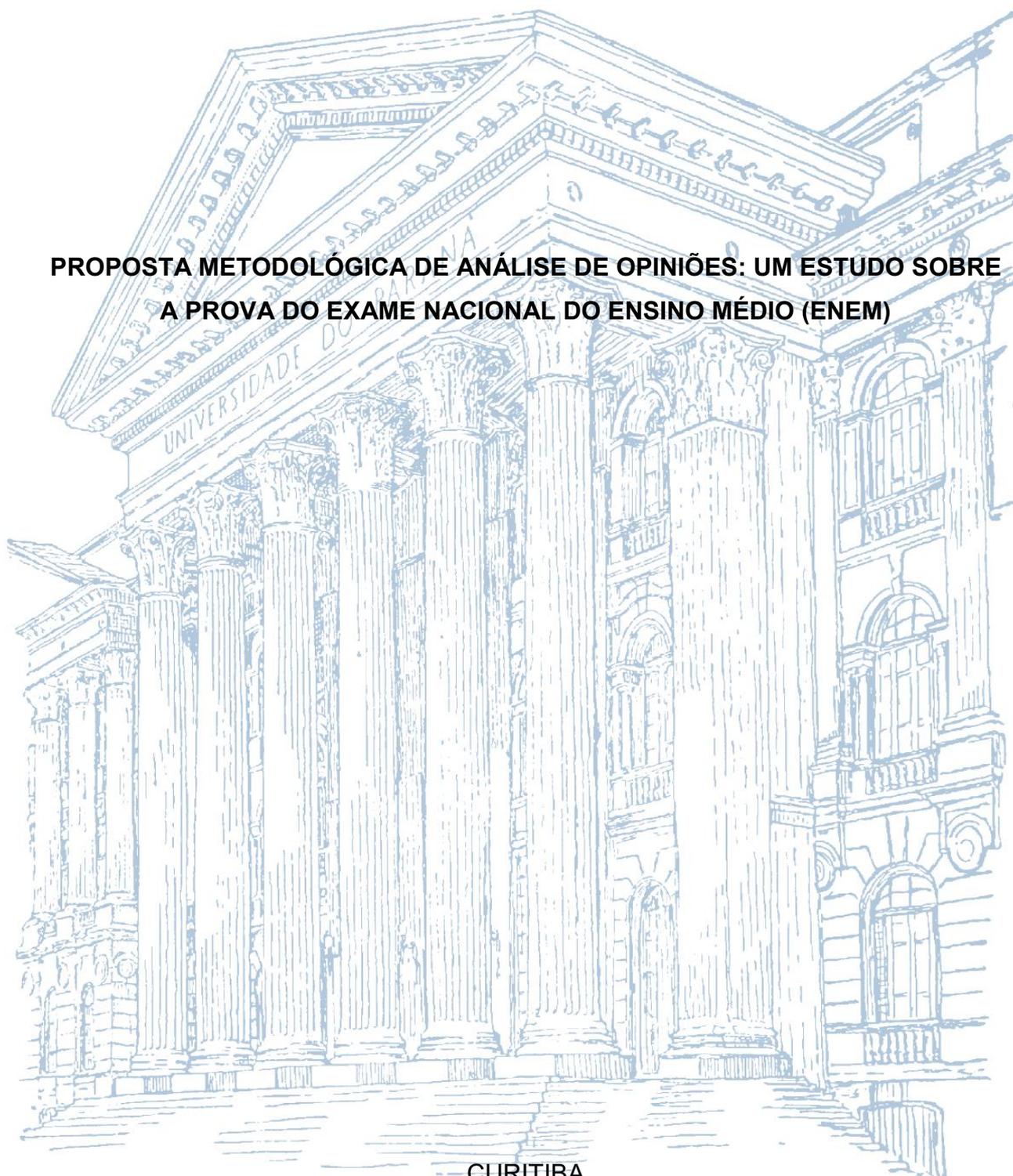


UNIVERSIDADE FEDERAL DO PARANÁ

ANA PAULA FANTIN ZELLNER

**PROPOSTA METODOLÓGICA DE ANÁLISE DE OPINIÕES: UM ESTUDO SOBRE
A PROVA DO EXAME NACIONAL DO ENSINO MÉDIO (ENEM)**



CURITIBA

2016

ANA PAULA FANTIN ZELLNER

**PROPOSTA METODOLÓGICA DE ANÁLISE DE OPINIÕES: UM ESTUDO SOBRE
A PROVA DO EXAME NACIONAL DO ENSINO MÉDIO (ENEM)**

Trabalho apresentado como requisito parcial à obtenção do grau de Bacharel em Gestão da Informação no curso de graduação em Gestão da Informação, Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná.

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

CURITIBA

2016

TERMO DE APROVAÇÃO

ANA PAULA FANTIN ZELLNER

“METODOLOGIA DE ANÁLISE DE OPINIÕES SOBRE A PROVA DO EXAME NACIONAL DO ENSINO MÉDIO (ENEM)”

Monografia aprovada como requisito parcial à obtenção do grau de Bacharel em Gestão da Informação no curso de graduação em Gestão da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná. Universidade Federal do Paraná, pela seguinte banca examinadora:

Profa. Dra. Denise Fukumi Tsunoda
(Orientadora / UFPR)

Prof. Dr. José Simão de Paula Pinto
(Examinador / UFPR)

Prof. Me. André José Ribeiro Guimarães
(Examinador / UFPR)

Curitiba, 06 de dezembro de 2016

*Dedico esse trabalho ao meu
companheiro, meus familiares e amigos
pela paciência e compreensão de minha
ausência e apoio dado para o
desenvolvimento desta pesquisa.*

AGRADECIMENTOS

À Professora Dra. Denise Fukumi Tsunoda, por toda a paciência nas orientações, pelas dicas, conselhos, recomendações, apoio emocional e confiança que foram fundamentais para que me motivasse a busca de conhecimento capaz de terminar este estudo e semear a vontade de seguir nesta área.

Aquele que eu amo e confio por estar ao meu lado me auxiliando e me motivando a continuar com meus sonhos.

Aos colegas e grandes amigos(as) do curso de Bacharelado em Gestão da Informação, pelas conversas, discussões e conselhos que auxiliaram o nascimento e conclusão dessa pesquisa.

A Universidade Federal do Paraná, seu corpo docente, direção e administração que oportunizaram ótimas estruturas físicas, professores e espaço para que fosse possível concluir este curso.

A todos meus professores pela passagem de conhecimento e de experiência de vida e que contribuíram de alguma forma para minha formação profissional e acadêmica. Vocês são fundamentais na formação de qualquer profissional e por isso meus professores têm o meu eterno agradecimento

Aos familiares e amigos que estiveram comigo, mesmo que por meio online foram capazes de transmitir apoio.

A todos que auxiliaram de alguma forma na conclusão deste estudo, respondendo pesquisa, dando opiniões e dicas para que pudesse ter se concluído, obrigada!

*“Todo o nosso conhecimento se inicia
com sentimentos”.*

Leonardo da Vinci

RESUMO

Trata-se de um estudo sobre mineração de opinião que utilizará de algoritmos de mineração de dados para descoberta de padrões e conhecimento nesta base de dados. O objetivo geral deste estudo é criar uma metodologia de análise de opiniões sobre a prova do Exame Nacional do Ensino Médio (ENEM). Para atingir esse objetivo, foram executadas as seguintes etapas: prospectar estudos relacionados ao tema da pesquisa por meio de pesquisa exploratória; propor uma base de dados para armazenar as opiniões coletadas de alunos e usuários do Facebook e Twitter de cursos preparatórios para a prova do ENEM da região central de Curitiba; e descobrir padrões de opiniões entre alunos e usuários do Facebook e Twitter pelo uso de métodos e técnicas de mineração de dados. A coleta de dados foi realizada de duas formas: por meio da aplicação de questionário para levantamento de opiniões e recuperação de opiniões em redes sociais (*Facebook* e *Twitter*). A proposta da metodologia de análise e mineração de opiniões permite a reaplicação deste estudo em outras pesquisas. Os resultados obtidos apresentaram tendenciosidade devido ao número de respostas pequeno. As amostras desejadas não foram atingidas e para não manter a tendenciosidade foram retirados alguns atributos. Pode-se identificar que os alunos tanto de cursos preparatório quanto da UFPR não se preparam menos de um ano para o ENEM e aqueles que se preparam mais aplicaram melhor o conhecimento adquirido na prova. As sentenças coletadas no Facebook e Twitter apresentaram em opiniões negativas dos usuários em relação ao (a): tempo de prova, questões cobradas, site de inscrição, resultado da prova e nível de dificuldade da prova dentre outros. As sentenças que apresentaram sentimentos positivos estavam relacionadas ao tema da redação do ENEM 2015: “a persistência da violência contra a mulher na sociedade brasileira”. Como proposta de trabalhos futuros sugere-se replicar a metodologia aqui proposta em outras pesquisas envolvidas na área de análise de sentimento, mineração de opiniões e mineração de dados. Outra sugestão é ampliar a amostra da pesquisa para outras regiões brasileiras.

Palavras Chave: Mineração de Opinião. Base de Dados. Redes Sociais.

ABSTRACT

This is a study on opinion mining that will use data mining algorithms to discover patterns and knowledge in this database. The general objective of this study is to create a methodology of analysis of opinions on the test of the National High School Examination (NHSC). To achieve this goal, the following steps were performed: prospect studies related to the research topic through exploratory research; to propose a database to store the collected opinions of students and users of Facebook and Twitter of preparatory courses for the NHSC test of the central region of Curitiba; and discover patterns of opinions among students and users of Facebook and Twitter through the use of methods and techniques of data mining. Data collection was done in two ways: through the application of a questionnaire to survey opinions and retrieve opinions on social networks (Facebook and Twitter). The proposal of the methodology of analysis and mining of opinions allows the reapplication of this study in other researches. The results obtained showed bias due to the small number of responses. The desired samples were not reached and in order not to maintain the bias, some attributes were removed. It is possible to identify that the students of both preparatory and UFPR courses do not prepare less than a year for the NHSC and those who prepare themselves more applied the knowledge acquired in the test. The sentences collected on Facebook and Twitter presented negative opinions from users regarding: test time, questions charged, registration site, test result and level of difficulty of the test among others. Sentences that present positive feelings were related to the theme of the writing of NHSC 2015: "the persistence of violence against women in Brazilian society". As a proposal of future work, it is suggested to replicate the methodology proposed here in other researches involved in the area of sentiment analysis, opinion mining and data mining. Another suggestion is to expand the sample of the research to other Brazilian regions.

Keywords: *Opinion Mining. Database. Social networks.*

LISTA DE FIGURAS

FIGURA 1 – CICLO DE VIDA DA INFORMAÇÃO.	25
FIGURA 2 – CICLO DE VIDA EVOLUTIVO DA INFORMAÇÃO.	26
FIGURA 3 – COMPONENTES DE UM PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO.	28
FIGURA 4 – UMA VISÃO GERAL DAS ETAPAS QUE COMPÕEM O PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS (KDD).	30
FIGURA 5 – A MINERAÇÃO DE DADOS EXEMPLIFICADA EM ETAPA BASEADAS NO PROCESSO DE DESCOBERTA DE CONHECIMENTO (KDD).	31
FIGURA 6 – PÁGINA DO TWITTER ONDE É POSSÍVEL OBSERVAR A PERGUNTA "O QUE ESTÁ ACONTECENDO?"	39
FIGURA 7 – LAYOUT DE DIVULGAÇÃO DE UMA PÁGINA NO FACEBOOK.	41
FIGURA 8 – RESULTADOS DA PESQUISA POR CURSOS PREPARATÓRIOS EM CURITIBA REALIZADA NO GOOGLE MAPS.	45
FIGURA 9 – PRÉVIA DAS COLUNAS ELABORADAS NO EXCEL 2013 COM BASE NO QUESTIONÁRIO APLICADO.	55
FIGURA 10 – PRÉVIA DA ANÁLISE REALIZADA COM OS DADOS COLETADOS DAS REDES SOCIAIS NA FERRAMENTA SEMANTRIA.	55
FIGURA 11 – VISUALIZAÇÃO DA BASE REPRESENTADA A PARTIR DA FERRAMENTA WEKA.	56
FIGURA 12 - PARÂMETROS DO APRIORI NO WEKA PARA MINERAR OS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIOS.	75
FIGURA 13 - ATRIBUTOS RESTANTES DEPOIS DE REMOVER "IDADE" E "NÚMERO DE QUESTÕES É COMPATÍVEL"	78
FIGURA 14 - PARÂMETROS DEFAULT DO J48 NO WEKA PARA MINERAR OS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIOS.	83
FIGURA 15 – ÁRVORE DE DECISÃO REPRESENTANDO OS DADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIO EXECUTADO PELO ALGORITMO J48 NO WEKA (2016).	85

FIGURA 16 – ÁRVORE DE DECISÃO REPRESENTANDO OS DADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DA UFPR EXECUTADO PELO ALGORITMO J48 NO WEKA (2016).....	87
FIGURA 17 - FLUXO DA METODOLOGIA PARA ANALISAR SENTIMENTOS E MINERAR OPINIÕES.	96

LISTA DE TABELAS

TABELA 1 – DADOS COLETADOS REFERENTES AOS PERIÓDICOS QUE ENVOLVAM O TEMA ENEM COLETADOS DA BASE WEB OF SCIENCE.	19
TABELA 2 – TÍTULOS DOS PERIÓDICOS RELACIONADOS AO TEMA ENEM COLETADOS DA BASE SCOPUS.	20
TABELA 3 – QUANTIDADE DE ALUNOS NOS CURSOS DE ADMINISTRAÇÃO, CIÊNCIAS CONTÁBEIS, CIÊNCIAS ECONÔMICAS E GESTÃO DA INFORMAÇÃO MATRICULADOS NO SEGUNDO SEMESTRE DE 2016.	46
TABELA 4 – PERFIS SELECIONADOS E NÃO SELECIONADOS DO <i>FACEBOOK</i> PARA A COLETA DE OPINIÕES.	47
TABELA 5 – NÚMERO DE ALUNOS DA UFPR DE ACORDO COM A IDADE.	60
TABELA 6 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO DE ACORDO COM A IDADE.	61
TABELA 7 – TOTAL DE RESPOSTAS ADICIONAIS DA QUESTÃO SE OS ALUNOS DOS CURSOS PREPARATÓRIO CONSIDERAM OU NÃO O CONHECIMENTO EM SALA DE AULA SUFICIENTE.	67
TABELA 8 – TOTAL DE RESPOSTAS ADICIONAIS DA QUESTÃO SE OS ALUNOS DOS CURSOS PREPARATÓRIO CONSIDERAM OU NÃO O CONHECIMENTO EM SALA DE AULA SUFICIENTE.	68
TABELA 9 – TOTAL DE RESPOSTAS ADICIONAIS A RESPOSTA DA QUESTÃO SE OS ALUNOS DA UFPR CONSIDERAVAM OU NÃO O ENEM COMO UMA FERRAMENTA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.	70
TABELA 10 – TOTAL DE RESPOSTAS ADICIONAIS A RESPOSTA DA QUESTÃO SE OS ALUNOS DOS CURSOS PREPARATÓRIO CONSIDERAVAM OU NÃO O ENEM COMO UMA FERRAMENTA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.	72
TABELA 11 - RESULTADOS DA COLETA DO QUESTIONÁRIO FEITO AOS ALUNOS DOS CURSOS PREPARATÓRIO POR MEIO DO ALGORITMO APRIORI EXECUTADO NO WEKA.	77
TABELA 12 - RESULTADOS APRESENTADOS PELO WEKA DOS DADOS COLETADOS DOS RESPONDENTES DOS CURSOS PREPARATÓRIOS APÓS RETIRAR OS ATRIBUTOS "IDADE" E " NÚMERO DE QUESTÕES É COMPATÍVEL".	80

TABELA 13 - RESULTADOS DA COLETA DO QUESTIONÁRIO FEITO AOS ALUNOS DA UFPR POR MEIO DO ALGORITMO APRIORI EXECUTADO NO WEKA.	82
TABELA 14 - MATRIZ DE CONFUSÃO DO ALGORITMO J48 OBTIDO COM A MINERAÇÃO DOS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIO.....	84
TABELA 15 - MATRIZ DE CONFUSÃO DO ALGORITMO J48 OBTIDO COM A MINERAÇÃO DOS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DA UFPR.....	85
TABELA 16 – CONTAGEM DE SENTENÇAS NEGATIVAS, NEUTRAS E POSITIVAS DAS REDES SOCIAIS TWITTER E FACEBOOK.....	88
TABELA 17 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO POSITIVAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.....	89
TABELA 18 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEGATIVAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.....	90
TABELA 19 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEGATIVAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.....	91
TABELA 20 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEUTRAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.....	92
TABELA 21 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEUTRAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.....	93

LISTA DE GRÁFICOS

GRÁFICO 1 – QUANTIDADE DE RESPONDENTES DE ACORDO COM O CURSO.	57
GRÁFICO 2 – QUANTIDADE DE ALUNOS MATRICULADOS E NÃO MATRICULADOS NA UFPR.	58
GRÁFICO 3 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS POR GÊNERO.	59
GRÁFICO 4 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS DE ACORDO COM O TIPO DE ESCOLA.	60
GRÁFICO 5 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS DE ACORDO COM O TEMPO DE PREPARAÇÃO.	62
GRÁFICO 6 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS QUE TIVERAM PROBLEMA COM O SISTEMA DE INSCRIÇÃO.	63
GRÁFICO 7 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS QUE CONSIDERAM OU NÃO O PREÇO COBRADO PARA REALIZAR O ENEM “JUSTO”.	64
GRÁFICO 8 – QUANTIDADE DE RESPOSTAS DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS CONSIDERANDO O CONTEÚDO COBRADO NA PROVA EM RELAÇÃO AO SEU CONHECIMENTO ADQUIRIDO.	65
GRÁFICO 9 – NÚMERO DE ALUNOS DA UFPR QUE CONSIDERAM OU NÃO QUE HÁ PREPARAÇÃO SUFICIENTE PARA REALIZAR A PROVA DO ENEM.	65
GRÁFICO 10 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO QUE CONSIDERAM OU NÃO QUE HÁ PREPARAÇÃO SUFICIENTE PARA REALIZAR A PROVA DO ENEM.	66
GRÁFICO 11 – NÚMERO DE ALUNOS DA UFPR QUE CONSIDERARAM OU NÃO O NÚMERO DE QUESTÃO DE ACORDO COM O TEMPO DE REALIZAÇÃO DE PROVA.	68
GRÁFICO 12 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO QUE CONSIDERARAM OU NÃO O NÚMERO DE QUESTÃO DE ACORDO COM O TEMPO DE REALIZAÇÃO DE PROVA.	69
GRÁFICO 13 – NÚMERO DE ALUNOS DA UFPR QUE CONSIDERAM OU NÃO O ENEM COMO UMA FORMA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.	69

GRÁFICO 14 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO QUE CONSIDERAM OU NÃO O ENEM COMO UMA FORMA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.	71
GRÁFICO 15 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS QUE CONCORDAM OU NÃO COM A CERTIFICAÇÃO DO ENSINO MÉDIO OBTENDO 450 PONTOS NO ENEM.	74

SUMÁRIO

1	INTRODUÇÃO	16
1.1	PROBLEMA DE PESQUISA	16
1.2	OBJETIVOS	17
1.2.1	Objetivo Geral	17
1.2.2	Objetivos Específicos	17
1.3	JUSTIFICATIVA	18
1.4	LIMITAÇÕES DA PESQUISA	21
1.5	ESTRUTURA DO DOCUMENTO	21
2	REFERENCIAL TEÓRICO	23
2.1	GESTÃO DA INFORMAÇÃO	23
2.1.1	Dado, informação e conhecimento	24
2.1.2	Recuperação da Informação	27
2.2	DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS (KDD)	29
2.2.1	Mineração de dados	32
2.2.2	Tarefas e métodos de mineração de dados	33
2.2.3	Mineração de opinião: definição	35
2.2.4	Banco de dados	36
2.3	REDE SOCIAL	37
2.3.1	A rede social Twitter	38
2.3.2	A rede social Facebook	40
2.4	EXAME NACIONAL DO ENSINO MÉDIO - ENEM	41
2.4.1	Formas de utilização do ENEM	42
2.4.2	Inscrição	43
2.4.3	A prova do ENEM	43
3	CARACTERIZAÇÃO E PROCEDIMENTOS METODOLÓGICOS	44
3.1	CARACTERIZAÇÃO DA PESQUISA	44
3.2	POPULAÇÃO E AMOSTRA	44
3.3	ETAPAS DE COLETA E ARMAZENAMENTO DOS DADOS	48
3.3.1	Coleta de dados por meio do questionário	48
3.3.2	Coleta de opiniões nas redes sociais	50
3.3.3	Armazenamento dos dados	50

3.4	ETAPAS DE CRIAÇÃO DA BASE DE DADOS COM OS DADOS COLETADOS POR MEIO DO QUESTIONÁRIO	51
3.5	ETAPAS PARA CRIAÇÃO DA BASE DE DADOS COM AS OPINIÕES COLETADAS DAS REDES SOCIAIS	52
3.5.1	Subjetividade	52
3.5.2	Extração.....	53
3.5.3	Classificação.....	53
3.5.4	Sumarização.....	54
3.6	MATERIAIS E MÉTODOS	54
4	RESULTADOS	57
4.1	ANÁLISE ESTATÍSTICA DA COLETA	57
4.2	MINERAÇÃO DE DADOS DA COLETA DO QUESTIONÁRIO	74
4.3	ANÁLISE DE SENTIMENTO DOS DADOS COLETADOS DAS REDES SOCIAIS	88
4.4	METODOLOGIA DE ANÁLISE DE OPINIÕES SOBRE A PROVA DO ENEM	93
5	CONSIDERAÇÕES FINAIS	97
5.1	VERIFICAÇÃO DOS OBJETIVOS PROPOSTOS	97
5.2	TRABALHOS FUTUROS.....	98
5.3	CONTRIBUIÇÕES DA PESQUISA.....	99
	REFERÊNCIAS	100
	APÊNDICE A – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO PRÉ-TESTE	105
	APÊNDICE B – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO APLICADA AOS ALUNOS DOS CURSOS PREPARATÓRIOS	107
	APÊNDICE C – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO APLICADA AOS ALUNOS DA UFPR	110
	APÊNDICE D – BASE DE DADOS “OPINIÕES ENEM”	113

1 INTRODUÇÃO

As pessoas em seu convívio social compartilham de ideias e opiniões sobre determinado acontecimento. Essas informações podem ser compartilhadas em forma de conversas ou textos em ambientes virtuais. Boas ou ruins, as opiniões não estão estruturadas formalmente, impossibilitando retirar análises para tomadas de decisões.

Métodos vêm sendo utilizados para recuperar dados em repositórios de dados, entende-se – no âmbito desta pesquisa - como repositório de dados as bases de dados e a própria Web. A Mineração de Dados é uma área que utiliza algoritmos para absolver essa quantidade de dados a fim de auxiliar o homem na análise e possivelmente tirar resultados conclusivos. Opiniões de usuários estão contidas nos repositórios de dados, e para coletá-las e analisá-las são utilizadas técnicas de Mineração de Opinião, ou também conhecida como Análise de Sentimento. A Mineração de Opinião possui como objeto de estudo palavras opinativas, ou seja, palavras que demonstram opinião do usuário sobre um acontecimento, por exemplo: legal, ruim, ótimo, péssimo entre outros.

Para a aplicação de métodos de Mineração de Opinião é necessário possuir um repositório de dados, assim é possível aplicar um algoritmo e dele retirar resultados que geram conhecimento. Este estudo tem por objetivo propor uma metodologia de análise de opiniões sobre a prova do ENEM. As técnicas escolhidas para este estudo são classificadas como algoritmos de associação e árvores de decisão, dentro destas categorias encontram-se o Apriori e o C4.5 por exemplo.

Esta pesquisa apresentará o referencial teórico da área de estudo, as etapas utilizadas para o desenvolvimento da base de opiniões e a sumarização dos resultados a partir da aplicação do algoritmo de Mineração de Opiniões.

1.1 PROBLEMA DE PESQUISA

A mineração de opinião começou a ganhar interesse em publicação de pesquisas no ano de 2001. Nos primórdios ela agregava temas relacionados ao processamento de linguagem natural, recuperação de informação, inteligência artificial e inteligência computacional (PANG; LEE, 2008).

As opiniões de um grupo de pessoas podem influenciar a forma com que outras pessoas pensam e se comportam. Essas opiniões acabam sendo expressas

em forma de postagens em redes sociais ou até mesmo em conversas em grupos. Isto faz com que as possibilidades de buscar opiniões aumentem devido à diversidade de plataformas que não estão mais limitadas a apenas opiniões de sua própria rede pessoas ou de especialistas (BECKER; TUMITAN, 2013). A internet é um recurso que tem sido explorado como fonte de opiniões devido à diversidade de seus usuários. Diante disso, surge a necessidade de construir técnicas que auxiliem a obtenção e recuperação desse tipo de informação.

O Exame Nacional do Ensino Médio (ENEM) é uma forma de avaliar o desempenho de alunos do último ano do ensino médio das escolas de todo o Brasil. Os alunos, em alguns casos, expressam suas opiniões referente à prova do ENEM na internet – em postagens nas redes sociais – ou por meio de conversas informais. Essas opiniões não são registradas em mecanismos formais que permitam a sua recuperação.

A partir do exposto busca-se resolver o principal problema deste estudo: como propor uma metodologia de análise de opiniões sobre a prova do ENEM?

1.2 OBJETIVOS

Após a análise e definição do problema de pesquisa foi possível estabelecer os objetivos a serem seguidos neste projeto. Nas seções a seguir estarão definindo o objetivo geral e específico do presente estudo.

1.2.1 Objetivo Geral

Este estudo tem por objetivo geral propor uma metodologia de análise de opiniões sobre a prova do ENEM.

1.2.2 Objetivos Específicos

Para atingir o objetivo geral deste estudo deverão ser atingidos os objetivos específicos a seguir:

- a) prospectar estudos relacionados ao tema da pesquisa;
- b) definir uma metodologia para criação desta base de dados;

- c) registrar as opiniões e os dados coletados referente aos estudantes e usuários do Facebook e Twitter que realizaram o ENEM em uma base de dados; e
- d) descobrir padrões de opiniões entre alunos e usuários do Facebook e Twitter pelo uso de métodos e técnicas de mineração de dados.

1.3 JUSTIFICATIVA

O tema mineração de opinião aplicada no levantamento de dados sobre o ENEM ainda é uma área que precisa ser explorada. Foi realizado um levantamento nas bases internacionais *Web of Science* da empresa Thompson e na base *Scopus* da empresa Elsevier referente ao tema “ENEM”. Essa busca objetivou-se analisar se houve publicações em periódicos da área da informação ou da área de tecnologia que envolvesse o tema deste estudo. Os resultados trazidos pela *Web of Science* resultaram em 46 títulos relacionados ao assunto.

No final da pesquisa realizada na base da *Web of Science*, pode-se verificar que existiram dois documentos (identificados na TABELA 1) publicados em periódicos da área da Tecnologia e outro da área de Educação. Porém nenhum dos dois abrange a área de Mineração de Opinião e/ou Análise de Sentimento. O documento do periódico da área da Tecnologia o “*Journal of Information Science and Engineering*” (TABELA 1 e TABELA 2) é referente ao artigo “*Elman network with embedded memory for system identification*”, onde o termo “ENEM” utilizado pelos autores remete a “*Elman Network with Embedded Memory*” um tipo de estrutura de rede neural (KALINLI; SAGIROGLU, 2006), do qual não é o foco deste estudo. E com relação ao periódico da área da Educação o “*Texto Livre-Linguagem e Tecnologia*”, ao pesquisar o artigo “*Technology And Semiotic Power: Writing, Today*” foi possível destacar que o mesmo se trata de um estudo sobre aspectos da produção de textos aplicado ao ENEM (RIBEIRO, 2005). Este tema não envolve a área da tecnologia apesar do periódico abranger a área.

TABELA 1 – DADOS REFERENTES AOS PERIÓDICOS QUE ENVOLVAM O TEMA ENEM COLETADOS DA BASE WEB OF SCIENCE.

Periódicos	Total
Ensaio: Avaliação e Políticas Públicas em Educação	8
Educação em Revista	4
Revista Brasileira de Ensino de Física	3
Ensaio Pesquisa em Educação em Ciências (Belo Horizonte)	2
Educar em Revista	2
Avaliação: Revista da Avaliação da Educação Superior (Campinas)	2
Bolema: Boletim de Educação Matemática	2
Educação & Sociedade Estudos Econômicos (São Paulo)	2
CEPAL REVIEW	1
Revista Brasileira de Economia	1
CHALLENGES OF EDUCATION IN BRAZIL	1
ENCOUNTERS IN THEORY AND HISTORY OF EDUCATION	1
Revista Brasileira de Educação Especial Ciência & Educação (Bauru)	1
Revista Brasileira de Estudos Pedagógicos	1
Cadernos de Pesquisa	1
Psicologia em Estudo	1
SOCIOINT14: International Conference on Social Sciences and Humanities	1
Revista Brasileira de Educação	1
STROJNISKI VESTNIK-JOURNAL OF MECHANICAL ENGINEERING	1
Revista Brasileira de Educação Médica	1
3RD INTERNATIONAL CONFERENCE OF EDUCATION, RESEARCH AND INNOVATION (ICERI201)	1
Revista Brasileira de Estudos de População	1
MOVIMENTO	1
Revista Brasileira de História	1
POSITIVITY	1
Proposições	1
TEXTO LIVRE-LINGUAGEM E TECNOLOGIA	1
JOURNAL OF INFORMATION SCIENCE AND ENGINEERING	1
Journal of the Japan Society of Coloproctology	1
Total Geral	49

FONTE: Elaborado pela autora com base nos dados da pesquisa (2016).

Na pesquisa realizada na Scopus, os resultados obtidos foram 35 registros. Desses 35 registros, um documento foi publicado em um periódico da área da ciência informação. Este documento é o mesmo que apareceu nos registros coletados da Web of Science.

TABELA 2 – TÍTULOS DOS PERIÓDICOS RELACIONADOS AO TEMA ENEM COLETADOS DA BASE SCOPUS.

Periódicos	Total
Ensaio	4
Educação e Sociedade	2
Revista Brasileira de Ensino de Física	2
Bolema – Mathematics Education Bulletin	2
Meta: Avaliação	2
Chemical and Pharmaceutical Bulletin	1
Research on Population Ecology	1
Diálogos	1
Cepal Review	1
Revista Brasileira de Educação	1
CEUR Workshop Proceedings	1
Revista degli Studi Orientali	1
Estudos Econômicos	1
Psicologia em Estudo	1
Journal of Information Science and Engineering	1
Revista Brasileira de Economia	1
Life Science Journal	1
Revista Brasileira de Educação Especial	1
Strojinski Vestnik/Journal of Mechanical Engineering	1
Revista Brasileira de Estudos de População	1
The RAINS model of acidification: science and strategies in Europe	1
Schweizerische Medizinische Wochenschrift	1
Zentralblatt für Allgemeine Pathologie und Pathologische Anatomie	1
Pediatric Infectious Disease Journal	1
Zeitschrift für Kinderheilkunde	1
Monatsschrift für Kinderheilkunde	1
Arzneimittel-Forschung/Drug Research	1
Movimento	1
Total Geral	35

FONTE: Elaborado pela autora com base nos dados da pesquisa (2016).

Em ambas as pesquisas, os periódicos com maiores registros de documentos referentes ao tema ENEM são da área de educação. Conforme os resultados apresentaram, nenhum dos periódicos anteriores abrange a área de Mineração de Opinião e/ou Análise de Sentimento. E não há evidências nessas bases de dados de estudos realizados que envolva o tema ENEM com a Mineração de Opinião.

Além dessa análise científica que justifica a falta de pesquisa nessas duas áreas a Mineração de Opinião e a Análise de Sentimento são áreas que despertaram a curiosidade da autora em experiências passadas, porém o estudo não foi realizado com profundidade. Este fato motivou o aprofundamento de pesquisas na área.

O ENEM é uma prova que é realizada em todo o Brasil, no ano de 2015 foram 7,7 milhões de inscritos (PORTAL BRASIL, 2015). E por ser um mecanismo de avaliação do ensino médio é importante que a prova expresse o conhecimento dos candidatos. Por esse motivo este estudo busca formalizar as opiniões dos envolvidos (alunos e usuário do Twitter e Facebook) de forma a, eventualmente, subsidiar outras edições da prova do ENEM.

1.4 LIMITAÇÕES DA PESQUISA

Esta pesquisa possui limitações quanto à sua coleta de dados, população e amostragem. Para este estudo estão limitadas apenas as opiniões dos realizadores do ENEM 2015 da região central de Curitiba. A escolha da amostra considerável de pequena proporção para a abrangência do ENEM foi devido à facilidade de acesso e contatos com os cursos preparatórios e com os alunos do Sociais Aplicadas da UFPR, se limitando ao município de Curitiba.

Conforme comentado por Izquierdo, Bevilaqua e Cammar (2006) após realizar uma prova de vestibular, ou no caso o ENEM, os conteúdos estudados não perduram na memória. Devido a esta constatação optou-se utilizar opiniões apenas dos realizadores do ENEM 2015 por ser a última edição desta prova.

A Mineração de Opinião possui um foco para diversos estudos e pesquisas. Porém, sua total explanação está fora do escopo desse trabalho, limitando-se apenas no objetivo geral deste estudo.

1.5 ESTRUTURA DO DOCUMENTO

Este documento está organizado em seções. O embasamento teórico desta pesquisa está apresentado na seção 2 e abrange conceitos aplicados à Gestão da Informação; Mineração de Dados; Descoberta de Conhecimento em Base de Dados

(KDD); Mineração de Opinião; Base de Dados, Redes Sociais por fim, conceitos relacionados ao ENEM e suas definições.

Na seção 3 estão apresentados a descrição e os procedimentos metodológicos escolhidos, a metodologia de pesquisa e sua população e amostragem.

Na seção 4 estão apresentados os resultados desta pesquisa assim como a análise dos resultados baseados nas tarefas de classificação e regras de associação da mineração de dados.

Por fim na seção 6 são descritas as considerações finais deste estudo e sugestões de pesquisas futuras na área.

2 REFERENCIAL TEÓRICO

Nas seções seguintes, serão apresentados conceitos que fundamentam a pesquisa e proposta deste estudo. Serão abordados os conceitos aplicados à gestão da informação; Descoberta de Conhecimento em Base de Dados (KDD); Data Mining; Mineração de Opinião; Redes Sociais, por fim conceitos relacionados ao ENEM e suas definições.

A proposta desta pesquisa foi desenvolvida com o auxílio desses conceitos e ajudaram no desenvolvimento da metodologia para a construção da base de dados.

2.1 GESTÃO DA INFORMAÇÃO

A informação passou a ganhar maior abrangência de armazenamento e recuperação a partir do surgimento de novas tecnologias. Esse avanço científico surgiu a principalmente a partir da segunda guerra mundial, e com isso a necessidade social, histórica, cultural e política do registro e transmissão de informação (PINHEIRO, 2002).

O marco inicial da utilização do termo gestão da informação na literatura pode ser encontrado nos trabalhos de Paul Otlet em 1934 e Vanevar Bush e Frederick von Hayek em 1945 (BARBOSA, 2008). Barbosa (2008) descreveu que Otlet (1934) em seu livro "*Traité de documentation*" abordou sobre a documentação, que era entendida como um processo de comunicação e transmissão do conhecimento, e da criação e evolução das instituições. Bush (1945) descreveu em seu artigo "*As we may think*" sobre uma máquina (*Memex*) capaz de armazenar livros e recuperá-los rapidamente, ideia considerada como precursora da Web e da gestão eletrônica de documentos (GED) (BARBOSA, 2008). Por fim, Von Hayek (1945) publicou o artigo "*The use of knowledge in society*" que o conhecimento nunca se encontra em forma concentrada ou integrada que garanta o melhor uso dos recursos conhecidos por cada um dos membros da sociedade (BARBOSA, 2008). A partir dessas obras é possível perceber a preocupação por traz do termo relacionado à informação.

O conceito de gestão da informação é geralmente apresentado sob um contexto organizacional, como aquela capaz de aplicar princípios encontrados na administração a partir do gerenciamento informacional: determinação das necessidades, obtenção, distribuição e uso da informação (DAVENPORT, 1998).

Marchiori (2002) apresenta a gestão da informação sob o enfoque da ciência da informação, como aquela que se ocupa em estudar a informação envolvendo a identificação, coleta, validação, representação, recuperação e uso para um usuário.

Para aplicar esses processos da gestão da informação há a necessidade de um profissional que apresente conhecimentos interdisciplinares, habilidades gerenciais, técnicas e políticas, além de atitude ética. Além disso, deve possuir habilidades de análise, condensação, interpretação, representação e estratégias de busca e apresentação/formatação da informação considerando diferentes suportes e canais que a informação está disponibilizada (MARCHIORI, 2002, p. 75).

Além do exposto, a gestão da informação engloba conceitos de dado, informação e conhecimento. A próxima seção abordará esses conceitos e definições.

2.1.1 Dado, informação e conhecimento

Davenport (1998) conceitua dado como simples observações sobre o estado do mundo que podem ser facilmente estruturados, facilmente obtido por máquinas, facilmente transferível e frequentemente quantificado. Para Dante (1998, *tradução da autora*) dados representam registros icônicos, simbólicos (fonéticos ou numéricos) ou signos (linguísticos, lógicos ou matemáticos) por meio dos quais se representam feitos, conceitos ou instruções.

O conceito de informação apresentado por Buckland (1991) é definido sob três enfoques: como processo; como conhecimento e como coisa, conforme a seguir:

- informação como processo: a informação é o ato de informar ou comunicar o conhecimento de algum fato ou ocorrência;
- informação como conhecimento: informação é usada para denotar aquilo que é percebido enquanto um processo e comunicar o conhecimento referente a algum fato ou evento;
- informação como coisa: o termo “informação” pode ser atribuído para objetos considerados como informativos.

Dante (1998) apresenta o conceito de informação como dados ou matéria informacional relacionada ou estruturada de maneira atual ou potencialmente significativa. E mediante a processos de análises (separação, validação, comparação) que agregam valor se converte em conhecimento informativo (DANTE, 1998, *tradução*

da autora). Para a autora a informação não é só dinâmica no sentido de sua permanência transferência e também no sentido da sua própria evolução decidindo à medida que vai fluindo e vai se modificando em forma permanente, e com base nisso, a autora ilustrou a informação em um ciclo, apresentado na FIGURA 1.

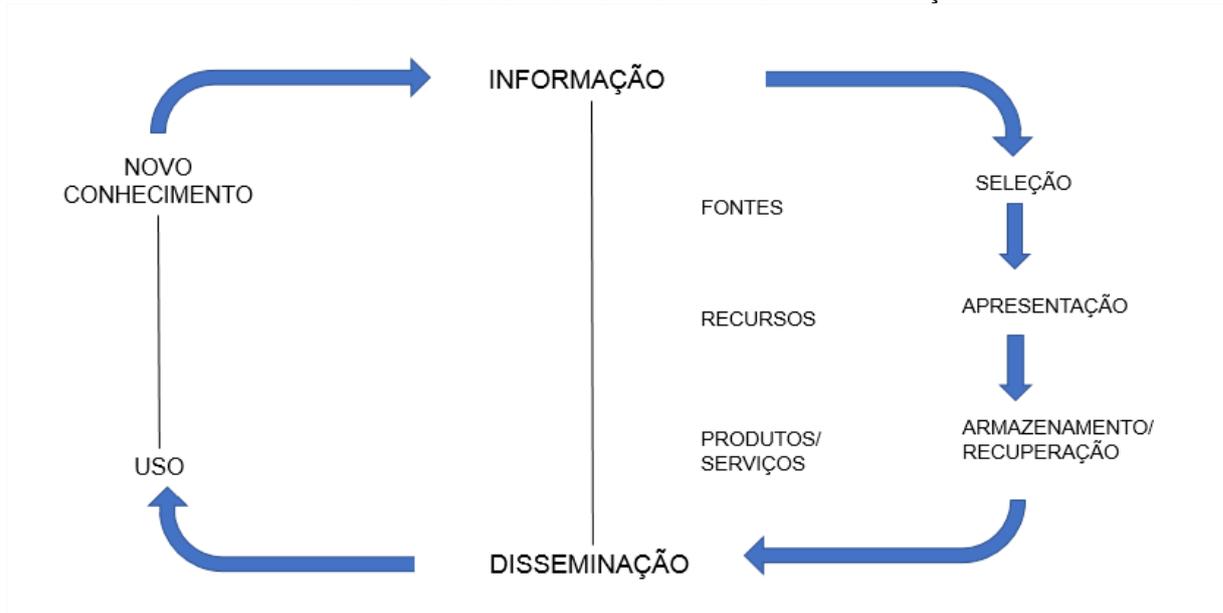
FIGURA 1 – CICLO DE VIDA DA INFORMAÇÃO.



FONTE: Adaptado de Dante (1998).

Além desse ciclo, a autora comenta que informação pode evoluir e ser transferida apresentado outro ciclo, que demonstra a evolução da informação representado na FIGURA 2.

FIGURA 2 – CICLO DE VIDA EVOLUTIVO DA INFORMAÇÃO.



FONTE: Adaptado de Dante (1998).

Este ciclo apresenta que a evolução da informação perpassa as fases: informação, seleção, apresentação, armazenamento/recuperação, disseminação, uso e novo conhecimento produzindo um gerador de três tipos de bens: as fontes, os recursos e os produtos/serviços de informação (DANTE, 1998).

Além disso, pode-se definir informação como dados dotados de relevância e propósito a partir de análises elaboradas por pessoas ou computadores (DAVENPORT, 1998). Davenport (1998) e Setzer (1999) argumentam que para a informação ser armazenada em um computador ela deve ser reduzida ou transformada em dado. A única forma de administrar a complexidade da informação é criar bancos de dados em computadores (DAVENPORT, 1998).

Setzer (1999) apresenta as diferenças básicas entre dado, informação e conhecimento. Para o autor os dados são uma sequência de símbolos quantificados ou quantificáveis e podem obviamente ser armazenados em um computador e processados por ele. A informação, para o autor, representa algo significativo para alguém através de textos, imagens, sons ou animação e não pode ser diretamente em um computador. Setzer (1999) apresenta o conhecimento como uma abstração interior, pessoal, de alguma coisa que foi experimentada por alguém que não pode ser inserido em um computador por meio de uma representação, pois senão foi reduzido a uma informação e conseqüentemente a dados. A partir disso, o autor descreve que é equivocado falar-se de uma “base de conhecimento” em um

computador, no máximo poderíamos abordar sobre “base de informação” e se de fato for processada, então estaríamos tratando de uma tradicional “base de dados”.

Os dados em um determinado sistema de informação mesmo que recolhidos e armazenados de uma forma sistemática e estruturada, não podem ser usados diretamente para fins de tomada de decisões (VERCELLIS, 2009). Esses dados precisam ser processados por meio de ferramentas de extração e métodos analíticos adequados, capazes de transformá-los em informação e conhecimento que podem ser posteriormente utilizados pelos tomadores de decisão (VERCELLIS, 2009).

Vercellis, (2009) define dado como uma codificação estruturada de entidades primárias individuais, bem como as operações que envolvem duas ou mais entidades primárias. Enquanto a informação é o resultado das atividades de extração e transformação dos dados, e é significativo para aqueles que o recebem em um domínio específico. E o conhecimento é gerado a partir da transformação da informação em conhecimento quando ele é usado para tomar decisões e desenvolver as ações correspondentes (VERCELLIS, 2009).

O conceito de dado adotado para esta pesquisa será o definido por Davenport (1998) onde os dados podem ser facilmente estruturados, facilmente obtido por máquinas, facilmente transferível e frequentemente quantificado. O conceito de informação adotado será o descrito por Vercellis (2009) como atividades de extração e transformação de dados. E o conceito de conhecimento será o descrito por Davenport (1998) como dados dotados de relevância e propósito a partir de análises elaboradas por pessoas ou computadores.

2.1.2 Recuperação da Informação

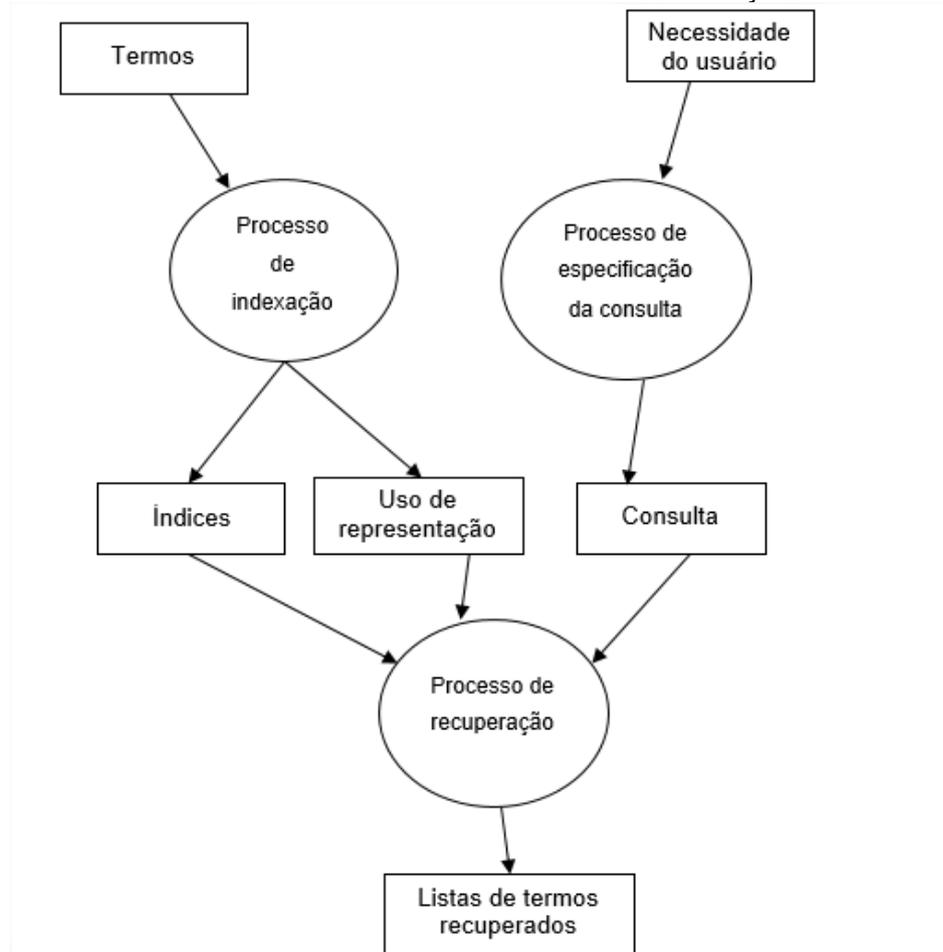
Para Cardoso (2000) a recuperação de informação é uma subárea da ciência da computação que estuda o armazenamento e recuperação automática de documentos, que são objetos de dados, geralmente textos. A recuperação da informação também por ser entendida como uma área que:

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

Ainda nesse sentido de facilidade de acesso, Ferneda (2006) argumenta que a recuperação de informação no ambiente da Internet é relativamente facilitada pelos mecanismos de busca (*search engines*), que coletam e indexam uma parte da imensa quantidade de páginas disponíveis na Web. E para facilitar a seleção dos itens recuperados, a maioria dos mecanismos de busca realiza um ordenamento dos resultados, utilizando algum algoritmo que tenta prever a relevância de cada item para a necessidade de informação do usuário (FERNEDA, 2006).

A necessidade de recuperação de informações armazenadas em grandes repositórios de informação disponíveis na Internet e de responder com maior precisão às buscas realizadas diretamente pelos usuários finais, tem levado a um esforço no sentido de criar sistemas que auxiliem nessa recuperação (BRÄSCHER, 2002). Um processo de recuperação de informação pode ser estruturado conforme a FIGURA 3.

FIGURA 3 – COMPONENTES DE UM PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO.



FONTE: Adaptado de Cardoso (2000).

Esse processo de recuperação de informação basicamente consiste em uma entrada de termos indexada e consulta formulada, e finalmente o processo de recuperação que, a partir das estruturas de dados e da consulta formulada, recupera uma lista de termos ou índices considerados relevantes (CARDOSO, 2000).

2.2 DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS (KDD)

A Descoberta de Conhecimento em Base de Dados, ou no inglês *Knowledge Discovery in Databases* (KDD) é um método tradicional de transformar dados em conhecimento e depende de análises manuais e interpretativas para tentar resolver o problema da era da informação conhecida como a sobrecarga de dados (FAYYAD, PIATETSKY-SHAPIRO; SMYTH, 1996). Em vários campos (ciência, marketing, finanças, saúde, varejo), a abordagem clássica para análise de dados baseia-se, fundamentalmente, em um ou mais analistas tornando-se intimamente familiarizado com os dados e servindo como uma interface entre os dados e os usuários e produtos (FAYYAD, PIATETSKY-SHAPIRO; SMYTH, 1996).

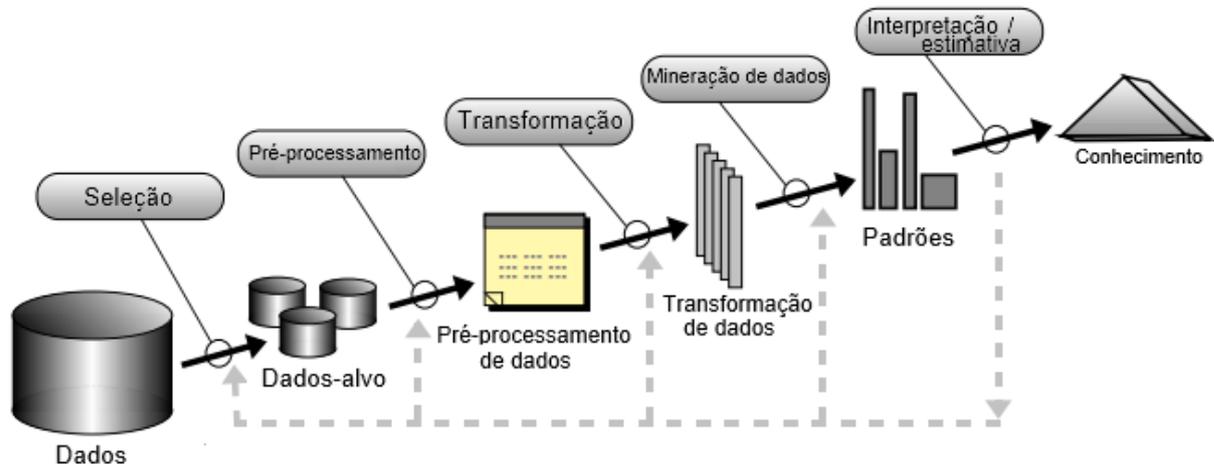
O modelo elaborado por Fayyad, Piatetsky-Shapiro e Smyth (1996) apresenta cinco etapas essenciais para garantir o conhecimento através dos dados, sendo elas:

- seleção: buscar os dados (banco de dados) necessários para a análise;
- pré-processamento e limpeza: Operações básicas que incluem a remoção de ruído, a coleta da informação necessária para modelar o ruído, decidindo sobre as estratégias para lidar com falta campos de dados, e representar informações de sequência de tempo e das alterações conhecidas.
- transformação: encontrar recursos úteis para representar os dados dependendo do objetivo da tarefa. Com métodos de redução de tamanho ou de transformação, o número efetivo de variáveis em consideração pode ser reduzido, ou invariáveis representações para os dados podem ser encontradas;
- mineração de dados: aplicação de métodos de mineração de dados (Algoritmos) para enumerar os padrões; e

- avaliação consideradas: avaliar os produtos de extração de dados para identificar o subconjunto de padrões enumerados considerado conhecimento.

Estas etapas também estão apresentadas na FIGURA 4.

FIGURA 4 – UMA VISÃO GERAL DAS ETAPAS QUE COMPÕEM O PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS (KDD).



FONTE: Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 41)

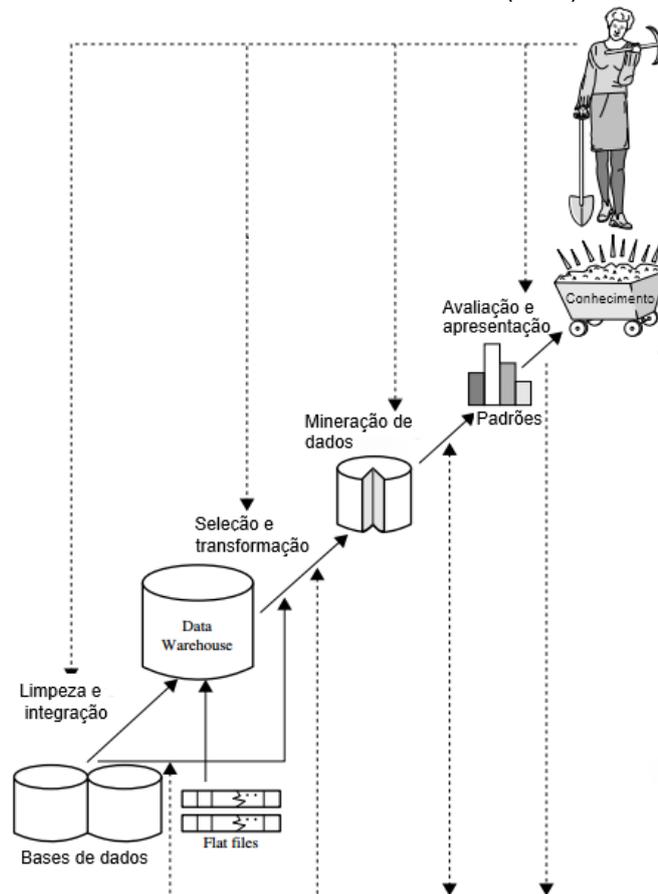
Porém, para Han, Kamber e Pei (2011) o KDD consiste em uma sequência iterativa de etapas, a considerar:

- limpeza de dados: consiste em remover o ruído e dados inconsistentes;
- integração de dados: etapa onde é combinada múltiplas fontes de dados. Os autores consideram a limpeza de dados e integração de dados como uma etapa de pré-processamento, onde os dados resultantes são armazenados em uma *data warehouse*;
- seleção de dados: é nesta etapa que os dados relevantes para a tarefa de análise são recuperados do banco de dados;
- transformação de dados: onde os dados são transformados e consolidadas em formas apropriadas para a mineração através da realização de índices ou operações de agregação. Para os autores essa etapa é realizada antes do processo de seleção de dados, em particular no caso de armazenamento de dados. A redução de dados pode também ser realizado para obter uma representação mais pequena dos dados originais, sem sacrificar a sua integridade;

- mineração de dados: um processo essencial onde os métodos inteligentes são aplicados para extrair padrões de dados;
- avaliação de padrões: para identificar os padrões verdadeiramente interessantes que representam conhecimento baseado em medidas mais interessantes; e
- apresentação do conhecimento: onde técnicas de visualização e representação do conhecimento são utilizados para apresentar o conhecimento minerado aos usuários.

Estas etapas estão apresentadas conforme a FIGURA 5.

FIGURA 5 – A MINERAÇÃO DE DADOS EXEMPLIFICADA EM ETAPA BASEADAS NO PROCESSO DE DESCOBERTA DE CONHECIMENTO (KDD).



FONTE: Han, Kamber e Pei (2011, p. 7).

É possível perceber semelhanças nos dois processos apresentados, ambos iniciam com dados, porém Han, Kamber e Pei (2011) colocam base de dados como primeiro elemento, e ao final, ambos os processos geram conhecimento.

Em ambos os modelos apresentados dissertam sobre a etapa da Mineração de Dados, da qual será melhor abordada na próxima seção.

2.2.1 Mineração de dados

Após a expansão do uso da internet como meio de lazer e trabalho, a web se transformou em um local de armazenamento de grande quantidade de informações e dados. Para poder absolver essa quantidade de dados foi desenvolvido técnicas de mineração de dados (*data mining*) porque é inviável para o homem analisar essa quantidade de dados sem o auxílio de ferramentas computacionais apropriadas (GOLDSCHMIDT; PASSOS, 2005).

Historicamente, a noção de encontrar padrões úteis em dados tem sido dada uma variedade de nomes, incluindo a mineração de dados, extração de conhecimento, descoberta de informações, a colheita de informações, arqueologia de dados e processamento padrão de dados (FAYYAD, PIATETSKY-SHAPIRO; SMYTH, 1996). A mineração de dados termo tem sido usado principalmente por estatísticos, analistas de dados e as comunidades de sistemas de informação de gestão, também popularizando no campo de banco de dados (FAYYAD, PIATETSKY-SHAPIRO; SMYTH, 1996).

Existem várias definições para o conceito de mineração de dados, Liu (2006) argumenta que a mineração de dados pode ser definida como um campo multidisciplinar que envolve a aprendizagem de máquina, estatísticas, bases de dados, inteligência artificial, recuperação de informação e visualização. Há muitas tarefas relacionadas à mineração de dados, como a aprendizagem supervisionada (ou classificação), a aprendizagem não supervisionada (ou agrupamento), a mineração de regras de associação e de mineração padrão sequencial (LIU, 2006). Entende-se por aprendizagem de máquina os algoritmos que aprendem padrões presentes nos dados fornecidos como entrada, esses padrões (complexos ou não) são usados para fazer previsões relativas a dados ainda não vistos e novos (BAEZA-YATES; RIBEIRO-NETO, 2013).

Conforme apresentado por Bing (2006) é possível realizar a mineração de dados a partir de três etapas principais, sendo elas:

- a) pré-processamento: os dados brutos geralmente não são adequados para a mineração, pois precisam ser limpos a fim de remover ruídos ou

anormalidades. Os dados também podem ser demasiadamente grandes e/ou envolver muitos atributos irrelevantes, que exigem redução de dados por meio de amostragem e atribuir seleção;

- b) a exploração dos dados: os dados processados são alimentados por meio de um algoritmo de mineração de dados que irá produzir padrões ou conhecimento;
- c) pós-processamento: em muitas aplicações, nem todos os padrões descobertos são úteis. Essa etapa identifica aqueles dados úteis para aplicações. Várias técnicas de avaliação e visualização são usadas para tomar a decisão.

Outra definição de mineração de dados é apresentada por Han, Kamber e Pei (2011) que comentam que a mineração de dados pode ser entendida como um passo essencial no processo de descoberta de conhecimento, pois revela padrões escondidos para avaliação, ou também como todo o processo de descoberta de conhecimento. Diante disso, os autores consentiram que a mineração de dados é o processo de descobrir padrões e conhecimentos interessantes em grandes quantidades de dados. Podendo considerar como fontes de dados os bancos de dados, *data warehouse*, a *Web* e outros repositórios de informação/dados que são transmitidos para o sistema de forma dinâmica.

Uma das técnicas utilizadas para analisar dados e informações em uma base de dados são as análises de sentimento ou mineração de opiniões que estarão apresentadas na próxima seção.

2.2.2 Tarefas e métodos de mineração de dados

Na mineração de dados, são definidas as tarefas algoritmos de acordo com os objetivos do estudo. As tarefas possíveis de um algoritmo de extração de padrões podem ser agrupadas em atividades preditivas e descritivas.

As atividades preditivas possuem técnicas de classificação, como os algoritmos de ID3, C4.5 e *Naive Bayes* e técnicas de regressão de dados (GALVÃO; MARIN, 2009). Entende-se por técnicas de classificação a aprendizagem de uma função que mapeia um item de dados em uma das várias classes predefinidas e por regressão a

aprendizagem de uma função que mapeia um item de dados a uma variável de previsão de valor real (FAYYAD, PIATETSKY-SHAPIRO; SMYTH, 1996).

As atividades descritivas englobam técnicas de sumarização, como os algoritmos genéticos, associação como os algoritmos *Apriori* e GSP (algoritmo de geração sequencial), agrupamento ou *clusters* (*K-means*) (GALVÃO; MARIN, 2009). A associação é um exemplo onde o uso de restrições e medidas mais interessantes pode garantir a integridade da mineração (HAN, KAMBER; PEI, 2011). O *Apriori* é um dos métodos mais conhecidos e emprega busca de profundidade e gera conjuntos de itens candidatos de n elementos a partir de um conjunto de itens (CASTRO; FERRARI, 2016). Uma base de dados com m itens pode gerar até $2^m - 1$ itens frequentes, excluindo conjuntos vazios. Uma forma de reduzir o número de conjuntos de itens candidatos é uma forma de reduzir a complexidade computacional. O agrupamento ou *clustering* é uma tarefa descritiva comum, onde se procura identificar um conjunto finito de categorias ou grupos para descrever os dados (FAYYAD, PIATETSKY-SHAPIRO; SMYTH, 1996).

Além disso, a mineração de dados utiliza tecnologias para poder identificar padrões em dados, as principais, de acordo com Galvão e Marin (2009) são: redes neurais, árvore de decisão, algoritmos genéticos (AGs), lógica nebulosa (*fuzzy logic*) e estatística.

Entende-se por redes neurais os algoritmos capazes de aprender a partir de repetidos dados que podem representar padrões semelhante são funcionamento das redes neurais humanas (HAYKIN, 2001).

As árvores de decisão são uma série de perguntas sistematicamente dispostas de modo que cada pergunta consulta um atributo e galhos com base no valor do atributo, nas folhas da árvore são colocadas previsões da variável de classe (WU; KUMAR, 2009). Um dos algoritmos de árvore de decisão é o J48. De acordo com Brilhadori e Lauretto (2013)

o J48 é um indutor *top-down* de árvores de classificação que no Weka atua junto com o algoritmo C4.5, a seleção da melhor partição dos nós e o critério de parada são baseados na entropia de Shannon, como é usual em parte da família de indução de árvores de classificação. O J48 também possui uma fase de pós-poda da árvore após a expansão, na qual são convertidas para folhas as sub-árvores que não representam ganhos de informação acima de um limiar especificado. O algoritmo é capaz de lidar com classes binárias, nominais e valores faltantes de classe. Suporta atributos binários, de data, nominais, numéricos e valores faltantes (BRILHADORI; LAURETTO, 2013, p. 98).

Os Algoritmos Genéticos (AGs) são um método de busca, muitas vezes aplicados à otimização e aprendizado em uma tentativa de simular computacionalmente os processos pelos quais a seleção natural opera, e aplicá-los para resolver problemas de negócios e de pesquisa (LAROSE, 2006).

A lógica nebulosa, ou *fuzzy logic* é utilizada para representar a extensão da lógica clássica para uma mais flexível com objetivo de formalizar conceitos imprecisos e a outra é onde se aplicam conjuntos *fuzzy* à diversas teorias e tecnologias para processar informações imprecisas como por exemplo, em processos de tomada de decisão (GOLDSCHMIDT; PASSOS, 2005).

Por fim, a estatística fornece modelos para análise e interpretação de dados (GALVÃO; MARIN, 2009).

2.2.3 Mineração de opinião: definição

Estudos relacionado com à área de Mineração de Opinião, ou também chamada de Análise de Sentimento, tiveram seu marco inicial a partir de 2001. Ela surgiu como uma nova área do conhecimento a partir de reunião de Recuperação de Informação (*Information Retrieval*) e Linguística Computacional no ambiente *on-line* (PANG, LEE; VAITHYANATHAN, 2002).

Seu foco de estudo se baseia em estudar a opinião expressa de determinado documento ou acontecimento (ESULI; SEBASTIANI, 2006). Sua utilização permite rastrear, por exemplo, opiniões de usuários sobre produtos ou sobre candidatos políticos expressas em fóruns *on-line* a partir de suas aplicações e métodos (LOPES; HIRATANI, 2008). Além disso, permite aplicar seus resultados para fins de crítica de filmes e produtos baseados em opiniões de usuários. E, pode ser utilizada em aplicações de *Business Intelligence* (LOPES; HIRATANI, 2008).

A Mineração de Opinião e Análise de Sentimento, resumidamente, analisa as opiniões, sentimentos, avaliações, apreciação descritas por pessoas referente a problemas, eventos, tópicos e seus atributos (PANG; LEE, 2008).

Com relação a formas de aplicações da Mineração de Opinião Pang e Lee (2008) descrevem que a tarefa fundamental utilizada é a classificação ou regressão. Para os autores, é por meio da classificação ou regressão que é possível determinar unidades textuais, incluir tomar uma decisão para uma frase particular ou documento,

ordenar um conjunto de textos, dando um único rótulo para uma coleção de documentos, e categorizar a relação entre duas entidades baseadas em evidência textual.

Liu (2012) aborda que a classificação é essencialmente um problema de classificação de texto onde é geralmente formulado como um problema de classificação de duas classes, sendo elas positivas ou negativas. Dentre as palavras positivas e negativas o autor descreveu que os indicadores mais importantes de sentimentos são “palavras sentimento”, também chamados de “palavras de opinião” que servem para expressar sentimentos. Por exemplo, “bom”, “maravilhoso” e “surpreendente” são palavras sentimento positivo, e “mau”, “pobre”, e “terrível” são palavras sentimento negativo (LIU, 2012).

Para aplicar técnicas de mineração de opinião são comumente utilizados dois tipos de algoritmos: os de aprendizagem supervisionada e os de aprendizagem não-supervisionada. Os algoritmos de aprendizagem supervisionada são aqueles que objetivam construir um modelo de distribuição de classes (categorias) em função das características dos dados em questão, como exemplo: *Naïve Bayesian*, e *Support Vector Machines* (SVM) (FERREIRA, 2010). Já os algoritmos de aprendizagem não-supervisionadas, as classes dos dados não são conhecidas, por exemplo os algoritmos de *clustering* (agrupamento) (FERREIRA, 2010).

2.2.4 Banco de dados

Para poder realizar a mineração de dados é necessário possuir os dados armazenados e estruturados em um banco de dados. Um banco de dados pode ser entendido como coleções de dados que contém informações importantes de acordo com a coleta, esta pode ser de uma empresa ou de determinado evento (SILBERSCHATZ, KORTH; SUDARSHAN, 2006). Pode ser definido como um conjunto de “dados” devidamente relacionados (SILBERSCHATZ, KORTH; SUDARSHAN, 2006)

De acordo com Silberschatz, Korth e Sudarshan (2006) o gerenciamento de dados envolve definir estruturas para armazenamento de informação e fornece mecanismos para a manipulação de informações. Para o autor um sistema de banco de dados precisa garantir: a segurança das informações armazenadas; evitar

resultados anômalos em consultas; facilitar o acesso aos dados; e garantir a integridade dos dados (atender às restrições de armazenamento).

Os dados recolhidos por grandes organizações são armazenados diariamente em bases de dados. Mas os administradores de banco de dados podem não estar dispostos a permitir que os mineradores de dados tenham acesso direto a essas fontes de dados, e podem apenas permitir acesso aos bancos de dados operacionais (usado para operações de rotina). Porém, mineradores de dados usam uma grande quantidade de dados o que poderiam interferir com as operações de negócios comuns. Bases de dados operacionais não estão organizadas para mineração de dados, isso pode dificultar a obtenção dos dados necessários e demandar de muito tempo (BROWN, 2014).

2.3 REDE SOCIAL

A rede (*network*) é um sistema de nodos e elos, uma estrutura sem fronteiras, uma comunidade não geográfica (MARTELETO, 2001). A rede social deriva deste conceito, passa a representar um conjunto de participantes autônomos, unindo ideias e recursos em torno de atores e interesses compartilhados (MARTELETO, 2001). Com relação ao ambiente *web* as redes sociais são serviços baseados na internet que permitem aos indivíduos: construir um perfil público ou semi-público dentro de um sistema limitado; articular uma lista de outros usuários com quem eles compartilham uma conexão; e visualizar e percorrer sua lista de conexões e aqueles feitos por outros dentro do sistema (ELLISON, 2007).

Os elementos fundamentais de uma rede social são: os atores, e os laços de relacionamento (NOGUEIRA, 2015). Atores são entidades sociais (coletivas, corporativo ou individual) com um conjunto de características ou atributos com comportamentos e objetivos em comum, podem ser do mesmo tipo (formando um nó-único) ou de tipos diferentes (formando uma rede composta) (NOGUEIRA, 2015). Laços de relacionamento ligam os atores uns aos outros, eles podem ser laços de amizades, profissionais ou familiares representando um conjunto de ligações de um tipo específico entre membros de um grupo (NOGUEIRA, 2015). Depois de aderir a um site de rede social, os usuários são solicitados a identificar outros usuários no sistema com quem tem um relacionamento (ELLISON, 2007). O rótulo para estes relacionamentos varia de acordo com os termos de sites populares incluem "Amigos",

"Contatos" e "Fãs" (ELLISON, 2007). O termo "Amigos" pode ser enganosa, porque a conexão não significa necessariamente amizade, no sentido conversar com todos os dias, e as razões para as pessoas se conectam são variados (ELLISON, 2007).

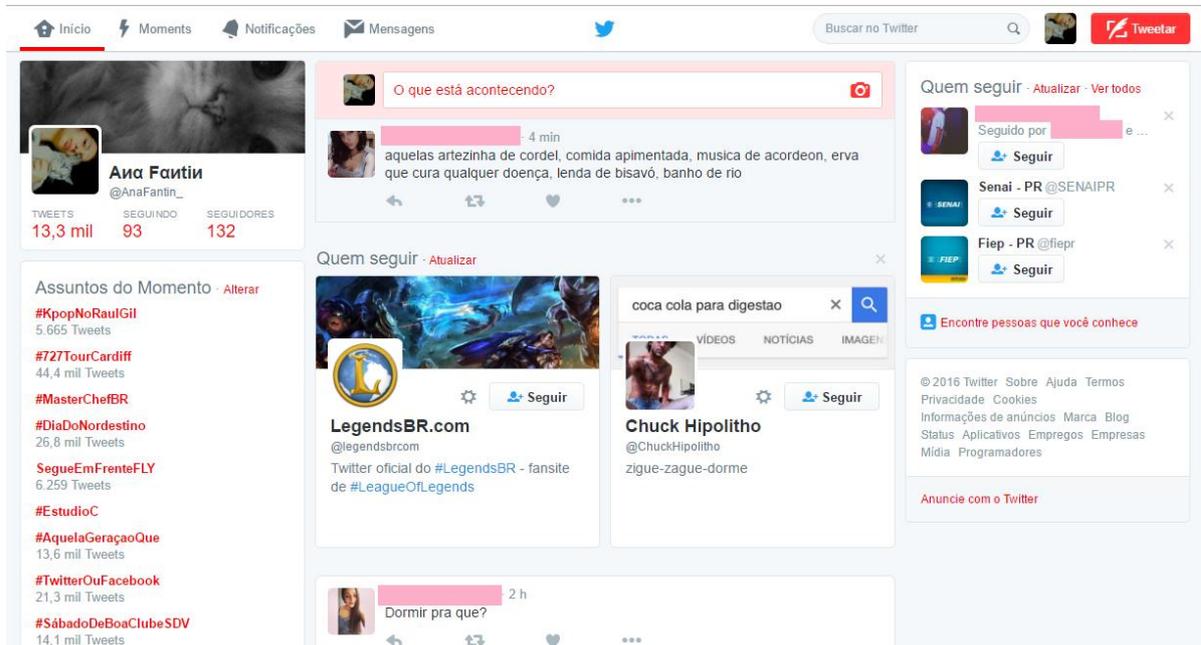
Para estudar como os comportamentos ou as opiniões dos indivíduos dependem das estruturas nas quais eles se inserem, a unidade de análise não são os atributos individuais (classe, sexo, idade, gênero), mas o conjunto de relações que os indivíduos estabelecem através das suas interações uns com os outros. A estrutura é apreendida concretamente como uma rede de relações e de limitações que pesa sobre as escolhas as orientações, os comportamentos, as opiniões dos indivíduos (MARTELETO, 2001, p 72).

Mesmo sendo informal as relações sociais, os efeitos das redes podem ser percebidos fora de seu espaço, nas interações com o Estado, a sociedade ou outras instituições representativas (MARTELETO, 2001).

2.3.1 A rede social Twitter

De acordo com Recuero (2009) o Twitter é um site popularmente denominado de um serviço de *microblogging* porque permite que sejam escritos pequenos textos de até 140 caracteres a partir da pergunta "O que está acontecendo? ", em inglês "*What's happening?*". Na FIGURA 6, por exemplo, é possível observar essa frase e a estrutura do Twitter. A partir dessa frase o usuário podem escrever breves narrativas sobre o que se estava fazendo no momento.

FIGURA 6 – PÁGINA DO TWITTER ONDE É POSSÍVEL OBSERVAR A PERGUNTA "O QUE ESTÁ ACONTECENDO?"



FONTE: TWITTER (2016).

Em virtude da rapidez com que se publicam *tweets* na rede, com frequência a rede Twitter consegue divulgar notícias com muito mais rapidez que qualquer meio jornalístico tradicional (PRIMO, 2008). Em pouco tempo, a forma como esta rede era utilizada foi ultrapassando sua proposta inicial, como a cobertura de eventos em andamento (congressos, jogos etc.) e propaganda política (PRIMO, 2008).

O Twitter é estruturado a partir de seguidores e pessoas a seguir, cada usuário pode escolher quem deseja seguir e ser seguido por outros. Outras formas de comunicação encontradas nesta ferramenta é a possibilidade de enviar mensagens em modo privado para outros usuários. A janela particular de cada usuário contém todas as mensagens públicas emitidas por aqueles indivíduos a quem ele segue. Para poder direcionar mensagens basta utilizar o símbolo “@” – arroba – antes do nome do destinatário (RECUERO, 2009).

De acordo com Berlim (2015) o Twitter é uma ferramenta de interessante interface que além de ser uma rede de relacionamentos e interatividade, permite a geração de estratégias para a tomada de decisão.

2.3.2 A rede social Facebook

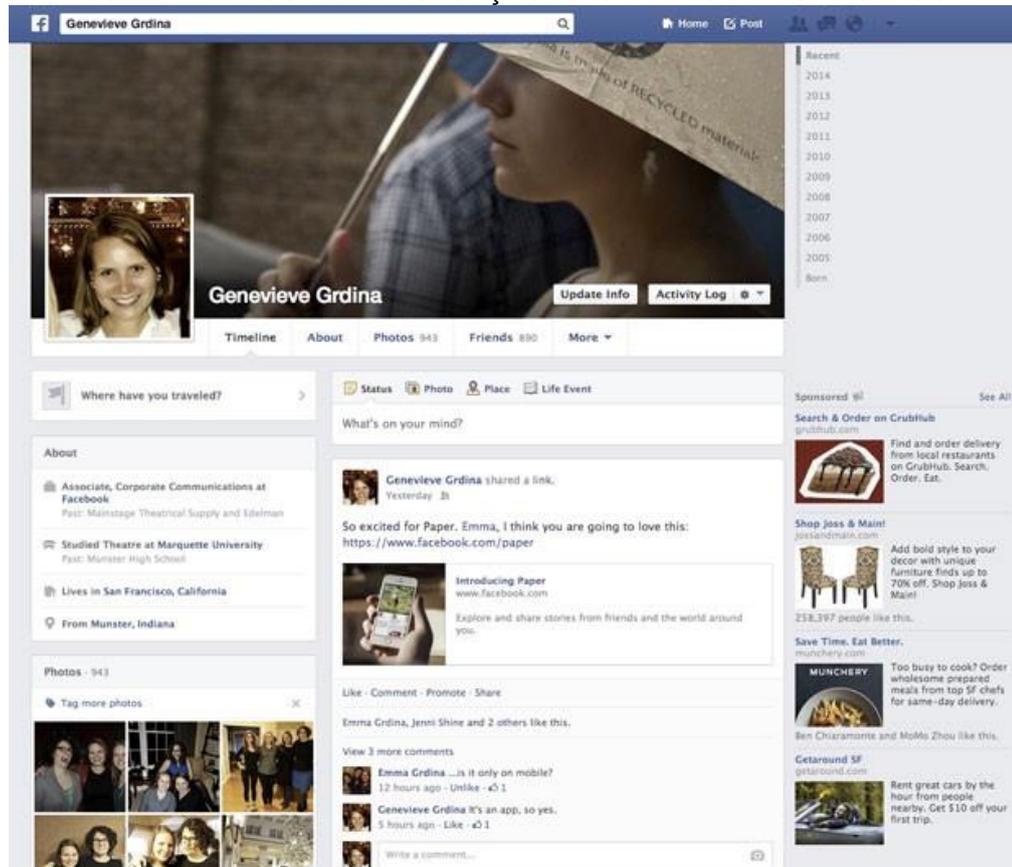
O Facebook se iniciou em 4 de fevereiro de 2004, com a missão de dar às pessoas o poder de compartilhar informações e fazer do mundo um lugar mais aberto e conectado (FACEBOOK, 2016). Atualmente milhões de pessoas usam o Facebook para compartilhar um número ilimitado de fotos, links, vídeos e conhecer mais as pessoas com quem você se relaciona (FACEBOOK, 2016).

Pinto et. al. (2011) descrevem que o *Facebook* é uma das redes sociais de maior destaque na atualidade, apresenta um crescimento expressivo e ativa participação de seus membros. Sua estrutura é dotada de ferramentas variadas que possibilitam a socialização, mas que pressupõe empenho no processo de aprendizagem sobre a utilização de seus recursos, em função de sua aparente complexidade (PINTO, et al., 2011).

O visual do Facebook organiza o conteúdo cronologicamente, como em um "*feed*" de notícias (G1, 2014). As publicações podem conter fotos, vídeos e possuir o maior destaque na página.

O *layout* de divulgação do Facebook apresentado na FIGURA 7 revela a página atual dessa rede social acessada através do computador.

FIGURA 7 – LAYOUT DE DIVULGAÇÃO DE UMA PÁGINA NO FACEBOOK.



FONTE: FACEBOOK (2014).

Além de páginas acessas via computador, o Facebook criou páginas para dispositivos móveis, tais como *smartphones* e *tablets* (G1, 2014). De acordo com o site do G1 (2014) em 2014 essa rede social atingiu 1,19 bilhão de usuários em todo o mundo.

2.4 EXAME NACIONAL DO ENSINO MÉDIO - ENEM

De acordo com o Ministério da Educação e Cultura - MEC (2016) o Exame Nacional do Ensino Médio, mais conhecido como ENEM foi criado em 1998 com o objetivo de avaliar o desempenho do estudante ao fim da escolaridade básica. Os estudantes que estão concluindo ou que já concluíram o ensino médio em anos anteriores podem participar deste exame.

A partir de 2009 passou a ser utilizado também como mecanismo de seleção para o ingresso no ensino superior. Foram implementadas mudanças no Exame que

contribuem para a democratização das oportunidades de acesso às vagas oferecidas por Instituições Federais de Ensino Superior (IFES), para a mobilidade acadêmica e para induzir a reestruturação dos currículos do ensino médio, conforme consta no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Sobre o Enem - INEP (2016).

2.4.1 Formas de utilização do ENEM

O ENEM, de acordo com o INEP (2016), é utilizado como critério de seleção para os estudantes que pretendem concorrer a uma bolsa no Programa Universidade para Todos (ProUni). Além disso, cerca de 500 universidades que utilizam o resultado do exame como critério de seleção para o ingresso no ensino superior, seja complementando ou substituindo o vestibular.

O ENEM tem como ênfase a avaliação do perfil dos egressos deste nível de ensino. Seu objetivo principal é proporcionar uma avaliação do desempenho dos alunos, ao término da escolaridade básica, segundo uma estrutura de competências associadas aos conteúdos disciplinares, que se espera tenha sido incorporada pelo aluno, para fazer frente aos crescentes desafios da vida moderna (CASTRO; TIEZZI, 2005, p. 17).

Além do ProUni, a nota do ENEM é usada para selecionar estudantes para vagas públicas e privadas de ensino superior e ensino técnico pelo Sistema de Seleção Unificada (Sisu) e pelo Sistema de Seleção da Educação Profissional e Tecnológica (Sisutec), respectivamente (PORTAL BRASIL, 2015). Além de ser exigência para financiamento pelo Fundo de Financiamento Estudantil (FIES). Para maiores de 18 anos, o ENEM pode ser usado ainda como certificação do ensino médio (PORTAL BRASIL, 2015).

Para poder tirar a certificação do ensino médio, o candidato terá que indicar a pretensão de utilizar os resultados de desempenho no exame para fins de certificação de conclusão do Ensino Médio, no ato da inscrição, bem como a Instituição Certificadora; possuir no mínimo 18 (dezoito) anos completos na data da primeira prova de cada edição do exame; atingir o mínimo de 450 (quatrocentos e cinquenta) pontos em cada uma das áreas de conhecimento do exame; e atingir o mínimo de 500 (quinhentos) pontos na redação (BRASIL, 2011).

2.4.2 Inscrição

Para fazer a inscrição no ENEM, o aluno pode se inscrever pelo ambiente online através do *site*: <http://enem.inep.gov.br/> no período permitido seguindo os oito passos indicados no *site*. Para 2016, a taxa de inscrição foi estabelecida em R\$ 68,00 (G1, 2016), superior em R\$ 50,00 se comparada ao ano anterior. Contudo, alguns estudantes, de baixa renda e/ou que recebem bolsa família são isentos do pagamento e outros ainda podem pedir isenção, alunos que estão terminando o ensino médio em escolas públicas também estão isentos (G1, 2016).

2.4.3 A prova do ENEM

O ENEM avalia cinco áreas do conhecimento, a ser consideradas: Ciências da Natureza e suas Tecnologias; Ciências Humanas e suas Tecnologias; Linguagens, Códigos e suas Tecnologias; Matemática e suas Tecnologias; e Redação, conforme apresentado pelo INEP (2011).

A prova do ENEM é composta de 180 questões objetivas e uma redação. Ela é realizada uma vez ao ano e por ser uma prova extensa é dividida em dois dias. O primeiro dia possui 45 questões de ciências da natureza e mais 45 questões de ciências humanas possuindo 4 horas e meia para sua realização. O segundo dia de prova é realizada 45 questões de linguagens e códigos, mais 45 questões de matemática por fim, uma redação com duração de 5 horas e meia.

Concluída a exposição dos conceitos que contemplam o tema estudado, a próxima seção apresentará os procedimentos metodológicos.

3 CARACTERIZAÇÃO E PROCEDIMENTOS METODOLÓGICOS

Neste capítulo estão apresentadas a caracterização da pesquisa, e os procedimentos metodológicos realizados para alcançar os objetivos propostos como a etapa de coleta, análise e concepção das bases de dados.

3.1 CARACTERIZAÇÃO DA PESQUISA

Esta pesquisa pode ser classificada como exploratória. Conforme apresenta por Malhotra (2006), uma pesquisa exploratória tem por objetivo explorar ou fazer uma busca em um problema ou situação para prover critérios e maior compreensão. Ela pode ser utilizada para desenvolver hipóteses ou para estabelecer prioridades para pesquisas posteriores.

Como abordagem, pode ser considerada como qualitativa e quantitativa porque a pesquisa se caracteriza como um esforço cuidadoso para a descoberta de novas informações para a verificação e ampliação do conhecimento existente (GODOY, 1995). No que diz respeito da abordagem qualitativa ela se encontra em variados tipos de investigação podendo ser realizada através da leitura de livros, obras de referência e artigos de pesquisa na área que está sendo realizada desde o início deste estudo (GODOY, 1995). E com relação à abordagem quantitativa, Godoy (1995) explica que este tipo de estudo se preocupa com a medição objetiva e a quantificação dos resultados, buscando a precisão, evitando distorções na etapa de análise e interpretação dos dados, garantindo assim uma margem de segurança em relação às inferências obtidas.

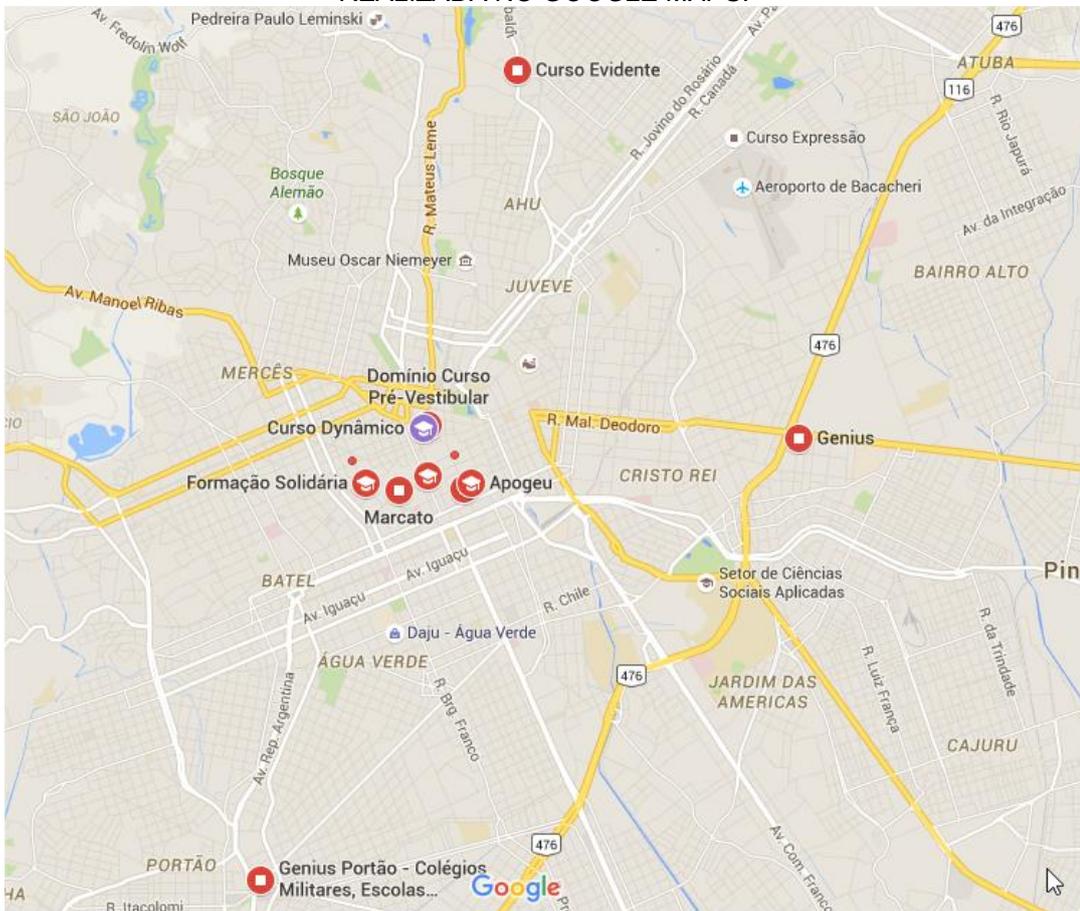
3.2 POPULAÇÃO E AMOSTRA

A população dos inscritos no ENEM 2015 no Brasil chegou a 7.746.057 (sete milhões e setecentos e quarenta e seis mil e cinquenta e sete) com base na divulgação do site da INEP (2015). O estado do Paraná teve 336.307 (trezentos e trinta e sei mil e trezentos e sete) inscritos.

A região escolhida para realizar a coleta de dados é a região central de Curitiba e seus arredores na qual detém de maior concentração de cursos de

preparação para o ENEM, de acordo com o resultado da pesquisa realizada na ferramenta *Google Maps* da Google em 11 de junho de 2016 foram retornados 8 (oito) locais na região central e 2 (dois) nas proximidades totalizando 10(dez) resultados. O termo utilizado para a pesquisa foi “cursos pré-vestibulares”. A amostra para esta pesquisa é considerada uma amostra por conveniência e será composta pelos 10 (dez) cursos preparatórios. Alguns desses cursos podem ser visualizados na FIGURA 8.

FIGURA 8 – RESULTADOS DA PESQUISA POR CURSOS PREPARATÓRIOS EM CURITIBA REALIZADA NO GOOGLE MAPS.



FONTE: GOOGLE MAPS (2016).

O questionário também será aplicado aos alunos do setor de Ciências Sociais Aplicadas (SCSA) da Universidade Federal do Paraná (UFPR) no Campus Jardim Botânico. Esse setor é composto por quatro cursos, a considerar: Administração, Ciências Contábeis, Ciências Económicas e Gestão da Informação. A intenção do questionário é conseguir atingir pelo menos 10% (dez por cento) dos alunos de cada curso para ter um número de respostas diversificado para realizar a mineração dos

dados. A TABELA 3 apresenta a quantidade de alunos matriculados no segundo semestre de 2016 por cursos.

TABELA 3 – QUANTIDADE DE ALUNOS NOS CURSOS DE ADMINISTRAÇÃO, CIÊNCIAS CONTÁBEIS, CIÊNCIAS ECONÔMICAS E GESTÃO DA INFORMAÇÃO MATRICULADOS NO SEGUNDO SEMESTRE DE 2016.

Curso	Alunos- matutino	Alunos - noturno	Total	Amostra
Administração	289	551	840	84
Ciências Contábeis	*	494	494	49
Ciências Económicas	318	481	799	80
Gestão da Informação	225	*	225	23
Total			1864	186

FONTE: Coordenação dos cursos de Administração, Ciências Contábeis, Ciências Econômicas e Gestão da Informação (set. 2016).

NOTA* o curso não possui alunos matriculados neste turno

Além da coleta dos dados pela aplicação do questionário também serão coletados por meio das redes sociais: *Facebook* e *Twitter*. No *Facebook*, os perfis considerados para a coleta foram selecionados com base em dois critérios: se as publicações são atuais, considerando o período entre outubro de 2015 e janeiro de 2016; e se houve comentários pelo menos nas cinco últimas publicações, com base nisso, a TABELA 4 apresenta as páginas consideradas para a coleta de dados e as rejeitadas. Esse período foi considerado porque está entre a inscrição e realização da prova do ENEM e a divulgação do resultado da prova. No *Twitter* as sentenças serão filtradas a partir da utilização da sua ferramenta de busca para coletar as opiniões dos usuários referentes ao tema ENEM.

TABELA 4 – PERFIS SELECIONADOS E NÃO SELECIONADOS DO FACEBOOK PARA A COLETA DE OPINIÕES.

Perfil	Endereço web	Observação
ENEM, Sisu e ProUni	https://www.facebook.com/ENEMsisuprouni/?ref=br_rs	Aceita
ENEM com você	https://www.facebook.com/ENEMcomvoce/?ref=br_rs	Aceita
ENEM	https://www.facebook.com/ENEMdicas/	Aceita
ENEM	https://www.facebook.com/onlinemundoENEM/?ref=br_rs	Rejeitada, é uma página de divulgação de aulas <i>online</i> .
Rede ENEM	https://www.facebook.com/RedeENEM/?ref=br_rs	Rejeitada, não possui comentários nas 5 últimas publicações.
ENEM da Depressão	https://www.facebook.com/ENEMDepressao/?ref=br_rs	Aceita
Total de páginas aceitas:		4
Total de páginas rejeitadas:		2
Total de páginas analisadas		6

FONTE: Elaborado pela autora (2016).

A criação das bases de dados passará por dois processos, primeiro, será concebida a base de dados a partir da aplicação do questionário e estruturação das opiniões coletadas dos estudantes dos cursos preparatórios. O segundo processo é à coleta de opiniões nas redes sociais, os dados dessa coleta serão inseridos em uma nova base de dados a elaboração da base de dados. Esta base terá dados coletados da rede social *Facebook* e *Twitter* da qual ocorrerão manualmente.

Após a aplicação do questionário, os dados serão estruturados conforme as operações utilizadas no processo de Descoberta de Conhecimento de Base de Dados, que consiste em: pré-processamento, mineração de dados e pós-processamento.

Para conceber a base com as opiniões coletadas nas redes sociais, serão seguidas as etapas de mineração de opinião, apresentadas por Silva (2010) e pela Berlim (2015) a considerar: subjetividade (seção 3.3.1), extração (seção 3.3.2), classificação (seção 3.3.3) e sumarização (seção 3.3.5).

Ambos os processos passarão pelas etapas de coleta de dados (seção 3.1.1) e armazenamento (seção 3.2.1).

3.3 ETAPAS DE COLETA E ARMAZENAMENTO DOS DADOS

A coleta de dados consiste em coletar os dados e selecionar os dados das fontes escolhidas para então poder armazená-los em uma base de dados.

Nesta etapa os dados serão coletados de três fontes: a primeira é por meio da aplicação de um questionário para os alunos de cursos preparatórios, a segunda é por meio da aplicação de um questionário para os alunos dos cursos do Setor de Ciências Sociais Aplicadas da UFPR e a terceira é por meio da coleta de comentários das publicações das páginas escolhidas. Para armazenar os dados, eles serão limpos, integrados por fim, formatados.

3.3.1 Coleta de dados por meio do questionário

Para criar o questionário foi realizada uma entrevista com dois professores de um curso preparatório a fim de elaborar as perguntas mais adequadas para a coleta de opiniões dos alunos.

A partir da entrevista realizada, foi possível verificar o interesse dos professores em saber a opinião dos alunos com relação ao conteúdo cobrado da prova com relação ao que o aluno viu em sala de aula, tempo de realização, sistema de inscrição, a opinião do aluno com relação a certificação do ensino médio se o candidato atinge a nota mínima, considera o valor cobrado para realizar a prova e se o aluno vê o ENEM como mecanismo de avaliação do ensino médio brasileiro.

Com base nas sugestões indicadas pelos professores foi possível criar a primeira versão do questionário (disponível no APÊNDICE A – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM). Após essa criação foi feito um pré-teste com pessoas que realizam o ENEM para verificar possíveis alterações e melhorias no questionário. O pré-teste foi concebido na ferramenta *Formulário Google* e pode ser visualizado no APÊNDICE A – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO PRÉ-TESTE.

O pré-teste foi realizado com oito pessoas que fizeram o ENEM. Dessas oito, apenas cinco propuseram sugestões. As sugestões aceitas consistiram em:

- alterar a palavra “sexo” para “gênero”;

- justificar o “não” da questão “Você considera justo que com apenas 450 pontos a certificação do ensino médio seja possível?”;
- alterar as questões: “Em que faixa etária você está?” para “Quantos anos você tinha na última vez que fez o ENEM?”;
- para padronizar os campos de respostas e deixa-las minúsculo, havia campos com opções Sim e Não e não e sim, foi optado por deixar sim e não;
- esclarecer melhor que a questão “Em que tipo de escola você estuda ou estudou?” remete ao último ano do ensino médio;
- modificar a questão “Quantos anos você tinha quando fez o ENEM?” indicando que é referente a idade da última vez em que o participante fez o ENEM;
- melhorar a questão “Você concorda que a certificação do ensino médio seja possível com 450 pontos?” na qual passou a ficar “Você concorda que 450 pontos sejam suficientes para a obtenção da certificação do ensino médio?”;
- melhorar a questão “Você acredita que o preço cobrado para realizar a prova é justo considerando que é uma prova que avalia o ensino médio brasileiro.” para “Você acredita que o preço cobrado para realizar a prova é justo?”.

As sugestões não aceitas foram:

- retirar os números das questões;
- deixar um campo aberto para o respondente; “Em que tipo de colégio que você estuda ou estudou?”.

Essas sugestões não foram aceitas pelos seguintes motivos: as questões ficam mais organizadas enumerando-as e deixar aberta a questão pode proporcionar ao respondente diversas formas de resposta como “estudei em escola pública e escola particular”, o foco dessa questão é saber qual tipo de colégio o respondente estudou pela última vez.

A versão final do questionário foi realizada no site Online Pesquisa que permitiu criar uma conta estudante gratuita utilizando e-mail da @ufpr.br. O questionário finalizado pode ser encontrado no APÊNDICE B – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO APLICADA AOS ALUNOS DOS CURSOS PREPARATÓRIOS. A versão final foi aplicada nos grupos

dos cursos preparatórios no *Facebook* e ficou disponível do dia 30/08/2016 ao dia 29/09/2016.

O questionário aplicado aos alunos da Universidade Federal do Paraná – Setor de Ciências Sociais Aplicadas se diferenciou pelas seguintes questões:

- Qual o seu curso?
- Você está atualmente matriculado na UFPR?
- Qual edição do ENEM você participou?

Essas questões foram realizadas para definir os cursos dos respondentes e também para permitir que apenas os alunos matriculados atualmente respondessem. Além disso para poder analisar quantos alunos fizeram a última edição do ENEM em 2015. Esse questionário ficou disponível do dia 30/09/2016 ao dia 29/10/2016. A versão desse questionário pode ser visualizada no APÊNDICE C – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO APLICADA AOS ALUNOS DA UFPR.

3.3.2 Coleta de opiniões nas redes sociais

As sentenças serão coletadas do Facebook nos perfis que foram aceitos conforme apresentado na TABELA 4. A coleta será realizada manualmente. As sentenças serão selecionadas das principais publicações de cada página, totalizando 50 (cinquenta) sentenças.

Na opção de busca do Twitter serão indexados os termos de pesquisa “ENEM” e “Exame Nacional do Ensino Médio”. A página retornará os *tweets* dos usuários que possuem perfil público e que utilizaram essas palavras. Serão coletadas 50 (cinquenta) sentenças devido ao número de retorno que a página de busca do Twitter apresenta. Elas serão coletadas manualmente para indexar na base de dados

3.3.3 Armazenamento dos dados

Nesta etapa, as sentenças consideradas apropriadas para minerar opiniões serão armazenadas em uma base de dados por meio da ferramenta Microsoft Excel. As sentenças coletadas tanto pelo questionário quanto pelas redes sociais terão identificadores únicos variando de 001 a *n*.

Cada questão do questionário corresponderá a um atributo na base de dados. Cada resposta adquirida pelo questionário resultará em um metadado. Os metadados da base poderão ser indexados nos formatos: texto e numéricos.

A limpeza dos dados poderá consistir das seguintes formas: se houver dados vazios na base os mesmos serão padronizados e substituídos por “vazio” ou “0” e se houver dados que estejam escritos incorretamente, como por exemplo, a colocação em um caractere na ordem errada: “insatisfatório” quando o correto é “insatisfatório” ou até mesmo a grafia incorreta.

3.4 ETAPAS DE CRIAÇÃO DA BASE DE DADOS COM OS DADOS COLETADOS POR MEIO DO QUESTIONÁRIO

A Descoberta do Conhecimento em Base de Dados compreende três etapas operacionais básicas: pré-processamento, mineração de dados e pós-processamento.

O pré-processamento consiste basicamente em preparar os dados para a etapa de mineração de dados. Para poder abrir a base de dados no Weka, foi necessário primeiro padronizar as respostas das questões abertas, como todas as questões abertas eram para justificar a resposta “não”, foi mantido apenas o “não” dessas respostas. O conteúdo das respostas abertas será analisado na seção de apresentação dos resultados.

Além disso, foi necessário retirar os espaçamentos entre as palavras, por exemplo, escola pública, se tornou escolapublica (sem espaço). Além disso, foi necessário retirar todos os acentos das palavras como: pública, não, aceitável, entre outras.

Outro ponto necessário para a mineração dos dados, foi a desratização da coluna “idade” na qual consistia em dados numéricos. Para obter significância estatística foi necessário dividir a idade em 2 grupos: A e B. O grupo A ficou com o valor de 15 a 17 totalizando 46 registros, e o grupo de B de 18 a 38 totalizando 24 (vinte e quatro) registros.

A mineração de dados é considerada a principal etapa do processo de KDD. Consiste basicamente em aplicar algoritmos sobre os dados em busca do conhecimento implícitos e úteis (GOLDSCHMIDT e PASSOS, 2005). Após a

elaboração da base de dados, serão utilizadas técnicas de mineração de opinião para apresentar os padrões encontrados na base de dados criada.

O pós-processamento dos dados abrange o tratamento do conhecimento obtido na mineração de dados. Tal tratamento tem como objetivo facilitar a interpretação dos dados e a avaliação da utilidade do conhecimento descoberto.

Nesta etapa, o conhecimento adquirido será organizado e para simplificar serão utilizados gráficos como forma de representação do conhecimento adquirido.

3.5 ETAPAS PARA CRIAÇÃO DA BASE DE DADOS COM AS OPINIÕES COLETADAS DAS REDES SOCIAIS

Para poder minerar opiniões serão utilizadas etapas complexas consideradas importantes para efetivar a descoberta de padrões (BERLIM, 2015).

Essas etapas se dividem em: subjetividade, extração, classificação e sumarização, que estão apresentadas nas próximas seções.

3.5.1 Subjetividade

Nessa etapa, será verificado se os dados coletados são objetivos ou subjetivos para então poder analisar a opinião. Por exemplo, na sentença: “*O céu é azul*”, é nítido que não existe uma opinião, então neste caso a frase é objetiva. Neste outro exemplo, “*O céu está lindo hoje*”, pode ser classificada pois contém opinião, então é uma frase subjetiva (BERLIM, 2015).

É necessário identificar a opinião e detectar a subjetividade, separando as sentenças objetivas das subjetivas. As únicas sentenças que terão utilidades para análise de sentimentos são as subjetivas (SILVA, 2010). As subjetivas apresentam sentimentos, emoções ou pensamentos relativos a algo, ou seja, representam as opiniões que serão analisadas (SILVA, 2010).

Berlim (2015) afirma que a fase subjetividade é de extrema importância para se filtrar apenas as sentenças que se configuram como opiniões, ao se utilizar uma sentença objetiva, perde-se o sentido do estudo. Por esse motivo, faz-se necessário separar as sentenças selecionadas neste estudo.

3.5.2 Extração

Esta etapa envolve separar e diferenciar os sentimentos em negativos e positivos. Para Silva (2010) a extração de características é uma das tarefas mais difíceis de ser realizada, e menos propensa a automação da mineração de opinião.

Esta etapa “tem por objetivo identificar características do objeto sob análise que são alvo de comentários como pares válidos a partir de um corpus representativo para o domínio escolhido” (SILVA, 2010, p. 11).

No exemplo “*A comida do Restaurante X é maravilhosa, e a sobremesa servida não é cara. O sushi no Restaurante Y é muito bom, mas dizer que o camarão é bom é piada.*”, os aspectos extraídos nesta sentença seriam: comida, carne e sushi, pois esses revelam características que podem ser analisadas quando comparados dois ou mais restaurantes (SILVA, 2010).

Primeiramente, as sentenças foram divididas em palavras, para aplicar etiquetas para cada uma de suas classes gramaticais; a próxima etapa é a análise da sentença que compreende uma fase de identificação dos grupos de substantivos, dos grupos de verbos, das expressões preposicionais e das outras estruturas simples; após isso é especificada um domínio de aplicação, pois essa é a etapa que identifica as entidades relevantes no texto, ou seja, identifica os aspectos a partir do grupo de substantivos anteriormente selecionados.

Berlim (2015) complementa que é importante seleccione previamente as características que realmente sejam pertinentes para análise, reduzindo assim o número de características e extraíndo-se apenas as que serão interessantes e válidas.

3.5.3 Classificação

Nesta etapa as sentenças que foram classificadas como subjetivas serão utilizadas para analisar se a mesma é uma opinião neutra, negativa ou positiva.

Utilizando o exemplo de Silva (2010), podemos analisar a sentença da seguinte forma: *A comida do Restaurante X é maravilhosa [positiva], e a sobremesa servida não é cara [neutra]. O sushi no Restaurante Y é muito bom [positiva], mas dizer que o camarão é bom é piada [negativa]*

Para poder realizar essa classificação, será utilizada uma ferramenta para avaliar se os dados da base a ser criada na seção 3.1.3 apresentam opiniões neutras, negativas ou positivas.

3.5.4 Sumarização

Nesta etapa, as opiniões filtradas na etapa de classificação poderão ser apresentadas, em formas de textos, gráficos, comparações. Para isso, é importante que se apresente as opiniões de mais de uma pessoa para poder apresentar uma perspectiva quantitativa.

O resultado desta etapa será apresentado na próxima seção, pois corresponde aos resultados deste estudo.

3.6 MATERIAIS E MÉTODOS

Para conceber a base de dados foi utilizada a ferramenta Microsoft Excel 2013. Esse programa é um editor de planilhas que possui uma interface intuitiva e ferramentas de cálculo e de construção de gráficos. Para esta pesquisa o Excel será utilizado para conceber as bases de dados, na qual serão indexados os registros das coletas realizadas nas redes sociais e pelo questionário.

Nesta ferramenta foram utilizadas as colunas: idade, gênero, tipo de escola, tempo de preparação, problemas com o sistema de inscrição, consideração sobre o preço da prova, avaliação do conhecimento, instrução oferecida em sala de aula foi suficiente, ENEM é uma ferramenta de avaliação do ensino médio, nota ao ENEM. A FIGURA 9 apresenta uma prévia de 6 (seis) colunas elaboradas no Excel 2013. Essas colunas foram nomeadas de acordo com as questões aplicadas no questionário.

FIGURA 9 – PRÉVIA DAS COLUNAS ELABORADAS NO EXCEL 2013 COM BASE NO QUESTIONÁRIO APLICADO.

gênero	tipo de escola	idade	tempo de preparação	teve problema	preço cobrado	conhecimento	instrução oferecida	número de questões	ENEM é uma boa preparação	a certificação	Que nota você tirou
feminino	escola pública	17	entre sete meses	não	sim	regular	Escolas públicas não preparam	sim			6
feminino	escola pública	17	entre sete meses	não	sim	regular	Não há preparação suficiente	sim			6
feminino	escola pública	18	entre sete meses	não	não	regular	sim	não	Muitas questões	não	7
feminino	escola pública	18	mais de um ano	não	sim	insatisfatório	É necessário	não	Porque em sala	sim	8
masculino	escola particular	19	mais de dois anos	não	sim	bom	A prova é extensa	não	Os candidatos	não	6
feminino	escola particular	18	entre quatro meses	não	não	muito bom	sim	sim	Porque nem todos	sim	9
masculino	escola pública	38	entre quatro meses	não	não	insatisfatório	A prova é muito extensa	não	Porque a ferr	não	3
feminino	escola pública	16	entre sete meses	não	sim	bom	sim	não	sim	sim	8
feminino	escola pública	18	mais de um ano	não	não	insatisfatório	Extremamente	sim	Por ser apenas	não	6
masculino	escola pública	17	menos de três meses	não	não	bom	sim	não	É injusto! As	não	7
feminino	escola pública	17	mais de dois anos	não	não	insatisfatório	sim	não	sim	sim	8
masculino	escola pública	17	entre quatro meses	não	sim	regular	sim	sim	sim	sim	7
masculino	escola pública	16	entre sete meses	não	não	regular	sim	sim	sim	não	6
feminino	escola pública	10	entre sete meses	não	sim	insatisfatório	Porque o ensino	não	sim	sim	8
feminino	escola pública	16	entre sete meses	não	sim	regular	sim	não	sim	sim	7
feminino	escola pública	18	mais de um ano	não	sim	bom	Por que estudar	não	Por que existem	não	7
masculino	escola particular	17	entre sete meses	não	não	muito bom	sim	não	sim	sim	7
feminino	escola pública	17	Estava no segundo ano	não	sim	insatisfatório		não		não	3
feminino	escola particular	27	entre sete meses	não	não	insatisfatório	Tenho a impressão	não	sim	não	7
masculino	escola pública	18	No ano de 2010	não	não	insatisfatório	O conteúdo que	não	Como disse, r	não	5
feminino	escola pública	20	entre sete meses	não	não	bom	Falta de professores	sim	sim	sim	8
feminino	escola pública	17	menos de três meses	não	sim	muito bom	Sem a colaboração	sim	sim	não	5

FONTE: Elaborado pela autora com base nos dados da coleta (2016).

Para analisar as opiniões coletadas e armazenadas no Excel, foi utilizado a Semantria. A Semantria é uma solução criada pela companhia Lexalytics, renomada no fornecimento de softwares de análise de texto, é uma ferramenta de análise de sentimentos que pode ser aplicada a qualquer tipo de documentos, sentenças ou frases (BERLIM, 2015). A Semantria pode ser aplicada no Excel, baseada em um *plugin*, possibilitando a análise e classificação dos sentimentos como positivos, negativos e neutros, conforme apresentado na FIGURA 10 a seguir (BERLIM, 2015).

FIGURA 10 – PRÉVIA DA ANÁLISE REALIZADA COM OS DADOS COLETADOS DAS REDES SOCIAIS NA FERRAMENTA SEMANTRIA.

Source Text	Summary	Detected Language	Detected Language Strength	Document Sentiment	Document Sentiment
Deveria haver um exame nacional	Deveria haver um exame nacional à exer	Portuguese	0,39285713	0	neutral
Inadmissível um Instituto desse porte	Inadmissível um Instituto desse porte, c	Portuguese	0,42857143	-0,80000012	negative
adianta muito liberar a nota incompleta	adianta muito liberar a nota incompleta.	Portuguese	0,28947368	0,80000012	positive
Aí o site pede meu CPF minha senha e só	Aí o site pede meu CPF minha senha e só	Portuguese	0,33734939	-0,33333313	negative

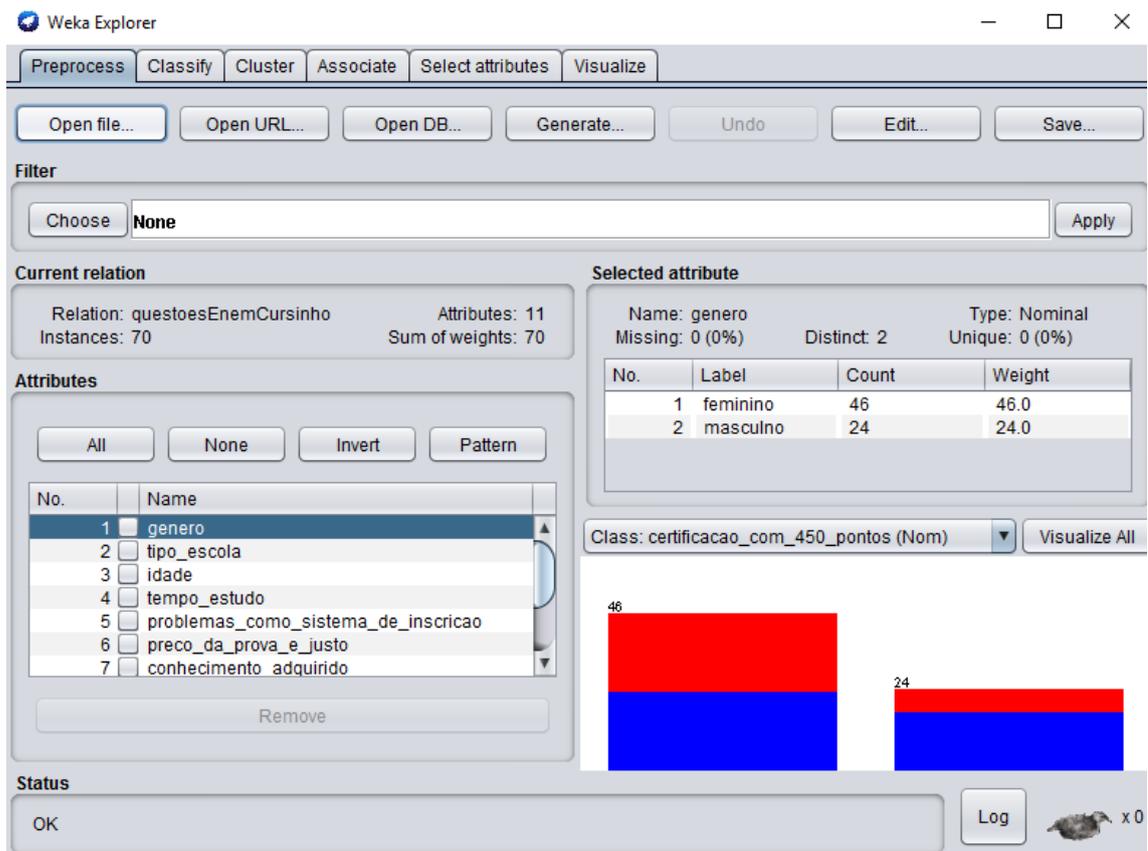
FONTE: Elaborado pela autora com base nos dados da coleta (2016).

A ferramenta Semantria apesar de possuir várias utilidades, para essa pesquisa será utilizada apenas a análise de sentimentos. A partir dos relatórios trazidos pela ferramenta é possível verificar que a ferramenta executa todas as etapas da mineração, que ao fim traz um relatório revelando a última etapa que é a sumarização (BERLIM, 2015). Porém, conforme destacado por Berlim (2015) a

ferramenta possui alguns pontos negativos, algumas ironias e gírias não são detectadas quando a análise é feita à nível de documento.

Para realizar a mineração dos dados coletados pelo questionário será utilizada o software Weka. O Weka suporta algoritmos de Classificação (aba *Classify*), *Cluster*, Associação (aba *Associate*) apresentado na FIGURA 11.

FIGURA 11 – VISUALIZAÇÃO DA BASE REPRESENTADA A PARTIR DA FERRAMENTA WEKA.



FONTE: Elaborado pela autora com base nos dados da coleta (2016).

Para este estudo serão utilizados dois métodos: um de classificação (J48) e um de associação (APRIORI) disponíveis na ferramenta Weka.

4 RESULTADOS

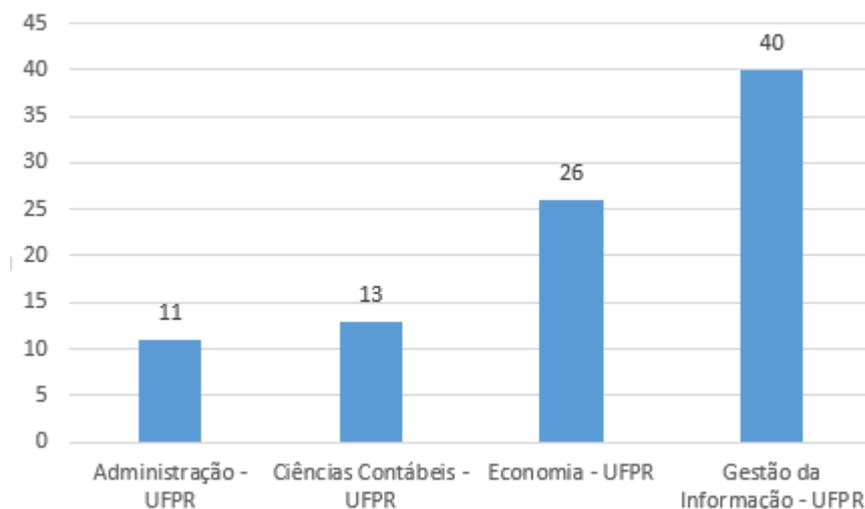
Os resultados desta pesquisa serão apresentados em duas partes: na primeira será apresentada os resultados da coleta e da mineração de dados nos dados coletados por meio do questionário e a segunda parte contempla a análise de sentimento dos dados coletados nas redes sociais (*Twitter e Facebook*).

4.1 ANÁLISE ESTATÍSTICA DA COLETA

Essa seção trará os resultados estatísticos da coleta de dados realizados com os alunos dos cursos do Setor de Ciências Sociais Aplicadas da UFPR e dos alunos dos cursos preparatórios.

O questionário aplicado aos alunos da UFPR obteve 90 (noventa) respostas, não atingido os 10% desejados da amostra. Para poder distinguir os alunos do Setor de Ciências Sociais Aplicadas da UFPR foi realizada uma questão para saber qual o curso o respondente era. Com base nisso foram obtidos 11 (onze) respostas de alunos do curso de Administração, 13 (treze) respostas dos alunos de Ciências Contábeis, 26 (vinte e seis) respostas dos alunos de Economia e 40 (quarenta) respostas dos alunos de Gestão da Informação da qual foi a única que atingiu a amostra desejada. O GRÁFICO 1 apresenta o resultado do total de respondentes.

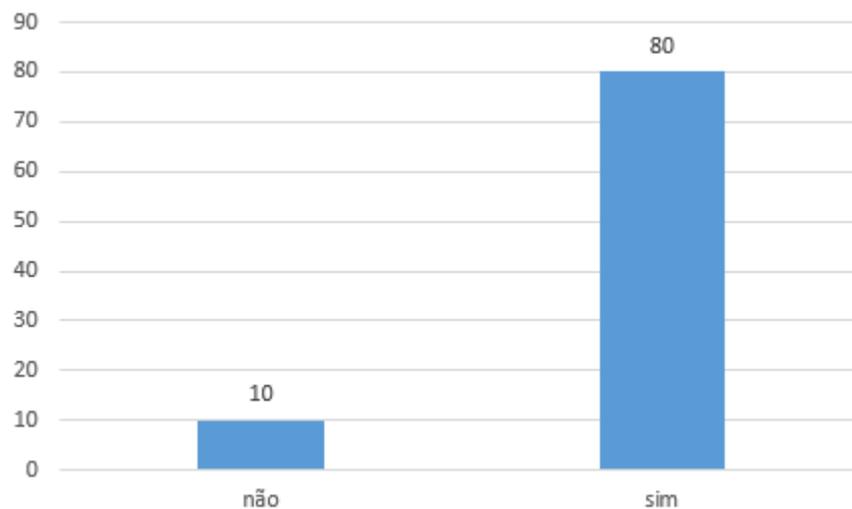
GRÁFICO 1 – QUANTIDADE DE RESPONDENTES DE ACORDO COM O CURSO.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

Além disso, para poder prosseguir com o questionário o respondente deveria estar matriculado e cursando atualmente na UFPR. Apenas 10 (dez) respondentes não estavam matriculados e 80 (oitenta) responderam que estavam matriculados e cursando, que prosseguiram com o questionário, porém dos 80 que responderam que estão matriculados, apenas 69 completaram 100% (cem por cento) do questionário. O GRÁFICO 2 apresenta o percentual desses respondentes.

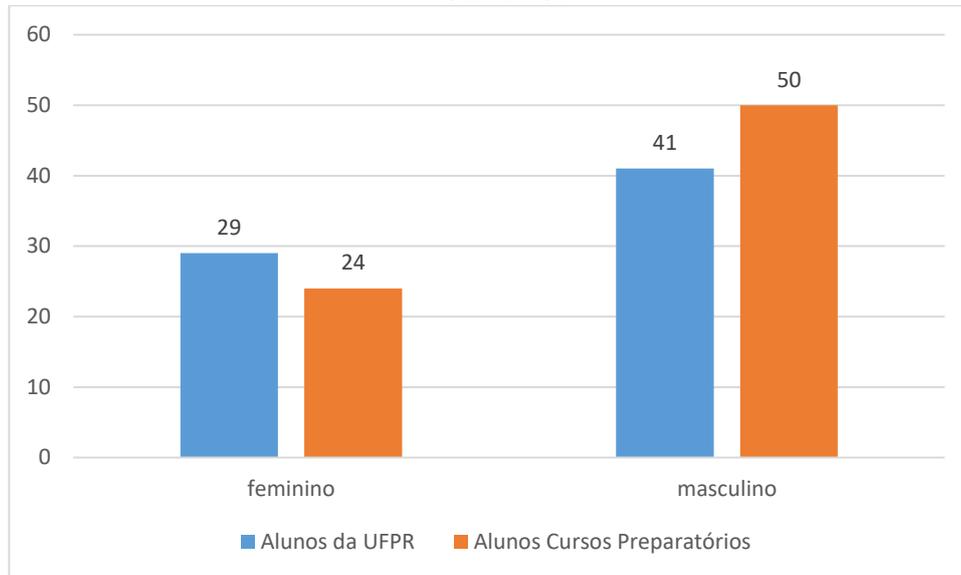
GRÁFICO 2 – QUANTIDADE DE ALUNOS MATRICULADOS E NÃO MATRICULADOS NA UFPR.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

Com relação ao questionário aplicado aos alunos dos cursos preparatórios foram obtidas 74 (setenta e quatro) respostas. Dessas 74 pessoas, 50 (cinquenta) são mulheres e 24 (vinte e quatro) homens. Com relação aos alunos da UFPR 41 (quarenta e um) dos respondentes são homens e 29 (vinte e nove) são mulheres, conforme apresentado no GRÁFICO 3.

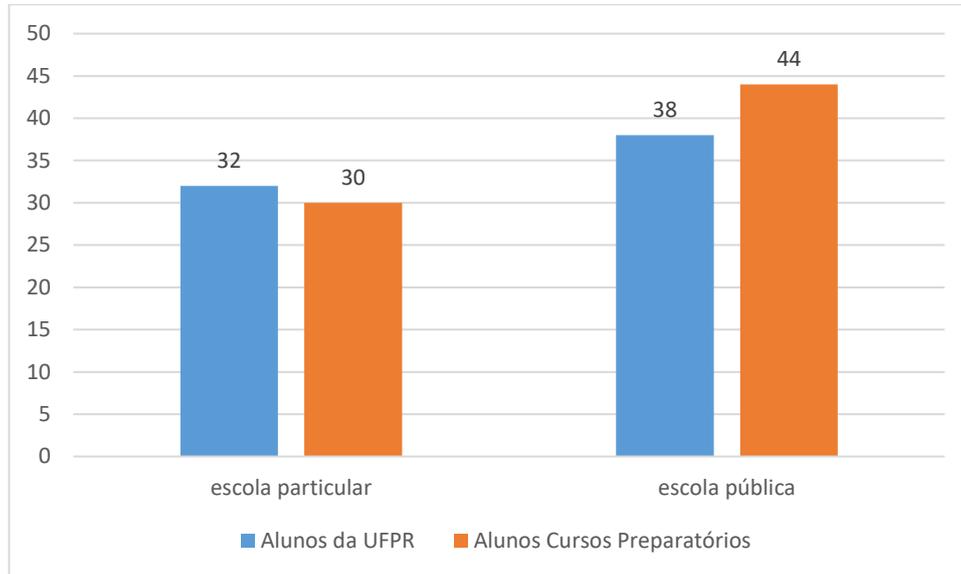
GRÁFICO 3 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS POR GÊNERO.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

Com relação ao tipo de escola que os respondentes da UFPR cursaram no Ensino Médio, a maioria vem de escola pública, são 38 (trinta e oito) alunos oriundos de escolas públicas e 32 (trinta e dois) de escolas particulares. O mesmo acontece com os alunos dos cursos preparatórios, a quantidade de respondentes de escolas públicas foi maior do que respondentes de escolas particulares. Foram 44 (quarenta e quatro) alunos oriundos de escola pública e 30 (trinta) alunos de escola particular. Esses dados estão apresentados no GRÁFICO 4.

GRÁFICO 4 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS DE ACORDO COM O TIPO DE ESCOLA.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

A TABELA 5 apresenta a relação da idade dos respondentes. A grande maioria dos respondentes está na faixa entre 17 (dezessete) e 18 (dezoito) anos, 33 (trinta e três) alunos possuíam 17 (dezessete) anos na época em que realizou o ENEM e 18 (dezoito) participantes tinham 18 (dezoito) anos, representando 63,75% da amostra. Para a mineração dos dados foi necessário discretizar esse atributo, conforme apresentado na metodologia desse trabalho.

TABELA 5 – NÚMERO DE ALUNOS DA UFPR DE ACORDO COM A IDADE.

Idade	Quantidade	Porcentagem
15	2	2,50%
16	3	3,75%
17	33	41,25%
18	18	22,50%
19	8	10,00%
20	6	7,50%
21	5	6,25%
22	2	2,50%
24	1	1,25%
30	1	1,25%
36	1	1,25%
Total	80	100,00%

FONTE: elaborado pela autora com base nos dados da pesquisa (2016).

Assim como acontece com os resultados dos alunos a UFPR, os respondentes dos cursos preparatórios em sua grande parte também possuem idades entre 17 (dezesete) e 18 (dezoito) anos, sendo 17 (dezesete) a idade da maioria com 30 (trinta) respondentes e 16 (dezesesseis) a segunda maior com 14 (catorze), representando 61,43% desta amostra.

A TABELA 6 apresenta o total e porcentagem das idades dos respondentes. Para a mineração dos dados, assim como foi feito com os resultados dos alunos da UFPR, também foi necessário discretizar esse atributo.

TABELA 6 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO DE ACORDO COM A IDADE.

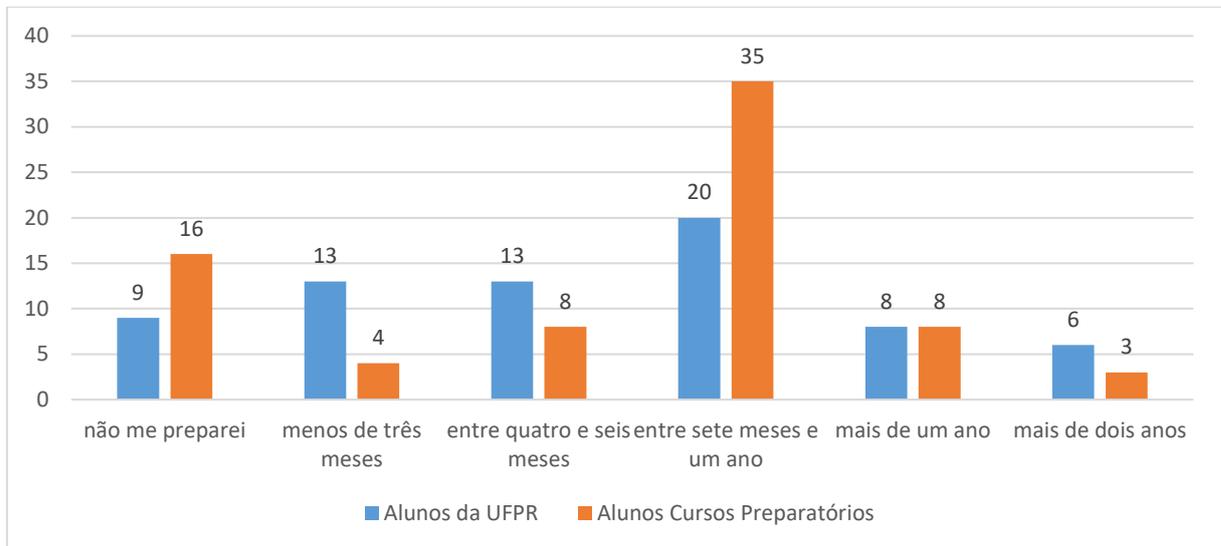
Idade	Quantidade	Porcentagem
15	2	2,86%
16	14	20,00%
17	30	42,86%
18	13	18,57%
19	4	5,71%
20	1	1,43%
21	1	1,43%
24	1	1,43%
25	1	1,43%
26	1	1,43%
27	1	1,43%
38	1	1,43%
Total	70	100,00%

FONTE: elaborado pela autora com base nos dados da pesquisa (2016).

Outro ponto destacado na coleta foi o período que o participante levou para se preparar para realização do ENEM. Essa foi uma questão considerada importante para avaliar se o desempenho do candidato influenciou ou não na resolução da prova. Com isso, a quantidade de alunos da UFPR que se preparam entre “sete meses e um ano” foi a maior com 20 (vinte) respondentes e a segunda maior é entre “quatro e seis meses” e “menos de três meses” ambas com 13 (treze) resultados. Assim como os alunos da UFPR, a quantidade de participantes dos cursos preparatórios que se preparam “entre sete meses e um ano” foi a de maior proporção dos resultados, com 35 (trinta e cinco) pessoas seguido de 16 (dezesesseis) que não se preparam (apresentada como “Outros” no gráfico). Apenas 3 (três) participantes se prepararam mais de dois anos para realizar o ENEM, esses dados estão apresentados no

GRÁFICO 5. Com base nos dois resultados apresentados, tanto os alunos da UFPR quanto os alunos dos cursos preparatórios não estudaram mais de um ano para a prova do ENEM.

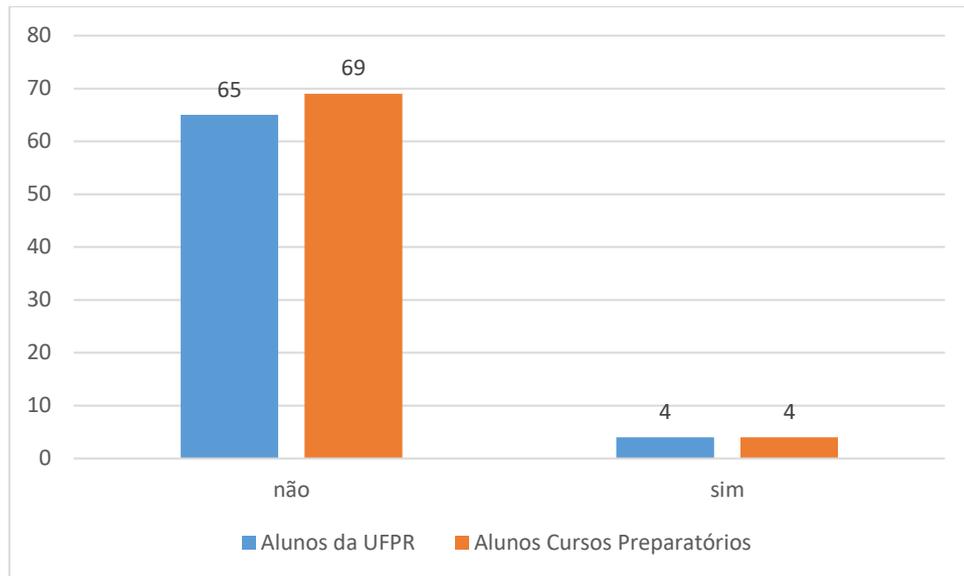
GRÁFICO 5 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS DE ACORDO COM O TEMPO DE PREPARAÇÃO.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

A questão realizada para descobrir se os respondentes tiveram problemas com o sistema de inscrição com relação aos alunos da UFPR obteve 65 (sessenta e seis) “não” e apenas 4 (quatro) “sim. Assim como apresentado anteriormente, o número de alunos que tiveram problemas com o sistema de inscrição é muito menor comparado aos alunos que não tiveram. Os alunos dos cursos preparatórios que não tiveram problemas com o sistema de inscrição totalizaram 69 (sessenta e nove) respostas e aqueles que tiveram foi de apenas 4 (quatro) alunos, conforme apresentado no GRÁFICO 6. Para não gerar tendenciosidade na mineração de dados esses resultados foram removidos no Weka.

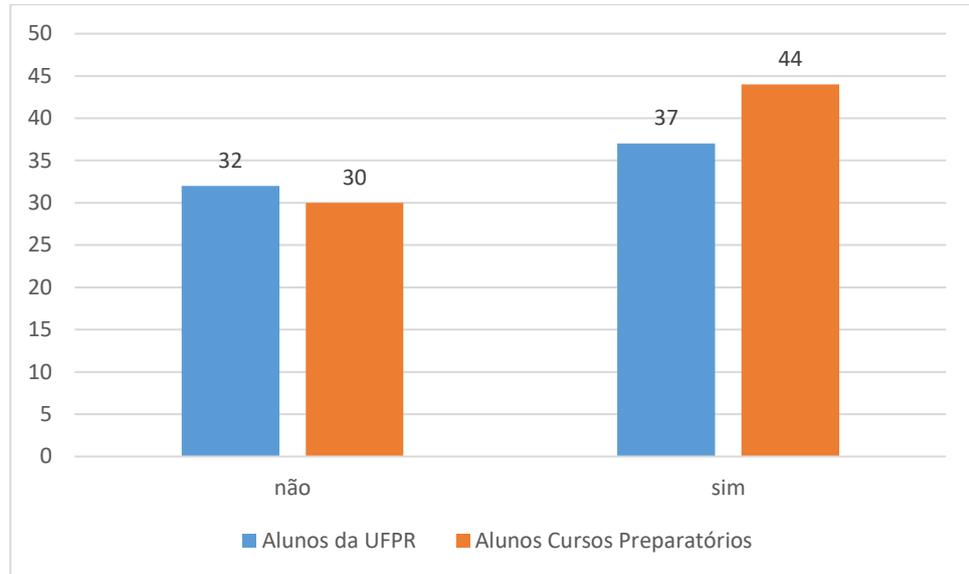
GRÁFICO 6 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS QUE TIVERAM PROBLEMA COM O SISTEMA DE INSCRIÇÃO.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

Quando perguntado aos alunos da UFPR se eles consideravam o preço cobrado pela prova “justo”, atualmente o valor é de R\$68,00 (sessenta e oito reais), obteve 37 (trinta e sete) respostas “sim” e 32 (trinta e duas) “não”. Isso demonstra que os respondentes ficaram quase divididos ao responder essa questão. O número de respondentes dos alunos dos cursos preparatório que consideram “justo” o preço cobrado para realizar a prova do ENEM foi de 44 (quarenta e quatro), e que não consideram foi de 30 (trinta) pessoas. O GRÁFICO 7 retrata a porcentagem dessas respostas.

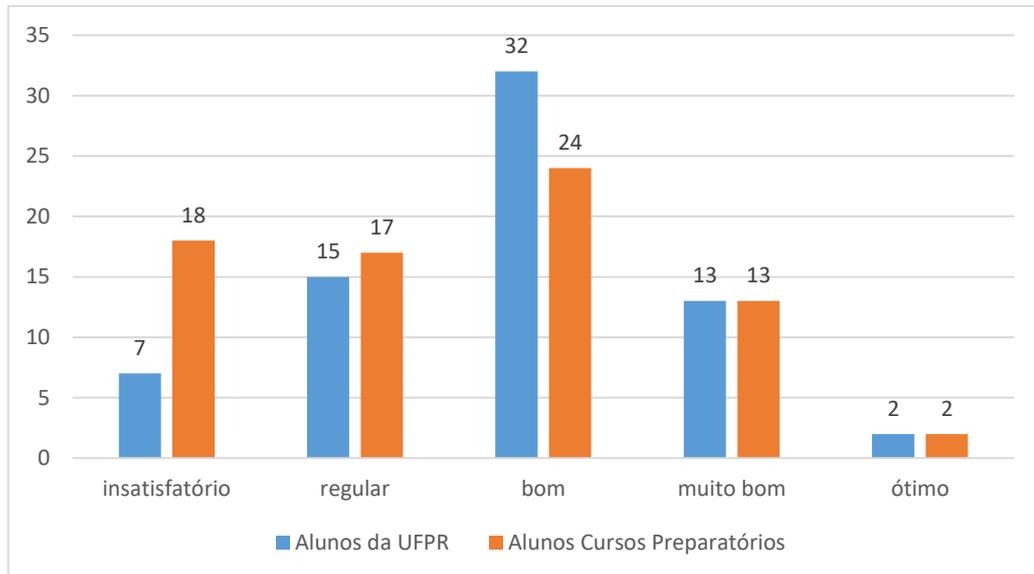
GRÁFICO 7 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS QUE CONSIDERAM OU NÃO O PREÇO COBRADO PARA REALIZAR O ENEM “JUSTO”.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

A questão que avalia o conteúdo cobrado na prova levando em consideração o conhecimento do respondente em sala de aula apresentou um dado interessante. A grande quantidade dos alunos da UFPR considerou o conhecimento aplicado na prova “bom” e “muito bom”, com 32 (trinta e duas) respostas para “bom” e 13 (treze) para “muito bom”. Assim como os alunos da UFPR, os alunos dos cursos preparatórios responderam em sua grande maioria ter tido um conhecimento “bom” aplicado na prova, foram 24 (vinte e quatro) “bom”, porém, ao contrário dos alunos da UFPR, 18 (dezoito) alunos responderam que consideraram o conhecimento adquirido “insatisfatório” o que pode demonstrar que a prova apresentava um nível de dificuldade acima do conhecimento real desses alunos. O GRÁFICO 8 apresenta essas e outras respostas dos alunos.

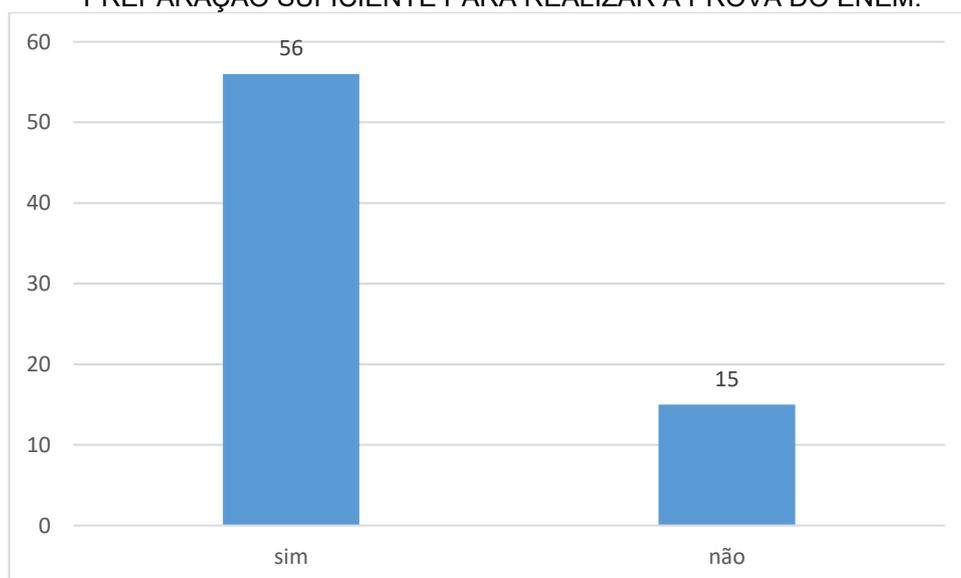
GRÁFICO 8 – QUANTIDADE DE RESPOSTAS DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS CONSIDERANDO O CONTEÚDO COBRADO NA PROVA EM RELAÇÃO AO SEU CONHECIMENTO ADQUIRIDO.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

E quando perguntado aos alunos se o conhecimento adquirido foi suficiente para realizar a prova a grande maioria 43 (quarenta e três) responderam que a preparação é sim suficiente para realizar o ENEM, apenas 26 (vinte e seis) responderam que não. O GRÁFICO 9 apresenta o percentual dessas respostas.

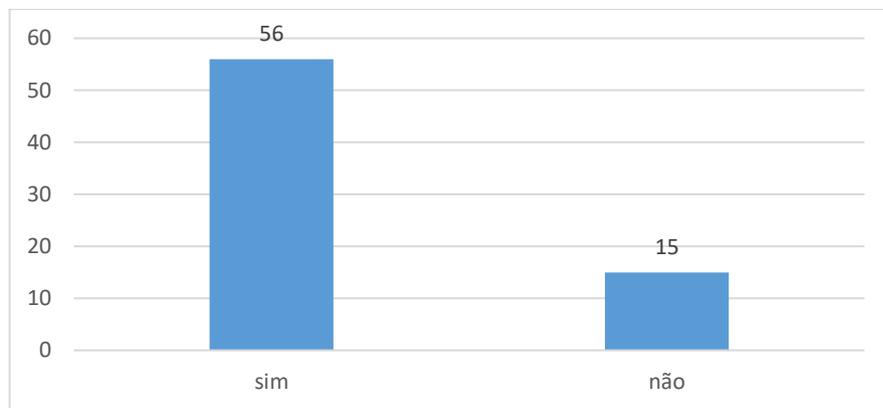
GRÁFICO 9 – NÚMERO DE ALUNOS DA UFPR QUE CONSIDERAM OU NÃO QUE HÁ PREPARAÇÃO SUFICIENTE PARA REALIZAR A PROVA DO ENEM.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

Ao contrário dos alunos da UFPR, 37 (trinta e sete) alunos dos cursos preparatório responderam que a preparação não é suficiente. E como essa era uma questão em que ele poderiam justificar o porquê da escolha não, muitos alunos descreveram que não há preparação suficiente para realizar esse tipo de prova, mesmo em cursos preparatórios e no ensino médio público e privado. E outros dois alunos descreveram que a prova é subjetiva e muito extensa o que cansa o estudante ao realiza-la. O GRÁFICO 10 e a TABELA 7 apresentam detalhadamente as respostas dos alunos.

GRÁFICO 10 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO QUE CONSIDERAM OU NÃO QUE HÁ PREPARAÇÃO SUFICIENTE PARA REALIZAR A PROVA DO ENEM.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

A TABELA 7 a seguir apresenta as respostas adicionais dadas pelos alunos ao escolher a resposta “não” da pergunta “Você acredita que apenas com a instrução oferecida em sala aula é suficiente para a realização da prova?”.

Agrupando as respostas, foi analisado que das 34 (trinta e quatro) respostas 18 (dezoito) respostas dizem a respeito à preparação insuficiente nas escolas públicas; 2 (duas) são referentes a necessidade de mais tempo de estudo para realizar a prova; 2 (duas) referente à falta de professores nas escolas públicas; 1 (uma) afirmando que a prova é subjetiva; 2 (duas) descrevendo que a prova é extensa; conteúdo cobrado no ENEM não é visto dentro de sala de aula 8 (oito) são referentes ao conteúdo cobrado no ENEM que não é visto dentro de sala de aula; e 1 (uma) diz respeito a falta de exemplos práticos em sala de aula.

TABELA 7 – TOTAL DE RESPOSTAS ADICIONAIS DA QUESTÃO SE OS ALUNOS DOS CURSOS PREPARATÓRIO CONSIDERAM OU NÃO O CONHECIMENTO EM SALA DE AULA SUFICIENTE.

continua

Respostas do campo adicional

- “Escolas públicas não preparam o suficiente, de modo que o cursinho é um complemento do que deveria ter sido aprendido”
- “Não há preparação suficiente na escola pública, tendo os alunos que optarem por um cursinho preparatório”
- “É necessário mais tempo de estudo para praticar com questões e exercícios”
- “A prova é extremamente subjetiva”
- “A prova é muita comprida, fora da realidade do estudante da escola pública”
- “Extremamente baixa a instrução da escola”
- “Porque o enem cobra conteúdos que muitas vezes não são ensinados em sala”
- “Por que estudei em colégio público, onde o ensino não é muito valorizado e sim a conclusão do período escolar”
- “Tenho a impressão que o que é passado em sala de aula é bem diferente do que o ENEM cobra. E vou mais longe, quando fazia o ensino médio, o que aprendia na escola ficava lá mesmo, pois não era passado/ensinado como aplicar o conteúdo fora da escola. Espero que as próximas gerações tenham mais sorte do que eu quanto ao ensino em sala de aula”
- “O conteúdo que é ensino médio não abrange todo conteúdo que é cobrado pelo ENEM”
- “Falta de professores causa retardo no aprendizado”
- “Sem a colaboração da turma e seu mal comportamento, o professor não consegue dar aula adequadamente”
- “Como aluna de escola pública, tive o desprazer de não ter professores em sala de aula e/ou a desqualificação de tal”
- “Ensino fraco, professor de história por exemplo ia pra colégio pra tomar chimarrao e passava só resumos para os alunos”
- “As escolas não são eficientes quanto ao assunto que cai no ENEM”
- “Muito conteúdo pra apenas passar em sala”
- “Estudei a vida inteira em colégio público, fui bolsista em um colégio particular havia muita coisa que não tinha aprendido e isso me afetou tanto na escola como na prova”
- “Seria necessário uma revisão geral dos 3 anos para que os alunos de rede pública lembrarem os conteúdos”
- “O ensino público é defasado demais”
- “Muitas vezes falta exemplos práticos, e aplicações do dia dia em matéria como física , química e biologia”
- “Principalmente em escolas públicas, o ensino deixa muito a desejar. Sem contar que, escolas particulares preparam o aluno desde o primeiro ano do ensino médio, e a pública deixa pra focar em vestibulares somente no terceiro ano pois nem sempre é o suficiente”
- “Ensino medio fraco”
- “É preciso buscar ajuda extra pois ha conteúdos cobrados muito específicos que acabam não tendo um foco tão grande em uma aula expositiva”
- “Como aluna de escola publica acredito que o conteúdo ministrado está um pouco devassado, falta desempenho das escolas, professores e claro dos alunos e buscar mais conhecimento”
- “Quando se diz respeito a escola pública não, porque a escola pública não prepara o aluno para enfrentar provas como ENEM e vestibulares afins”
- “Eu esqueci todo o conteúdo depois da prova, nunca realmente aprendi aquilo”
- “Meu cursinho em particular prepara mais pra outros estilos de prova”
- “Pois os alunos não constroem conhecimento suficiente e não desenvolvem ao longo da vida escolar habilidades que permitam usufruir desse conhecimento, como o pensamento crítico e a capacidade de fazer relações e ligações entre informações, nem são capazes de usar este conhecimento na vida prática”
- “Falta de conteúdo”
- “Nem todo conteúdo foi dado na escola publica”

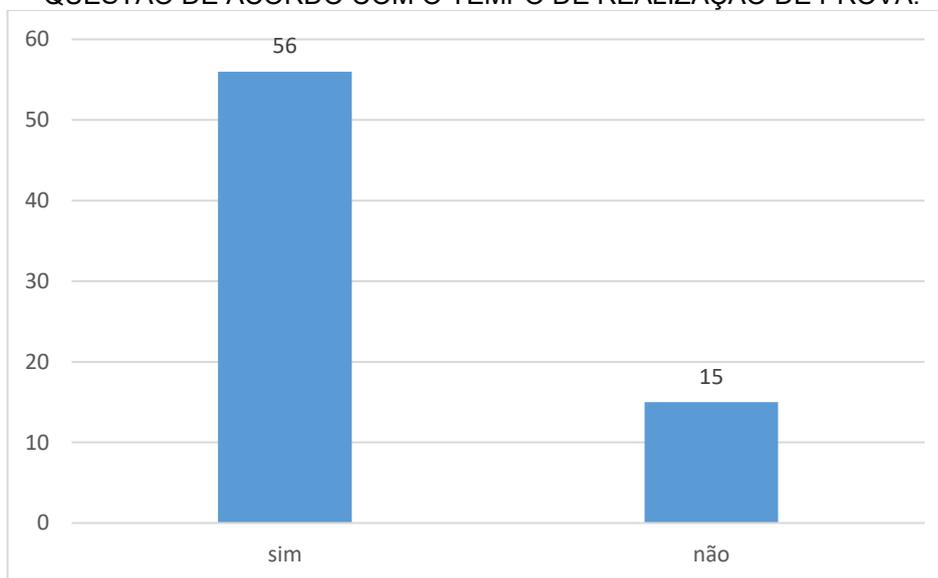
TABELA 7 – TOTAL DE RESPOSTAS ADICIONAIS DA QUESTÃO SE OS ALUNOS DOS CURSOS PREPARATÓRIO CONSIDERAM OU NÃO O CONHECIMENTO EM SALA DE AULA SUFICIENTE. conclusão

Respostas do campo adicional	
“Escolas publicad tendem a ter ensino insatisfatório nao abrangendo todos os assuntos cobrados no enem”	
“O enem cobra coisas especificas de acordo com as disciplinas, fundindo da realidade das escolas”	
“Falta melhor preparação para o aluno”	
“Conhecimentos de atualidade”	
Total de respostas adicionais:	34

FONTE: Dados da pesquisa (2016).

Outra questão realizada aos alunos foi para verificar se a quantidade de questões aplicadas na prova estava de acordo com o tempo disponível para sua realização. Nessa questão, os alunos da UFPR que não consideram de acordo com o tempo de 38 (trinta e oito) e 31 (trinta e um) e alunos consideraram de acordo o número de questão com o tempo de prova. O GRÁFICO 11 apresenta o porcentual desse resultado.

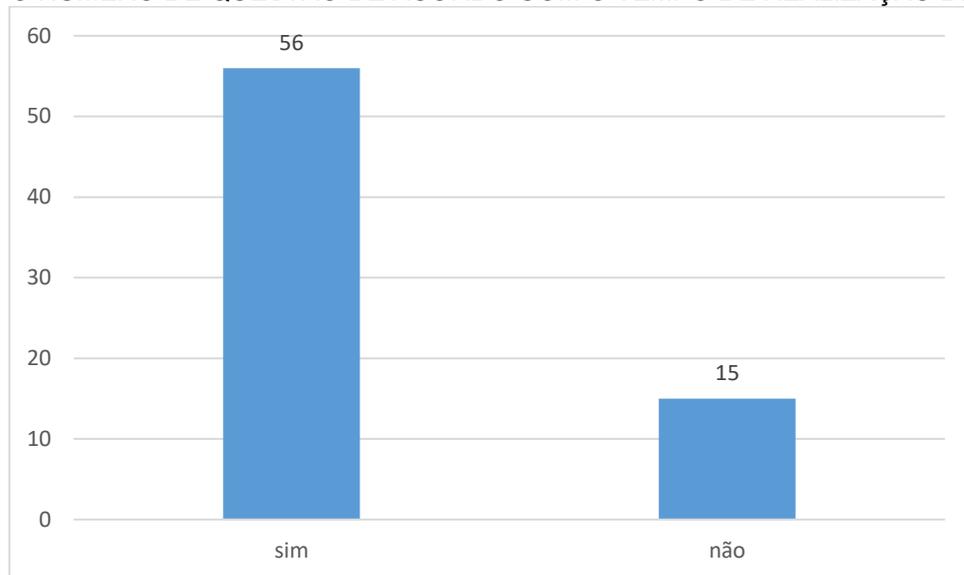
GRÁFICO 11 – NÚMERO DE ALUNOS DA UFPR QUE CONSIDERARAM OU NÃO O NÚMERO DE QUESTÃO DE ACORDO COM O TEMPO DE REALIZAÇÃO DE PROVA.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

Para esse caso, assim como os alunos da UFPR, a grande maioria dos alunos dos cursos preparatório, 56 (cinquenta e seis) alunos, consideram que o tempo não é suficiente e apenas 15 (quinze) consideraram que número de questões estava de acordo com o tempo de prova conforme apresentado no GRÁFICO 12.

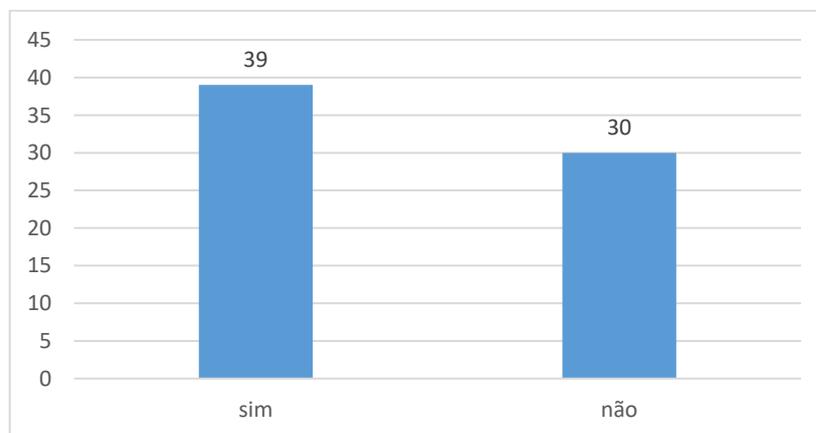
GRÁFICO 12 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO QUE CONSIDERARAM OU NÃO O NÚMERO DE QUESTÃO DE ACORDO COM O TEMPO DE REALIZAÇÃO DE PROVA.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

Foi perguntado aos alunos da UFPR se eles consideravam o ENEM como uma ferramenta de avaliação do ensino médio brasileiro, e 46 (quarenta e seis) respondentes disseram que “sim” e 22 (vinte e dois) que não. A porcentagem desse resultado está apresentado no GRÁFICO 13. Os alunos que escolheram a opção “não” poderiam deixar um complemento para sua resposta. A TABELA 8 apresenta as respostas complementares dos alunos da UFPR.

GRÁFICO 13 – NÚMERO DE ALUNOS DA UFPR QUE CONSIDERAM OU NÃO O ENEM COMO UMA FORMA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

A TABELA 8 apresenta as respostas adicionais dadas pelos alunos da UFPR ao escolher a resposta “não” da pergunta “Você concorda que o ENEM é uma ferramenta de avaliação da qualidade do ensino médio?”. Analisando as respostas por agrupamento, pode se observar que das 22 (vinte e duas), 7 (sete) são referente ao conteúdo do ENEM que não está alinhado com o que o aluno estuda em sala de aula; 3 (três) descrevem sobre as diferenças entre o ensino em escola pública e particular; 2 (duas) colocaram as questões como subjetivas; 1 (uma) diz respeito que a prova é manipulada; 1 (uma) descreveu que a prova é exaustiva; 4 (quatro) respondentes descreveram que o ENEM não avalia o ensino médio como um todo; 1 (uma) pessoas descreveu que nem todos que fazem o ENEM estão no ensino médio; 1 (uma) respondente descreveu que o ENEM é propaganda do governo; e 2 (duas) pessoas descreveram que a prova objetiva e confusa, e uma prova objetiva não avalia o ensino médio.

TABELA 8 – TOTAL DE RESPOSTAS ADICIONAIS A RESPOSTA DA QUESTÃO SE OS ALUNOS DA UFPR CONSIDERAVAM OU NÃO O ENEM COMO UMA FERRAMENTA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.

continua

Respostas do campo adicional:

"Nao condiz com a realidade do ensino com o que é cobrado"

"Não, a maioria dos assuntos aprendidos em sala de aula não são os que aparecem nas questões do ENEM. O que pode ser avaliado é somente a qualidade dos cursinhos, pois eles sim lhe passam a verdadeira matéria que é pedida no enem"

"Há uma discrepância considerável entre os sistemas público e privado de educação, e o sistema de pesos não prioriza necessariamente o acerto por "saber solucionar a questão"

"Questoes subjetivas fazem com que a prova tenha fama de facil em alguns aspectos, mas um minimo de complicações basicas impostas gera um alarde imenso. entao sim"

"Há a manipulação dos dados de forma a direcionar a avaliação da qualidade do ensino"

"Sim, mas esquecem de que há diferentes ensinos (particular e público) com qualidades diferentes e o enem faz uma média desses ensinos e traz uma avaliação que não é tão realista"

"Acho que exige pouquíssimo em relação a todo conteúdo visto durante o ensino médio"

"Porque as questões são muito mais interpretativas do que com conteúdos de dificuldade media/alta"

"É difícil avaliar todo um sistema educacional, de depende de diversos fatores, através de uma prova que costuma ser mais exaustiva do que avaliativas de fato. deve" se levar em conta muitos outros fatores para analisar a qualidade do ensino médio"

"ele não aborda vários assuntos estudados no ensino médio e é pouco tempo para avaliar todos os conteúdos"

"Nao é atraves de UMA prova que avalia-se, mas sim o desempenho durante os anos como um todo"

"Porque nem todos que fazem o ENEM necessariamente está matriculado ou cursando o ensino médio"

"O enem é mais um teste de paciencia do que uma ferramenta de avaliao"

"propaganda do governo/ideológica"

"Nao é possível avaliar todo o contexto do ensino médio em apenas uma prova"

"provas objetivas são muito confusas e pessoas muito inteligentes podem acabar indo mal e vice versa"

TABELA 8 – TOTAL DE RESPOSTAS ADICIONAIS A RESPOSTA DA QUESTÃO SE OS ALUNOS DA UFPR CONSIDERAVAM OU NÃO O ENEM COMO UMA FERRAMENTA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.

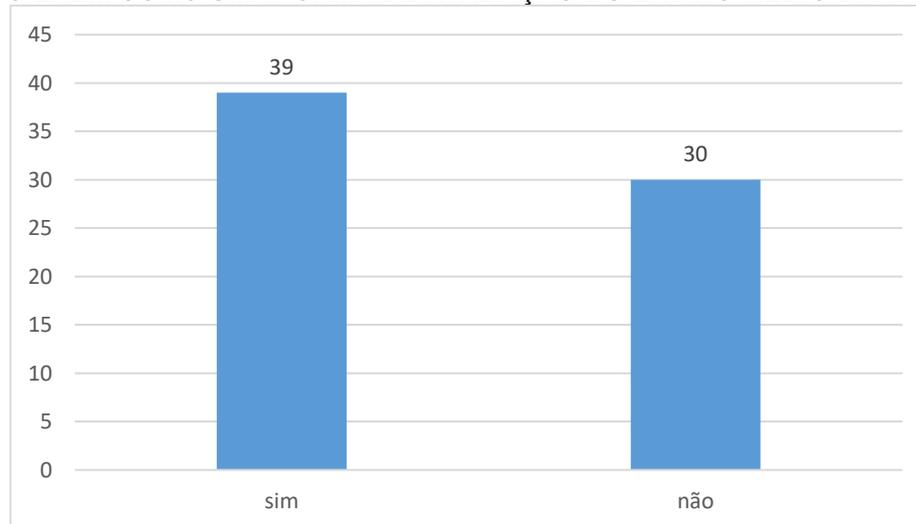
conclusão

Respostas do campo adicional	
"a prova cobra muito menos que o que se aprende no ensino médio"	
"nenhuma prova avalia bem o aluno"	
"A prova não avalia o conhecimento e sim a capacidade de interpretação"	
"O ENEM é basicamente interpretação de texto, ele não é um indicador de qualidade do ensino médio, pois depende da dedicação do aluno e não da qualidade de ensino da escola"	
"Não considero uma prova objetiva capaz de quantificar o quanto um aluno aprendeu durante seu ensino médio"	
"Pois algumas escolas públicas não focam no que é avaliado na prova"	
Total	22

FONTE: Dados da pesquisa (2016).

O total de alunos dos cursos preparatório que consideram o ENEM como uma ferramenta de avaliação do ensino médio brasileiro foi de 39 (trinta e nove) respondentes e 30 (trinta) alunos responderam que não o consideram com uma ferramenta que avalia o Ensino Médio. Essa também foi uma questão onde o respondente poderia complementar sua resposta. O GRÁFICO 14 e a TABELA 9 apresentam detalhadamente as respostas dos alunos.

GRÁFICO 14 – NÚMERO DE ALUNOS DOS CURSOS PREPARATÓRIO QUE CONSIDERAM OU NÃO O ENEM COMO UMA FORMA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

A TABELA 9 apresenta as respostas adicionais dadas pelos alunos ao escolher a resposta “não” da pergunta “Você concorda que o ENEM é uma ferramenta de avaliação da qualidade do ensino médio?”.

Analisando as respostas em grupos foi verificado que das 29 (vinte e nove) respostas, 2 (duas) eram referentes as questões desnecessárias, em excesso; 10 (dez) descreviam que o ENEM não cobra o que foi estudado em sala de aula; 2 (duas) que as questões objetivas não servem para avaliar; 3 (três) descreviam que cada aluno possui suas particularidades; 1 (uma) referente a necessidade de melhorar o ensino; 3 (três) referente ao nível de dificuldade da prova e que ela é extensa; 2 (duas) sobre o tempo que é pouco por cada questão; 3 (três) respondentes escreveram que a prova não avalia inteligência; 1 (uma) que o ENEM não é uma ferramenta de avaliação pois é usado para ingresso no ensino superior; e 1 (uma) que as notas são avaliadas de formas diferentes para cada área de conhecimento.

TABELA 9 – TOTAL DE RESPOSTAS ADICIONAIS A RESPOSTA DA QUESTÃO SE OS ALUNOS DOS CURSOS PREPARATÓRIO CONSIDERAVAM OU NÃO O ENEM COMO UMA FERRAMENTA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.

continua

Respostas do campo adicional:

"Muitas questões desnecessárias"

"Porque em sala de aula as matérias não são contextualizadas como a prova é"

"Os candidatos não se baseiam apenas no conhecimento adquirido no ensino médio em sala, pelo contrário o conteúdo é adquirido muito mais fora da sala de aula das escolas"

"Porque nem tudo o que é cobrado no enem é estudado em sala de aula"

"Porque a ferramenta deve ser construída em cima do que realmente se em cima e não em uma UTOPIA de ensino"

"Por ser apenas questões optativas, podendo haver acertos no "chute"

"É injusto! As escolas públicas não mostram todos os assuntos cobrados"

"Por que existem milhares de fatores que influenciam na hora da prova o que faz com que você não se concentre e realmente faça a prova, além disso ainda tem o fator subjetividade, que faz com que cada aluno seja diferente o que deveria ter uma prova diferente para cada aluno, o que seria inviável"

"Como disse, muito dos conteúdos presente no ENEM não é ensinado no ensino médio em muitas escolas"

"Porque nem todos tem a mesma condição para pagar um estudo bom"

"O ensino teria que melhorar muito pra dar pra fazer a prova somente com ensino médio"

"A prova cobra um ensino no qual, não é dado nas escolas públicas e nas particulares não é possível pagar um preço tão caro para tentar ter um ensino decente"

"O nível da prova é muito superior"

"Pois apresenta questões de alto nível, em pouco tempo e em um ritmo massante."

"Porque a forma que eles cobram (áreas dos conhecimento) não é da mesma forma tradicional aplicada nas escolas"

"Tempo por questão é menor do que o normal incapacitando o aluno ter desenvolver que tos e direcionando ele para o "chutes de várias questões"

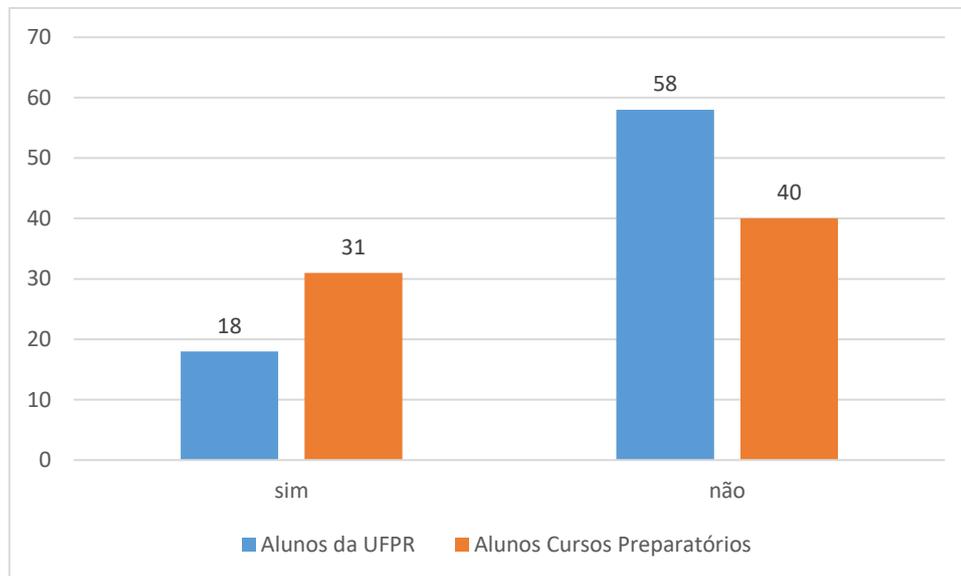
TABELA 9 – TOTAL DE RESPOSTAS ADICIONAIS A RESPOSTA DA QUESTÃO SE OS ALUNOS DOS CURSOS PREPARATÓRIO CONSIDERAVAM OU NÃO O ENEM COMO UMA FERRAMENTA DE AVALIAÇÃO DO ENSINO MÉDIO BRASILEIRO.

	conclusão
Respostas do campo adicional:	
"O aluno pode ser ótimo mas no dia da prova, por algum motivo, ir mal"	
"Acho que uma prova não seja capaz de definir o nível de inteligência"	
"Deixou de ser uma ferramenta de avaliação quando foi utilizado pela primeira vez para o ingresso para o nível superior"	
"Pois ele avalia mais a resistência do aluno do que a inteligência em si. Prova muito extensa e cansativa"	
"Textos grandes e 180 questões não avaliam o conhecimento que o aluno adquiriu durante sua formação e sim que ele "sobreviva "a uma prova muito extensa"	
"Pois acho inválido o método de conhecimento juntamente com o tempo absurdo de prova"	
"O número de questões desgasta absolutamente o aluno que por vezes se acaba perdendo seu rendimento no decorrer da prova"	
"Pois é múltipla escolha, é possível chutar, mesmo com o sistema de compensação de notas (injusto, por sinal)"	
"Pelo tempo ser curto acaba cansando as pessoas"	
"O enem trabalha com resistencia não com inteligente"	
"Os conteúdos são cobrados de maneira diferente do que é aplicado em sala de aula"	
"Por que as notas são extremamente valorizadas em ciencias exatas quando comparadas com as notas de humanas"	
"Em escolas publicas se vê muito menos do conteúdo do que é cobrado"	
Total:	29

FONTE: Dados da pesquisa (2016).

Por fim, foi perguntado aos alunos se eles concordavam ou não com a certificação do ensino médio com 450 (quatrocentos e cinquenta) pontos. Com relação aos alunos da UFPR houve 58 (cinquenta e oito) respostas “não” e apenas 18 (dezoito) “sim”. Com relação aos alunos dos cursos preparatório foram obtidos 40 (quarenta) alunos que responderam que não concordam e 31 (trinta e um) concordam que a certificação ocorra com essa pontuação, conforme apresentado no GRÁFICO 15.

GRÁFICO 15 – NÚMERO DE ALUNOS DA UFPR E DOS CURSOS PREPARATÓRIOS QUE CONCORDAM OU NÃO COM A CERTIFICAÇÃO DO ENSINO MÉDIO OBTENDO 450 PONTOS NO ENEM.



FONTE: Elaborado pela autora a partir da ferramenta Online Pesquisa com os dados da pesquisa (2016).

A subseção a seguir trará os resultados da mineração de dados realizada na ferramenta Weka.

4.2 MINERAÇÃO DE DADOS DA COLETA DO QUESTIONÁRIO

Foi utilizado com confiança 0,9 e suporte 0,2 o algoritmo Apriori para minerar os resultados do questionário aplicado para os alunos de cursinhos. A FIGURA 12 apresenta os parâmetros utilizados nessa mineração. Com isso foram obtidas 15 regras distintas em 70 (setenta) instâncias analisadas pela ferramenta.

FIGURA 12 - PARÂMETROS DO APRIORI NO WEKA PARA MINERAR OS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIOS.

car	False
classIndex	-1
delta	0.05
doNotCheckCapabilities	False
lowerBoundMinSupport	0.2
metricType	Confidence
minMetric	0.9
numRules	100
outputItemSets	False
removeAllMissingCols	False
significanceLevel	-1.0
treatZeroAsMissing	False
upperBoundMinSupport	1.0
verbose	False

FONTE: Retirado do Weka 3.6.13.

Todas as regras descobertas apresentam um dos dois valores no consequente: (a) Idade igual a "A" ou seja, que o aluno está na faixa etária de 15 a 17 anos e (b) número de questões não é compatível com a prova, sendo 5(cinco) resultaram em "Idade A" e 10 (dez) "número de questões é compatível = não".

As regras que apresentaram "idade A" foram: (1) se o respondente considera o conhecimento adquirido foi suficiente e considera o ENEM como uma ferramenta de avaliação da qualidade do ensino médio; (2) se o conhecimento adquirido foi suficiente, o número de questões não for compatível e o aluno considerar o ENEM como ferramenta de avaliação da qualidade do ensino médio; (3) se o respondente for do gênero feminino e considera o preço da prova "justo"; (4) se o respondente considerar o preço da prova "justo" e considerar o ENEM como ferramenta de avaliação do ensino médio; (5) por fim, se o aluno for de escola particular e considerar

o ENEM como ferramenta de avaliação do ensino médio ele também estará nessa faixa etária.

As regras que tiveram “número de questões é compatível = não” foram: (1) se o gênero for masculino e se o aluno não considerar justa a certificação com 450 pontos; (2) se o respondente considerar suficiente o conhecimento e não considerar justa a certificação com 450 pontos; (3) se o respondente considerar “justo” o preço cobrado pela prova, considerar o ENEM como ferramenta de avaliação do ensino médio e não considerar justa a certificação com 450 pontos; (4) se o respondente considerar “justo” o preço da prova e não considerar o ENEM como ferramenta de avaliação do ensino médio; (5) se o tipo de escola for particular, e o respondente não considerar “justo” a certificação com 450 pontos; (6) se o tipo de escola for pública, o conhecimento adquirido não for suficiente e não considerar o ENEM como ferramenta de avaliação do ensino médio se o tempo de estudo for entre sete meses e um ano e o respondente não considerar “justo” a certificação com 450 pontos; (7) se o conhecimento adquirido não foi suficiente e o respondente não considerar o ENEM como uma ferramenta de avaliação do ensino médio; (8) se o respondente não considerar o ENEM como uma ferramenta de avaliação do ensino médio e não considerar justa a certificação com 450 pontos; (9) por fim se o tipo de escola for pública e o respondente não considerar o ENEM como uma ferramenta de avaliação do ensino médio. A TABELA 10 apresenta as regras encontradas pelo Weka.

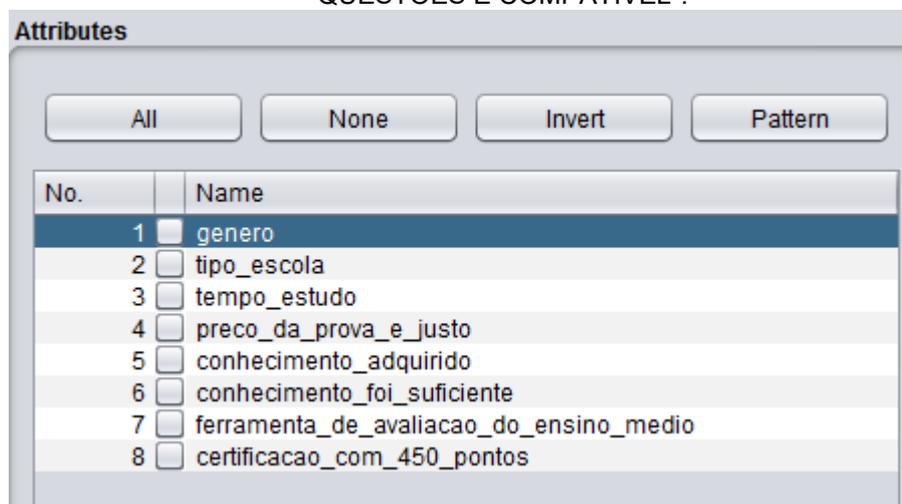
TABELA 10 - RESULTADOS DA COLETA DO QUESTIONÁRIO FEITO AOS ALUNOS DOS CURSOS PREPARATÓRIO POR MEIO DO ALGORITMO APRIORI EXECUTADO NO WEKA.

ID	Antecedente I	Antecedente II	Antecedente III	Consequente
1	conhecimento_foi_suficiente = sim	ferramenta_de_avaliacao_do_ensino_medio=sim (Sup 20)		idade=A (Sup 20)
2	genero=masculino	certificacao_com_450_pontos=nao (Sup 17)		numeros_de_questoes_e_compativel=nao (Sup 17)
3	conhecimento_foi_suficiente=sim	certificacao_com_450_pontos=nao (Sup 16)		numeros_de_questoes_e_compativel=nao (Sup 16)
4	preco_da_prova_e_justo=sim	ferramenta_de_avaliacao_do_ensino_medio=nao	certificacao_com_450_pontos=nao (Sup 16)	numeros_de_questoes_e_compativel=nao (Sup 16)
5	conhecimento_foi_suficiente=sim	numeros_de_questoes_e_compativel=nao	ferramenta_de_avaliacao_do_ensino_medio=sim (Sup 15)	idade=A (Sup 15)
6	genero=feminino	preco_da_prova_e_justo=sim	ferramenta_de_avaliacao_do_ensino_medio=sim (Sup 14)	idade=A (Sup 14)
7	preco_da_prova_e_justo=sim	ferramenta_de_avaliacao_do_ensino_medio=sim (Sup 20)		idade=A (Sup 19)
8	preco_da_prova_e_justo=sim	ferramenta_de_avaliacao_do_ensino_medio=nao (Sup 20)		numeros_de_questoes_e_compativel=nao (Sup 19)
9	tipo_escola=escolaparticular	certificacao_com_450_pontos=nao (Sup 16)		numeros_de_questoes_e_compativel=nao (Sup 15)
10	tipo_escola=escolapublica	conhecimento_foi_suficiente=nao	ferramenta_de_avaliacao_do_ensino_medio=nao (Sup 16)	numeros_de_questoes_e_compativel=nao (Sup 15)
11	tipo_escola=escolaparticular	ferramenta_de_avaliacao_do_ensino_medio=sim (Sup 15)		idade=A (Sup 14)
12	tempo_estudo=entresetemese seumano	certificacao_com_450_pontos=nao (Sup 15)		numeros_de_questoes_e_compativel=nao (Sup 14)
13	conhecimento_foi_suficiente=nao	ferramenta_de_avaliacao_do_ensino_medio=nao (Sup 24)		numeros_de_questoes_e_compativel=nao (Sup 22)
14	ferramenta_de_avaliacao_do_ensino_medio=nao	certificacao_com_450_pontos=nao (Sup 24)		numeros_de_questoes_e_compativel=nao (Sup 22)
15	tipo_escola=escolapublica	ferramenta_de_avaliacao_do_ensino_medio=nao (Sup 21)		numeros_de_questoes_e_compativel=nao (Sup 19)
Total de Regras		15		

FONTE: Algoritmo Apriori executado no Weka a partir dos dados da pesquisa (2016).

Devido à grande quantidade de respostas aos atributos “Idade A” e “número de questões é compatível = não” o conhecimento não foi surpreendente, conforme esperado. Por esse motivo, optou-se por retirar as colunas “idade” e “número de questões é compatível” do Weka. Foi retirado primeiramente o atributo “número de questões é compatível”, mantendo os parâmetros anteriores, porém foram obtidas as mesmas regras com o consequente “Idade A”. E se retirando o atributo “idade” os consequentes também seriam os mesmos que os anteriores. Com isso, optou-se por retirar os dois atributos ao mesmo tempo e manter apenas os atributos apresentados na FIGURA 13.

FIGURA 13 - ATRIBUTOS RESTANTES DEPOIS DE REMOVER “IDADE” E “NÚMERO DE QUESTÕES É COMPATÍVEL”.



FONTE: Imagem da base de dados da pesquisa retirada do Weka (2016).

Aplicando novamente o algoritmo Apriori agora com os parâmetros *default* (os parâmetros usados anteriormente não retornaram regras) foram obtidas 49 (quarenta e nove) regras, dessas 26 (vinte e seis) apresentaram confiança 100% e 23 (vinte e três) abaixo de 90%.

Analisando as 10 (dez) primeiras regras descobertas, foi identificado que elas apresentam um dos seis valores no consequente: (a) “gênero = feminino”; (b) certificação com 450 pontos = não; (c) ferramenta de avaliação do ensino médio = não; (d) conhecimento foi suficiente = não; (e) ferramenta de avaliação do ensino médio = sim; e (f) conhecimento foi suficiente = sim.

As dez primeiras instâncias obtidas apresentaram as seguintes regras: (1) se o preço da prova não é “justo”, se o conhecimento adquirido não for suficiente, e se o

respondente concordar com a certificação com 450 pontos, então esse respondente pertence ao gênero feminino; (2) se o tipo de escola for particular; se o respondente considerar o preço da prova “justo” e não considerar o ENEM uma ferramenta de avaliação do ensino médio, então o respondente não concorda com a certificação com 450 pontos; (3) se o tempo de estudo do respondente for mais de um ano então o então ele(a) não concorda com a certificação com 450 pontos; (4) se o tipo de escola for pública e o respondente não ter se preparado para a prova, então o conhecimento do respondente não foi suficiente; (5) se o tipo de escola for pública e o respondente não considerar “justo” o preço cobrado pela prova, então o respondente é do gênero feminino; (6) se o respondente for do gênero feminino e estudou entre sete meses e um ano, então o conhecimento adquirido foi bom; (7) se o tempo de estudo foi entre sete meses e um ano; se o conhecimento adquirido não foi suficiente e o respondente não considerar o ENEM uma ferramenta de avaliação do ensino médio então o respondente pertence ao gênero feminino; (8) se o respondente for de escola particular; se teve o conhecimento adquirido em sala de aula bom para realizar o ENEM e se o respondente considerar o ENEM como uma ferramenta de avaliação do ensino médio então o conhecimento que ele(a) possuía foi suficiente para realizar a prova; (9) se o respondente não considerar “justo” o preço da prova, se tiver o conhecimento adquirido insatisfatório e se o respondente não considerar o ENEM como uma ferramenta de avaliação do ensino médio então o conhecimento do respondente não foi suficiente para realizar a prova do ENEM; por fim (10) se o respondente for do gênero feminino e tiver o tempo de estudo mais de um ano, então o respondente não considera o ENEM como uma ferramenta de avaliação do ensino médio. Essas regras estão apresentadas na TABELA 11.

TABELA 11 - RESULTADOS APRESENTADOS PELO WEKA DOS DADOS COLETADOS DOS RESPONDENTES DOS CURSOS PREPARATÓRIOS APÓS RETIRAR OS ATRIBUTOS "IDADE" E " NÚMERO DE QUESTÕES É COMPATÍVEL".

ID	Antecedente I	Antecedente II	Antecedente III	Consequente
1	preco_da_prova_e_justo=nao	conhecimento_foi_suficiente=nao	certificacao_com_450_pontos=sim	genero=feminino
2	tipo_escola=escolaparticular	preco_da_prova_e_justo=sim	ferramenta_de_avaliacao_do_ensino_medio=nao	certificacao_com_450_pontos=nao
3	tempo_estudo=maisdeumano			ferramenta_de_avaliacao_do_ensino_medio=nao
4	tipo_escola=escolapublica	tempo_estudo=naomepreparei		conhecimento_foi_suficiente=nao
5	tipo_escola=escolapublica	preco_da_prova_e_justo=nao	certificacao_com_450_pontos=sim	genero=feminino
6	genero=feminino	tempo_estudo=entresetemesesehumano	conhecimento_adquirido=bom	ferramenta_de_avaliacao_do_ensino_medio=sim
7	tempo_estudo=entresetemesesehumano	conhecimento_foi_suficiente=nao	ferramenta_de_avaliacao_do_ensino_medio=sim	genero=feminino
8	tipo_escola=escolaparticular	conhecimento_adquirido=bom	ferramenta_de_avaliacao_do_ensino_medio=sim	conhecimento_foi_suficiente=sim
9	preco_da_prova_e_justo=nao	conhecimento_adquirido=insatisfatorio	ferramenta_de_avaliacao_do_ensino_medio=nao	conhecimento_foi_suficiente=nao
10	genero=feminino	tempo_estudo=maisdeumano		ferramenta_de_avaliacao_do_ensino_medio=nao
Total de regras		10 de 49 geradas		

FONTE: Algoritmo Apriori executado no Weka a partir dos dados da pesquisa (2016).

Com relação aos resultados da coleta do questionário aplicado aos alunos da UFPR foram utilizados no Apriori os parâmetros: confiança 0,9 e suporte 0,2 para minerar os dados. Com isso, diferentemente do obtido com os dados dos alunos dos cursos preparatórios, foram obtidas 9 (nove) regras apresentando um dos quatro consequentes: (a) ferramenta de avaliação = sim; (b) 450 pontos suficientes = não; (c) instrução suficiente = sim; e (d) gênero masculino.

As regras obtidas foram: (1) se gênero é masculino, e se o aluno concordar com a certificação do ensino médio com a obtenção de 450 pontos, então o respondente considera o ENEM uma ferramenta de avaliação do ensino médio; (2) se o curso do aluno for Gestão da Informação e se o respondente for do gênero feminino, então a aluna não concorda com a certificação do ensino médio; (3) se o respondente for do gênero feminino e estudou em escola particular então a respondente não concorda com a obtenção da certificação do ensino médio com 450 pontos; (4) se o curso do aluno for Gestão da Informação e se for do gênero masculino então o aluno considera o ENEM como ferramenta de avaliação do ensino médio; (5) se o respondente for do gênero feminino e se considerou a instrução oferecida em sala de aula suficiente para realizar o ENEM então o respondente não concorda com a certificação do ensino médio com 450 pontos; (6) se a idade do participante estiver dentro da faixa “A” e considerar o número de questões compatíveis com o tempo disponibilizado para a realização da prova então o respondente considerou a instrução oferecida em sala de aula suficiente para realização da prova do ENEM; (7) se o gênero do respondente for masculino e a faixa etária for entre 15 a 17 anos então o aluno considerou a instrução oferecida em sala de aula suficiente para realização da prova do ENEM; (8) se a faixa etária for “B”, se o respondente considerou o preço da prova “justo” e se o respondente considerar o ENEM uma ferramenta de avaliação do ensino médio então o respondente pertence ao gênero masculino; por fim (9) se o respondente não considerar o ENEM uma ferramenta de avaliação do ensino médio então o respondente não concorda com a certificação do ensino médio com a obtenção de 450 pontos. Essas regras podem ser visualizadas na TABELA 12.

TABELA 12 - RESULTADOS DA COLETA DO QUESTIONÁRIO FEITO AOS ALUNOS DA UFPR POR MEIO DO ALGORITMO APRIORI EXECUTADO NO WEKA.

ID	Antecedente I	Antecedente II	Antecedente III	Consequente
1	genero=masculino	450_pontos_suficientes=sim		ferramenta_de_avaliao=sim
2	curso=Gestao_da_Informacao_UFPR	genero=feminino		450_pontos_suficientes=nao
3	genero=feminino	tipo_de_escola=escola_particular		450_pontos_suficientes=nao
4	curso=Gestao_da_Informacao_UFPR	genero=masculino		ferramenta_de_avaliao=sim
5	genero=feminino	instrucao_suficiente=sim		450_pontos_suficientes=nao
6	idade=A	numero_de_questoes_compativel=sim		instrucao_suficiente=sim
7	genero=masculino	idade=A		instrucao_suficiente=sim
8	idade=B	preco_da_prova_justo=sim	ferramenta_de_avaliao=sim	genero=masculino
9	ferramenta_de_avaliao=nao			450_pontos_suficientes=nao
Total de regras		9		

FONTE: Algoritmo Apriori executado no Weka a partir dos dados da pesquisa (2016).

Outro algoritmo utilizado foi o J48 para poder visualizar a árvore de decisão. Ele foi utilizado no seu parâmetro *default* conforme apresentado na FIGURA 14. Com base nesses parâmetros foram obtidas 39 (trinta e nove) instâncias classificadas corretamente, cerca de 55% e 31 (tinta e uma) instâncias classificadas incorretamente, cerca de 44%.

FIGURA 14 - PARÂMETROS DEFAULT DO J48 NO WEKA PARA MINERAR OS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIOS.

weka.gui.GenericObjectEditor

weka.classifiers.trees.J48

batchSize	100
binarySplits	False
collapseTree	True
confidenceFactor	0.25
debug	False
doNotCheckCapabilities	False
doNotMakeSplitPointActualValue	False
minNumObj	2
numDecimalPlaces	2
numFolds	3
reducedErrorPruning	False
saveInstanceData	False
seed	1
subtreeRaising	True
unpruned	False
useLaplace	False
useMDLcorrection	True

FONTE: Retirado do Weka 3.6.13.

O atributo meta (ou ainda classe ou objetivo) utilizado para descobrir padrões foi o “certificação com 450 pontos”, o qual verifica se o aluno considera “justo” ou não a certificação do ensino médio obtendo 450 (quatrocentos e cinquenta) pontos na prova. Esse atributo foi escolhido por possuir a maior porcentagem de classificação corretas. Com base na matriz de confusão gerada pelo algoritmo, foi verificado que

ele obteve 29 acertos das respostas “não” e 10 acertos para as respostas “sim”. A matriz de confusão está apresentada na TABELA 13.

TABELA 13 - MATRIZ DE CONFUSÃO DO ALGORITMO J48 OBTIDO COM A MINERAÇÃO DOS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIO.

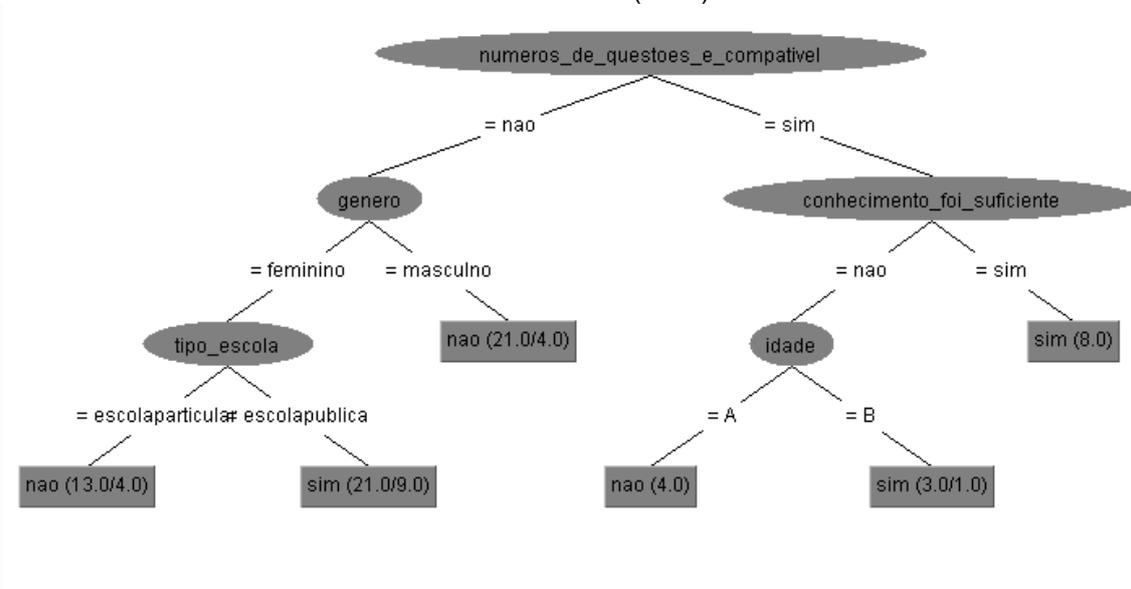
Matriz de confusão		
a	b	
29	11	a = não
20	10	b = sim
29 não classificados corretamente		
10 sim classificados corretamente		

FONTE: Retirado do Weka 3.6.13.

O algoritmo gerou uma árvore com base nas regras obtidas e o atributo da raiz que verifica se o número de questões é (ou não) compatível trouxe as seguintes regras: se o respondente for do gênero feminino, e tipo de escola particular então o respondente não está de acordo com a certificação do ensino médio obtendo 450 pontos; se o respondente for do gênero feminino, e tipo de escola pública então o respondente está de acordo com a certificação do ensino médio obtendo 450 pontos; por fim, se o gênero for masculino então o respondente não está de acordo com a certificação do ensino médio obtendo 450 pontos.

O atributo da raiz que testa se o número de questões é compatível trouxe as seguintes regras: se o conhecimento adquirido não for suficiente e a idade for “A”, então o respondente não está de acordo com a certificação do ensino médio obtendo 450 pontos; se o conhecimento adquirido não for suficiente e a idade for “B” então o respondente está de acordo com a certificação do ensino médio obtendo 450 pontos; por fim se o conhecimento obtido for suficiente então o respondente está de acordo com a certificação do ensino médio obtendo 450 pontos. A árvore gerada através do algoritmo J48 se encontra apresentada na FIGURA 15.

FIGURA 15 – ÁRVORE DE DECISÃO REPRESENTANDO OS DADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DOS CURSOS PREPARATÓRIO EXECUTADO PELO ALGORITMO J48 NO WEKA (2016).



FONTE: Criado pelo Weka com os resultados da pesquisa (2016).

Aplicando o algoritmo J48 nos dados da pesquisa coletada com os alunos da UFPR no seu padrão *default* foram retornadas 49 (quarenta e nove) instâncias classificadas corretamente, cerca de 71% e 20 (vinte) instâncias classificadas incorretamente, cerca de 28%. Foram analisadas 69 (sessenta e nove) instâncias.

O atributo meta (ou ainda classe ou objetivo) utilizado para descobrir padrões foi “tipo de escola”, esse atributo verifica se o aluno que realizou a prova do ENEM cursou o ensino médio em escola pública ou particular. A partir disso, o algoritmo retornou uma matriz de confusão contendo 24 acertos as respostas “escola particular” e 25 acertos a respostas “escola pública”. A matriz de confusão está apresentada na TABELA 14.

TABELA 14 - MATRIZ DE CONFUSÃO DO ALGORITMO J48 OBTIDO COM A MINERAÇÃO DOS RESULTADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DA UFPR.

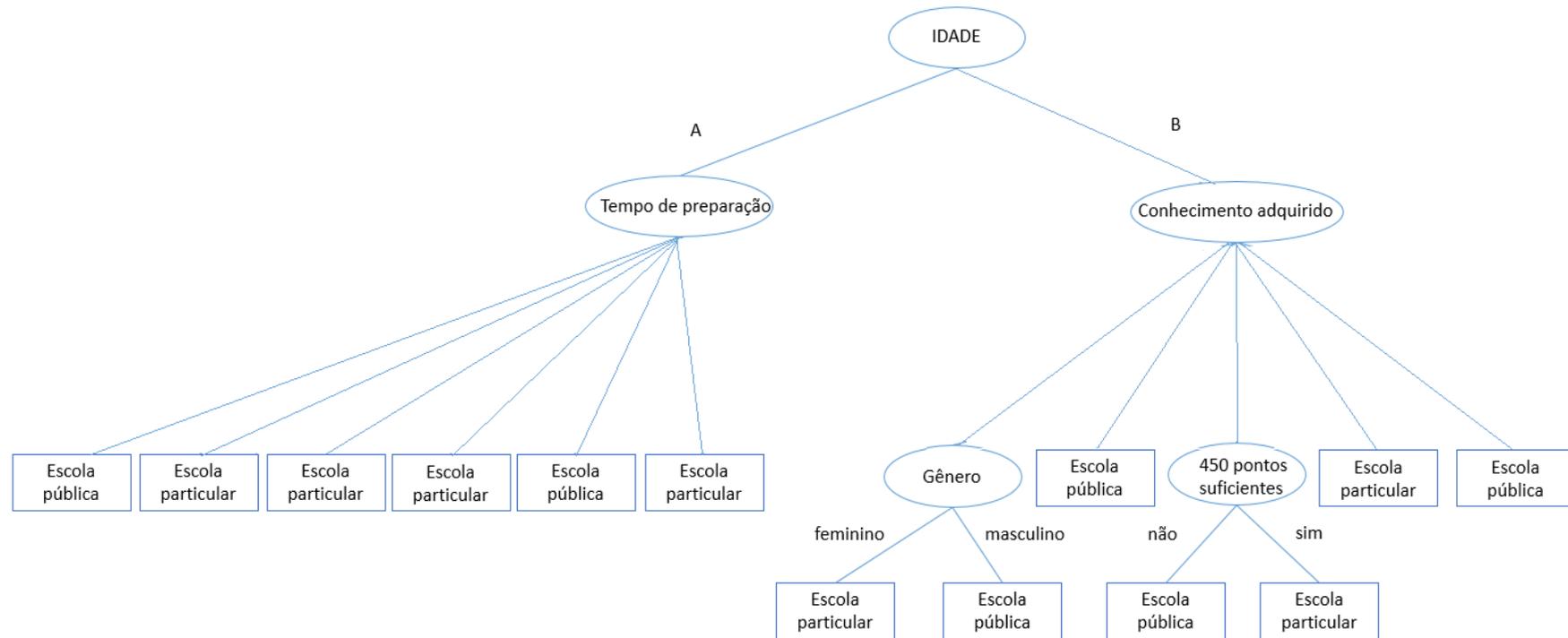
Matriz de confusão		
a	b	
24	8	a = escola particular
12	25	b = escola pública
24 “escola particular” classificados corretamente		
25 “escola pública” classificados corretamente		

FONTE: Retirado do Weka 3.6.13.

O algoritmo J48 com os dados da pesquisa aos alunos da UFPR conseguiu elaborar uma árvore com base nas regras obtidas, o atributo da raiz que testa se a idade do respondente corresponde a faixa “A” apresentou as seguintes regras: (1) se tempo de preparação for entre quatro e seis meses então o aluno estudou em escola pública; (2) se tempo de preparação for entre sete meses e um ano então o aluno estudou em escola particular; (3) se tempo de preparação for mais de dois anos, então o aluno estudou em escola particular; (4) se tempo de preparação for mais de um ano, então o aluno estudou em escola particular; (5) se tempo de preparação for menos de três meses, então o aluno estudou em escola pública; por fim (6) se o respondente não se preparou então ele(a) estudou em escola particular.

O atributo da raiz que testa se a idade do respondente corresponde a faixa “B” apresentou as seguintes regras: (1) se o conhecimento adquirido foi bom para realizar o ENEM e se o respondente for do gênero feminino então o respondente estudou em escola particular; (2) se o conhecimento adquirido foi bom para realizar o ENEM e se o respondente for do gênero masculino então o respondente estudou em escola pública; (3) se o conhecimento adquirido foi insatisfatório para realizar o ENEM então o aluno estudou em escola pública; (4) se o conhecimento adquirido foi muito bom para realizar o ENEM e o aluno não concordar com a certificação do ensino médio com a obtenção de 450 pontos então o aluno é de escola pública; (5) se o conhecimento adquirido foi muito bom para realizar o ENEM e o aluno concordar com a certificação do ensino médio com a obtenção de 450 pontos então o aluno é de escola particular; (6) se o conhecimento adquirido foi ótimo então o aluno é de escola pública; por fim (7) se o conhecimento adquirido foi regular então o aluno é de escola pública. Essas regras estão apresentadas na árvore de decisão mostrada na FIGURA 16.

FIGURA 16 – ÁRVORE DE DECISÃO REPRESENTANDO OS DADOS DO QUESTIONÁRIO APLICADO AOS ALUNOS DA UFPR EXECUTADO PELO ALGORITMO J48 NO WEKA (2016).



FONTE: Elaborado pela autora com os dados da pesquisa (2016).

A seção a seguir trará os resultados da coleta de opinião das redes sociais analisados a partir da ferramenta *Semantria*.

4.3 ANÁLISE DE SENTIMENTO DOS DADOS COLETADOS DAS REDES SOCIAIS

Foram coletados 100 (cem) sentenças, dessas, 50 (cinquenta) sentenças de cada rede social: Twitter e Facebook, os dados dessa coleta está disponível no APÊNDICE D – BASE DE DADOS “OPINIÕES ENEM”. Das sentenças coletadas e analisadas na ferramenta *Semantria*, foram encontradas 16 (dezesesseis) negativas, dessas 9 (nove) eram da rede social Facebook (56%) e 7 (sete) da rede social Twitter (44%). Das sentenças neutras, a ferramenta localizou 52 (cinquenta e duas), dessas 23 (vinte e três) pertencem ao Facebook (44%) e 29 (vinte e nove) ao Twitter (56%). As sentenças positivas totalizaram 32 (trinta e duas), sendo 18 (dezoito) coletadas do Facebook (56%) e 14 (catorze) do Twitter (44%). Esses dados estão apresentados na TABELA 15.

TABELA 15 – CONTAGEM DE SENTENÇAS NEGATIVAS, NEUTRAS E POSITIVAS DAS REDES SOCIAIS TWITTER E FACEBOOK.

Rede social	Tipo de sentimento						Total Geral
	Negativo	%	Neutro	%	Positivo	%	
Facebook	9	56%	23	44%	18	56%	50
Twitter	7	44%	29	56%	14	44%	50
Total	16	100%	52	100%	32	100%	100

FONTE: Elaborado pela autora com dados da pesquisa (2016).

Para facilitar o entendimento desses números, os quadros a seguir apresentam as principais palavras categorizadas pela ferramenta como “positivas”, “negativas” e “neutro e se a análise realizada pela ferramenta está de acordo com a mensagem repassada. Essas palavras influenciaram na classificação da sentença. A TABELA 16 apresenta as principais palavras categorizadas como positivas e se a análise da ferramenta está correta ou incorreta.

TABELA 16 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO POSITIVAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.

Palavras	Análise
"Correta"	Classificação incorreta, a pessoa participante na sentença não se referia a palavra correta como algo positivo, mas sim como reclamação para que os funcionários repassassem as informações corretamente.
"Tranquilidade"	Classificação correta.
"Não ruim"	Classificação incorreta, a pessoa participante na sentença utilizou de ironia
"Acertar"	Classificação correta.
"três ultimos"	Classificação incorreta, a pessoa participante na sentença utilizou de ironia
"Fácil"	Classificação correta.
"Ganhem"	Classificação incorreta. A pessoa participante na sentença descreveu que gostaria da anulação da prova do ENEM o que pode categorizar como uma sentença negativa devido a palavra "anulação".
"Alunos loco"	Classificação incorreta, utilização de gírias na sentença.
"Sonho"	Classificação incorreta, utilização de sacarmos na sentença.
"Bem"	Classificação incorreta, a palavra "bem" interpretada pela ferramenta remete algo positivo, porém nesse caso é ao desempenho dos participantes na prova, no caso, remete a algo negativo informando que "não foram bem" na sentença.
"Descente"	Classificação incorreta, a pessoa participante na sentença utilizou a palavra como forma de protesto para que os professores tenham uma remuneração mais alta.
"Remunera professor"	Classificação incorreta, a pessoa participante na sentença utilizou a palavra como forma de protesto para que os professores tenham uma remuneração mais alta.
"Funcional"	Classificação incorreta, a pessoa participante na sentença utilizou de ironia.
"Tranquila"	Classificação correta.
"Assim eu"	Classificação correta.
"Resolvi todas"	Classificação correta.
"Querida"	Classificação incorreta, a pessoa participante na sentença utilizou de ironia se referindo a ex-presidente Dilma.
"Qualidade"	Classificação incorreta, a pessoa participante na sentença utilizou de ironia.
"Lógico"	Classificação incorreta, a pessoa participante na sentença utilizou de ironia.
"Justa"	Classificação correta.
"Oportunidade"	Classificação correta.
"Ótima"	Classificação incorreta, neste caso a pessoa participante na sentença escreveu a sentença com duplo sentido.
"Feliz"	Classificação correta.
"Acertei"	Classificação correta.
Total de palavras	24
Classificadas incorretamente	14
Classificadas corretamente	10

FONTE: Elaborado pela autora com os dados da pesquisa (2016).

Das 24 (vinte e quatro) palavras 14 (catorze), cerca de 58% (cinquenta e oito por cento), apresentaram inconsistência na análise realizada pela ferramenta. Nesses casos, a ferramenta não conseguiu identificar a utilização de: gírias, frases no sentido irônico, e frases sarcásticas, classificando-as incorretamente. A TABELA 17 apresenta as principais palavras categorizadas como negativas e se a análise da ferramenta está correta ou incorreta.

TABELA 17 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEGATIVAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.

continua

Palavras	Análise
“(“	Classificação correta.
“Absurdo“	Classificação correta.
“Bem“	Classificação correta.
“Carência“	Classificação incorreta, na sentença o participante demonstra uma dúvida. Poderia estar classificada como neutro.
“Difícil“	Classificação correta.
“Ensino“	Classificação correta.
“Errei“	Classificação incorreta, na sentença o participante demonstra uma dúvida. Poderia estar classificada como neutro.
“Erro“	Classificação correta.
“Escrúpulos“	Classificação correta.
“Essa prova vo aguarda“	Classificação incorreta, o participante na sentença não demonstrou sentimento na frase, apenas ofereceu uma informação. Poderia estar classificada como neutro.
“Falha“	Classificação incorreta, na sentença o participante demonstra uma dúvida. Poderia estar classificada como neutro
“Ganhei“	Classificação correta.
“Horível“	Classificação correta.
“Impossível“	Classificação correta.
“Inadmissível“	Classificação correta.
“Inesperada“	Classificação incorreta, na sentença o participante demonstra uma dúvida. Poderia estar classificada como neutro.
“Irreversível“	Classificação correta.
“Lixo“	Classificação correta.
“Mesmo assim“	Classificação correta.
“Morri“	Classificação correta.
“Não consegui“	Classificação incorreta, na sentença a pessoas participante demonstra uma dúvida. Poderia estar classificada como neutro.
“Não consigo“	Classificação incorreta, na sentença a pessoas participante demonstra uma dúvida. Poderia estar classificada como neutro.
“Não ter“	Classificação correta.
“Não tinha“	Classificação correta.
“Negativa“	Classificação incorreta, na sentença a pessoas participante demonstra uma dúvida. Poderia estar classificada como neutro.
“Nojenta“	Classificação correta.
“Obrigada senhor“	Classificação incorreta, frase é positiva. O participante na sentença descreveu que ganhou uma boa nota, não há sentimento negativo.

TABELA 17 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEGATIVAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.

conclusão

Palavras	Análise
“Organizada“	Classificação correta.
“Pagina redireciona“	Classificação incorreta, na sentença a pessoas participante apenas demonstra uma dúvida. Poderia estar classificada como neutro.
“Palhaçada“	Classificação correta.
“Pessoa estuda“	Classificação correta.
“Piorar“	Classificação incorreta, na sentença a pessoas participante apenas demonstra uma dúvida. Poderia estar classificada como neutro.
“Positiva“	Classificação incorreta, na sentença a pessoas participante apenas demonstra uma dúvida. Poderia estar classificada como neutro.
“Prejudicada“	Classificação correta.
“Problema“	Classificação correta.
“Prova tava dificil demais“	Classificação correta.
“Qualidade“	Classificação correta.
“Quebrando“	Classificação correta.
“Questionário enorme“	Classificação correta.
“Ruim“	Classificação correta.
“Soubesse“	Classificação correta.
“Suportar“	Classificação correta.
“Tanta crítica“	Classificação incorreta, a pessoa participante na sentença utilizou gírias, mas quis repassar que gostou da prova.
“Taxa“	Classificação incorreta, na sentença o participante apenas demonstra uma dúvida. Poderia estar classificada como neutro.
“Tempo“	Classificação correta.
“Vingou“	Classificação correta.
“Violência“	Classificação incorreta. A pessoa respondente na sentença salientou o tema da redação do ENEM 2015, no final da sentença está a palavra “Amei”.
“Virava mulher“	Classificação correta.
Total de palavras	48
Classificadas incorretamente	15
Classificadas corretamente	33

FONTE: Elaborado pela autora com os dados da pesquisa (2016).

Das 48 (quarenta e oito) palavras 15 (quinze), cerca de 31% (trinta e um por cento), apresentaram inconsistência na análise realizada pela ferramenta. Isso ocorreu com sentenças na qual a pessoa participante demonstrou dúvida, ou utilizou de gírias. Houve apenas um caso em que o respondente se referiu ao tema da prova do ENEM 2015: “a persistência da violência contra a mulher na sociedade brasileira”, na qual comentou que “amou” o tema, mas devido a palavra violência a ferramenta classificou como negativa. A TABELA 18 apresenta as principais palavras categorizadas como neutras e se a análise da ferramenta está correta ou incorreta.

TABELA 18 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEUTRAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.

continua

Frases	Análise
"Não faz sentido"	Classificação correta.
"Minha inscrição"	Classificação correta.
"Sinceramente"	Classificação incorreta. O respondente demonstra frustração com o sistema de inscrição.
"Falhas"	Classificação incorreta. O respondente demonstra frustração com o sistema de inscrição. Deveria ser uma palavra negativa.
"Saco cheio"	Classificação incorreta. O respondente demonstra frustração com o sistema de inscrição. Deveria ser uma palavra negativa.
"Não demorei"	Classificação correta.
"Mais escrota"	Classificação incorreta, o participante na sentença utilizou gírias. Deveria ser classificada como uma palavra negativa.
"Gostaria"	Classificação correta.
"Errado"	Classificação correta.
"Ajuda"	Classificação correta.
"Algum problema"	Classificação correta.
"Certo seria"	Classificação correta.
"Logo pra nao ter"	Classificação correta.
"Não bastasse"	Classificação correta.
"Reclamam"	Classificação incorreta. O respondente na sentença demonstra frustração com a escrita dos participantes do ENEM. Deveria ser uma palavra negativa.
"Realmente fico chocada"	Classificação incorreta. O respondente na sentença demonstra frustração com a escrita dos participantes do ENEM. Deveria ser uma palavra negativa.
"Bem"	Classificação correta.
"Não achei"	Classificação correta.
"Sou tão"	Classificação correta.
"Gramática todos"	Classificação correta.
"Vence"	Classificação correta.
"Cansaço"	Classificação correta.
"Ótima idéia"	Classificação incorreta. O respondente na sentença demonstra frustração com nível de dificuldade da prova. Deveria ser uma palavra negativa.
"Prova somos massacrados"	Classificação incorreta. O respondente na sentença demonstra frustração com nível de dificuldade da prova. Deveria ser uma palavra negativa.
"Parabéns"	Classificação incorreta. A pessoa respondente na sentença salientou o tema da redação do ENEM 2015, parabenizando o tema.

TABELA 18 – PRINCIPAIS PALAVRAS CATEGORIZADAS PELO SEMANTRIA COMO NEUTRAS E SE A ANÁLISE DA FERRAMENTA ESTÁ CORRETA OU INCORRETA.

		conclusão
Frases	Análise	
"Espancadas"	Classificação incorreta. A pessoa respondente na sentença salientou o tema da redação do ENEM 2015, parabenizando o tema.	
"Estupradas"	Classificação incorreta. A pessoa respondente na sentença salientou o tema da redação do ENEM 2015, parabenizando o tema.	
"Rebaixadas profissionalmente"	Classificação incorreta. A pessoa respondente na sentença salientou o tema da redação do ENEM 2015, parabenizando o tema.	
"Aquele prova"	Classificação incorreta. A pessoa respondente na sentença salientou o tema da redação do ENEM 2015, parabenizando o tema.	
Total de palavras	29	
Classificadas incorretamente	13	
Classificadas corretamente	16	

FONTE: Elaborado pela autora com os dados da pesquisa (2016).

Das 29 (vinte e nove) palavras 13 (treze), cerca de 45% (quarenta e cinco por cento), apresentaram inconsistência na análise realizada pela ferramenta. Essas inconsistências surgiram em sentenças que apresentavam gírias e houve casos em que o respondente se referiu ao tema da prova do ENEM 2015 assim como na TABELA 17.

Pode se observar que a maior parte das sentenças coletadas possui alguma palavra categorizada pela ferramenta como negativa. Isso demonstra que, dos dados coletados grande parte dos realizadores do ENEM não estavam satisfeitos com algum critério (dos observados nos comentários e *tweets*): tempo de prova, questões cobradas, site de inscrição, resultado da prova, nível de dificuldade da prova.

4.4 METODOLOGIA DE ANÁLISE DE OPINIÕES SOBRE A PROVA DO ENEM

Para resolver o problema de pesquisa proposto nesse projeto de conclusão de curso essa subseção trará os passos para elaborar uma metodologia de análise de opiniões sobre a prova do ENEM.

O primeiro passo realizado foi a definição do público alvo e amostra para realizar a coleta de opiniões. Nesse projeto foram definidos três públicos: alunos de cursos preparatórios; alunos da Universidade Federal do Paraná – Setor de Ciências

Sociais Aplicadas que realizaram a prova do ENEM; e usuários das páginas do Facebook e *tweets* do Twitter para coletar opiniões.

O segundo passo da metodologia é a criação de um instrumento para coletar opiniões dos alunos dos cursinhos e da UFPR. Esse questionário deve levar em consideração que as questões realizadas passarão por uma mineração de dados para busca de padrões. Por esse motivo as questões realizadas devem procurar definir o perfil do respondente e buscar forma de coletar a opinião do respondente sem manipulá-lo, aliená-lo ou confundi-lo. Questões abertas são difíceis de minerar por não possuir padrões, por esse motivo a utilização de questões fechadas ou questões de escalas para possibilitar a busca de padrões na mineração de dados.

O terceiro passo é realizar uma pesquisa para localizar potenciais ferramentas de análise de sentimento ou mineração de opinião e ferramentas de mineração de dados. Para essa pesquisa foram utilizados o Semantria e o Weka, mas existem outras como Text2data e o RapidMiner que poderiam ser usadas.

O quarto passo é estruturar os dados da coleta para que possa ser analisado nas ferramentas de análise de sentimento ou mineração de opinião e ferramentas de mineração de dados. Por exemplo, no Semantria era necessário que a base de dados fosse elaborada no Excel da Microsoft. E para utilizar o Weka foi necessário retirar acentos, espaços, discretizar o atributo idade e converter a base criada no Excel para o formato *.arff* para poder abri-la.

O quinto passo é referente à mineração de dados, nesse ponto é necessário ter em vista qual ou quais algoritmos utilizar para minerar a base de dados. É necessário abrir a base de dados a ferramenta antes de procurar um algoritmo para descobrir quais algoritmos estão habilitados e quais estão disponíveis para uso.

O sexto passo corresponde a processar as bases criadas nas ferramentas escolhidas. A partir disso procurar formas de apresentar o conteúdo gerado nas ferramentas, como por exemplo: tabelas, imagens ou gráficos conforme utilizado nesse estudo.

Por fim, o sétimo passo é apresentar os resultados das opiniões dos realizados do ENEM, conforme a definição dos objetivos do estudo. Nesse passo serão apresentados os padrões de regras e conhecimentos adquiridos com a mineração de dados e as opiniões analisadas.

Na FIGURA 17 estão apresentadas as etapas descritas acima da metodologia proposta. O processo de retroalimentação acontece a partir da segunda etapa, ao criar

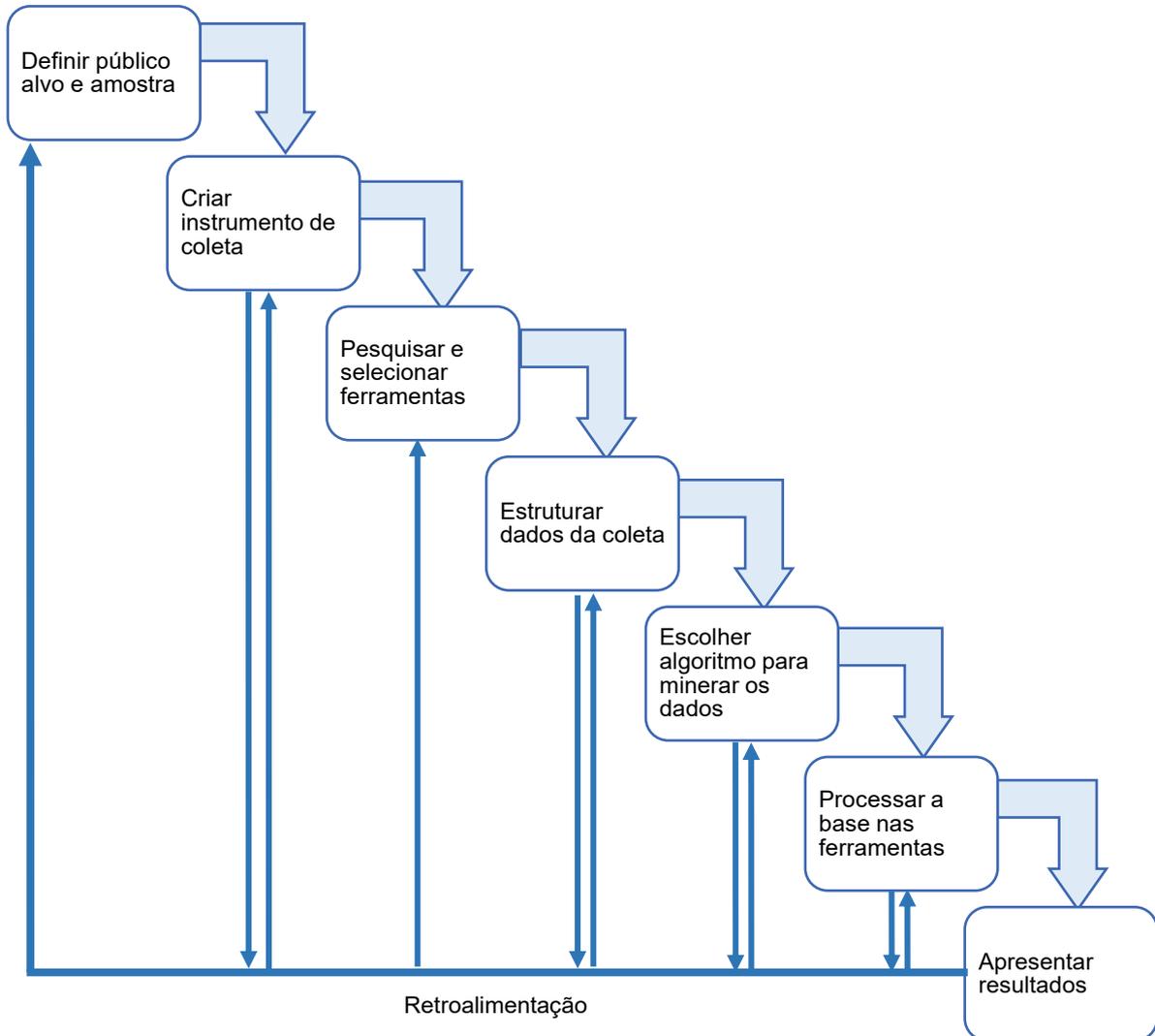
um instrumento de coleta deve ter definido o público alvo da pesquisa se for necessário recriar o instrumento de coleta, então deverá ser reavaliado o público alvo e a amostra desejada e adaptar o novo instrumento.

Na etapa “estruturar dado da coleta” se os dados coletados não forem suficientes para prosseguir com a próxima etapa, nesse caso talvez seja necessário retornar ao processo de “criar instrumento de coleta” e reavaliar o instrumento para que atinja um público maior ou voltar a definir o público para que este englobe mais pessoas.

Na etapa seguinte, em “escolher algoritmo para minerar os dados” se os dados não forem estruturados corretamente, a etapa deverá ser realizada novamente até que possa ser minerado. Se depois de estruturado os dados e após escolher o algoritmo de mineração não for possível processar a base nas ferramentas escolhidas, então deverá ser retornado para as etapas anteriores onde será possível escolher outro algoritmo, reestruturar os dados ou escolher outras ferramentas para processar a base. Após apresentar os resultados, o processo pode retornar para a primeira etapa caso o pesquisador deseje executar a metodologia novamente.

Para que não seja necessário voltar as etapas é necessário realizar um estudo do instrumento de coleta, ferramentas e algoritmos escolhidos para que estes atendam a necessidade do objetivo do estudo.

FIGURA 17 - FLUXO DA METODOLOGIA PARA ANALISAR SENTIMENTOS E MINERAR OPINIÕES.



FONTE: Elaborado pela autora (2016).

Para poder obter padrões de conhecimento significativos é importante atingir a amostra desejada, se isso não ocorrer, procurar outro público para o estudo. Neste estudo, as amostras não foram 100% atingidas e as regras da mineração de dados não trouxeram resultados significantes devido ao número de respondentes algumas questões ficaram com pouca distinção de respostas, como exemplo a questão que questionava se o aluno teve problema com o sistema de inscrição e mais de 80% responderam que não, e por isso tiveram que ser desconsideradas da mineração de dados.

5 CONSIDERAÇÕES FINAIS

As opiniões são formas de expressar determinado sentimento sobre determinado produto ou acontecimento. Elas podem auxiliar a medir o grau de satisfação do cliente ou usuário a partir do que ele expressa por meio de frases positivas, neutras ou negativas. A internet é um ambiente que compreende uma gama grande de opiniões de usuário em páginas na *web* como as redes sociais. Atualmente existem ferramentas que auxiliam na recuperação de opiniões dos usuários além da utilizada nesta pesquisa. Porém as ferramentas possuem recursos limitados para contas gratuitas, como realizar busca em um determinado período. Esse foi um fator que impediu que essa ferramenta fosse usada nesse estudo por não trazer resultados de cinco meses atrás, apenas resultados recentes.

É difícil encontrar uma pesquisa que apresente formas de criar bases de dados com opiniões de usuários e aplicações de ferramentas e algoritmos para permitir a descoberta de padrões e como consequência conhecimento. Essa foi uma justificativa que motivou a realização deste estudo, por isso o principal objetivo era de propor uma metodologia de análise de opiniões sobre a prova do ENEM. Além disso, esse trabalho tornou-se uma contribuição para os interessados em realizar estudos de mineração de dados. A partir da metodologia aqui proposta o pesquisador poderá replicá-la em seus estudos.

5.1 VERIFICAÇÃO DOS OBJETIVOS PROPOSTOS

Para atingir o objetivo específico de prospectar estudos relacionados ao tema da pesquisa foram realizados estudos na área que envolvessem o tema: Mineração de Dados, Mineração de Opinião e o próprio objeto de estudo: o ENEM que apresentados no decorrer deste documento. Essa etapa foi atingida por meio da pesquisa exploratória. Outro objetivo era de definir uma metodologia para criação desta base de dados na qual foram apresentados sete passos a serem seguidos, a considerar: (1) definir público alvo e amostra; (2) criar instrumento de coleta; (3) pesquisar e selecionar ferramentas; (4) estruturar dados da coleta; (5) escolher algoritmo para minerar os dados; (6) processar a base nas ferramentas; por fim (7) apresentar resultados. A partir desses passos, é possível replicar essa metodologia em outros estudos.

Para registrar as opiniões e os dados coletados referente aos estudantes dos cursos preparatórios, alunos do Setor de Ciências Sociais Aplicadas e usuários do Facebook e Twitter foi realizada a criação de três bases de dados, uma a primeira foi composta com as opiniões dos alunos e ex-alunos de cursos preparatórios, a segunda foi composta com as opiniões dos alunos matriculados nos cursos do Setor de Ciências Sociais da UFPR e a terceira com opiniões coletadas das redes sociais: *Facebook* e *Twitter* parte dessa metodologia foi baseada em estudos anteriores na área da mineração de opinião.

Como forma de descobrir padrões de opiniões entre alunos e usuários do Twitter e Facebook pelo uso de métodos e técnicas de mineração de dados foi utilizada a ferramenta Weka para aplicar os algoritmos de associação: Apriori e de árvore de decisão: J48. Com base nisso foram encontradas regras e apresentadas em tabelas e na árvore de decisão gerada. As regras apresentadas trouxeram uma tendência devido à grande quantidade de resultados a apenas uma das respostas.

Devido à amostra desta pesquisa não ter sido atingida, houveram respostas com tendenciosidade para a mineração, por exemplo, no questionário aplicado aos alunos da UFPR foram obtidas apenas 4 (quatro) “sim” para a questão “você já teve problemas com o sistema de inscrição”, esse dado levou o algoritmo a trazer regras com essa tendência e por esse motivo foi necessário retirar esse atributo da mineração pois o conhecimento adquirido não foi tão surpreendente.

Com relação à coleta de sentenças das redes sociais pode analisar que essas sentenças tanto do Facebook quanto do Twitter apresentaram em sua grande parte negatividade dos usuários em relação: tempo de prova, questões cobradas, site de inscrição, resultado da prova, nível de dificuldade da prova. As sentenças que apresentaram sentimentos positivos estavam relacionadas ao tema da redação do ENEM 2015: “a persistência da violência contra a mulher na sociedade brasileira”.

5.2 TRABALHOS FUTUROS

Uma das propostas deste estudo é que ele possa ser replicado em outras pesquisas, por isso, sugere-se expandir este estudo para outras cidades/regiões do país para realizar comparações e análises de opiniões tornando possível comparar qualidade de ensino em diferentes regiões do país a partir do desempenho dos alunos.

E além disso, sugere-se aplicá-lo em outras redes sociais além das duas aqui consideradas.

Como sugestão de trabalhos futuros seria analisar outras edições do ENEM, por exemplo do ENEM 2016 para avaliar se os realizadores continuam em sua grande parte descontentes ou se algum dos processos foram melhorados.

5.3 CONTRIBUIÇÕES DA PESQUISA

Essa pesquisa é uma contribuição para os pesquisadores que desejam trabalhar na área de Mineração de Opinião por apresentar as etapas para seguir um estudo. É uma contribuição para a ciência por reunir os referenciais da área da Mineração de Dados e Mineração de Opinião.

E essa pesquisa contribui para outros pesquisadores interessados em estudar sobre o Exame Nacional do Ensino Médio, por apresentar resultados referentes a prova como nível de dificuldade, número de questões, tempo de prova. E os comentários aqui disponibilizados podem ser utilizados em outros estudos sem necessariamente levar o pesquisador a realizar uma coleta de dados.

REFERÊNCIAS

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

BARBOSA, R. Gestão da informação e do conhecimento: origens, polêmicas e perspectivas. **Inf. Inf.**, Londrina, v. 13, n. esp, p. 1-25, 2008. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1843/1556>>. Acesso em: 23 abr. 2016.

BECKER, K.; TUMITAN, D. Introdução à Mineração de Opiniões: conceitos, aplicações e desafios. **Lectures of the 28th Brazilian Symposium on Databases**, Pernambuco, n. 1, 2013. 27-52. Disponível em: <http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd_versaosubmetida.pdf>. Acesso em: 6 abr. 2016.

BERLIM, J. B. **Análise de ferramentas de mineração de opinião para aplicação em redes sociais**. TCC (Graduação) - Curso de Gestão da Informação, Universidade Federal do Paraná. Curitiba, p. 147. 2015.

BRÄSCHER, M. A ambiguidade na recuperação da informação. **DataGramZero**, v. 3, n. 1, Fevereiro 2002. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/284/1/BrascherDTZ2002.pdf>>. Acesso em: 31 mai. 2016.

BRASIL. ENEM por Escola. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira**, 2011. Disponível em: <<http://portal.inep.gov.br/web/enem/enem-por-escola>>. Acesso em: 22 mai.2016.

BRILHADORI, A; LAURETTO, M. S. Estudo comparativo entre algoritmos de árvores de classificação e máquinas de vetores suporte, baseados em ensembles de classificadores. **Escola de Artes, Ciências e Humanidades - Universidade de São Paulo (USP)**, São Paulo, p. 12, 2013. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2013/009.pdf>>. Acesso em: 02 nov. 2016.

BROWN, M. S. **Data mining for dummies**. 1. ed. New Jersey: John Wiley & Sons, 2014.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**, Berkeley, v. 45, n. 5, p. 351-360, 1991. Disponível em: <<http://inls151f14.web.unc.edu/files/2014/08/buckland1991-informationasthing.pdf>>. Acesso em: 23 abr. 2016.

BUSH, V. As we may think. **The Atlantic Monthly**, Boston, v. 176, n. 1, p. 101-108, July 1945. Disponível em: <[http://worrydream.com/refs/Bush%20-%20As%20We%20May%20Think%20\(Life%20Magazine%209-10-1945\).pdf](http://worrydream.com/refs/Bush%20-%20As%20We%20May%20Think%20(Life%20Magazine%209-10-1945).pdf)>. Acesso em: 23 abr. 2016.

CARDOSO, O. N. P. Recuperação da informação. **Infocomp**, Olinda, v. 2, 2000. ISSN 1. Disponível em:

<<http://www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/46/31>>. Acesso em: 31 mai. 2016.

CASTRO, L. N. D.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. 1^a. ed. São Paulo: Saraiva, 2016.

CASTRO, M. H. G. D.; TIEZZI, S. A reforma do ensino médio e a implantação do ENEM no Brasil. In: BROCK, C.; SCHWARTZMAN, S. Os desafios da educação no Brasil. Rio de Janeiro: **Nova Fronteira**, 2005. p. 119-152. Disponível em: <<http://www.schwartzman.org.br/simon/desafios/4ensinomedio.pdf>>. Acesso em: 22 mai. 2016.

DANTE, G. P. **Gestion de información en las organizaciones: principios, conceptos y aplicaciones**. 1. ed. Santiago, 1998.

DAVENPORT, T. H. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação**. Tradução de Bernadette Siqueira Abrão. São Paulo: Futura, 1998. Disponível em: <<http://amormino.com.br/livros/20141114-ecologia-informacao.pdf>>. Acesso em: 23 abr. 2016.

ELLISON, N. B. Social network sites: Definition, history, and scholarship. 1. ed. **Journal of Computer-Mediated Communication**, v. 13, 2007. 210-230 p. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full>>. Acesso em: 25 set. 2016.

ESULI, A.; SEBASTIANI, F. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. **Proceedings of the Fifth International Conference on Language Resources and Evaluati**, Genova, p. 417-422, 2006.

FACEBOOK. **Sobre**. Disponível em: <<https://www.facebook.com/pg/FacebookBrasil/about/>>. Acesso em 12 nov. 2016.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. 3. ed. **AI Magazine**, v. 17, 1996. Disponível em: <<http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>>. Acesso em: 18 mai. 2016.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação da informação. **Ci. Inf**, Brasília, v. 35, p. 25-30, Janeiro 2006. ISSN 1. Disponível em: <<http://www.scielo.br/pdf/%0D/ci/v35n1/v35n1a03.pdf>>. Acesso em: 31 mai. 2016.

FERREIRA, E. D. B. A. **Análise de sentimento em redes sociais utilizando influência das palavras**. Recife: Tese (Graduação em Ciência da Computação) - Centro de informática, Universidade Federal de Pernambuco, 2010. Disponível em: <<http://www.cin.ufpe.br/~tg/2010-2/ebaf.pdf>>. Acesso em: 18 mai. 2016.

G1. **Facebook completa 10 anos: veja a evolução da rede social**. **G1**, 2014. Disponível em: <<http://g1.globo.com/>>. Acesso em: 16 out. 2016.

G1. **ENEM 2016: inscrição começa nesta segunda e vai até dia 20**. Educação, São Paulo, 2016. Disponível em:

<<http://g1.globo.com/educacao/enem/2016/noticia/enem-2016-inscricoes-comecam-nesta-segunda-8.ghtml>>. Acesso em: 07 jun. 2016.

GALVÃO, N. D.; MARIN, H. D. F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paul Enferm**, São Paulo, v. 22, n. 5, p. 686-690, 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-21002009000500014>. Acesso em: 22 mai. 2016.

GODOY, A. S. Introdução à pesquisa qualitativa e suas possibilidades. **Revista de Administração de Empresas**, São Paulo, v. 35, p. 57-63, Mar./Abr. 1995. ISSN 2. Disponível em: <<http://www.scielo.br/pdf/rae/v35n2/a08v35n2.pdf>>. Acesso em: 28 mai. 2016.

GOLDSCHMIDT, R.; PASSOS, E. Data mining: um guia prático. **Gulf Professional Publishing**, 2005. 256 p.

HAN, J.; KAMBER, M.; PEI, J. **Data mining: Concepts and Techniques**. 3. ed. San Francisco: Elsevier, 2011. Disponível em: <[http://www.cse.hcmut.edu.vn/~chauvtn/data_mining/Texts/\[1\]%20Data%20Mining%20-%20Concepts%20and%20Techniques%20\(3rd%20Ed\).pdf](http://www.cse.hcmut.edu.vn/~chauvtn/data_mining/Texts/[1]%20Data%20Mining%20-%20Concepts%20and%20Techniques%20(3rd%20Ed).pdf)>. Acesso em: 22 mai. 2016.

HAYKIN, S. **Redes neurais: princípios e práticas**. 2. ed. Porto Alegre: Bookman, 2001.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Sobre o Enem - INEP. **Sobre o ENEM**. 2016. Disponível em: <<http://portal.inep.gov.br/web/enem/sobre-o-enem>>. Acesso em 12 nov. 2016.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Sobre o Enem - INEP. **Microdados para download**. 2015. <<http://portal.inep.gov.br/basicas-levantamentos-acessar>>. Acesso em 06 nov. 2016.

IZQUIERDO, I.; BEVILAQUA, L. R. M.; CAMMAR, M. A arte de esquecer. **Estud. av.**, São Paulo, v. 20, n. 58, Setembro/Dezembro 2006. ISSN 1806-9592. Disponível em: <http://www.scielo.br/scielo.php?pid=S0103-40142006000300024&script=sci_arttext&tlng=pt>. Acesso em: 10 abr. 2016.

KALINLI, A.; SAGIROGLU, S. Elman Network with Embedded Memory for System Identification. **Journal of Information Science and Engineering**, v. 22, p. 1555-1568, 2006. Disponível em: <http://www.iis.sinica.edu.tw/page/jise/2006/200611_16.pd>. Acesso em: 6 mai. 2016.

LAROSE, D. T. **Data mining methods and models**. 1. ed. John Wiley & Sons, 2006. 344 p.

LIU, B. **Web Data Mining: exploring hyperlinks, contents, and usage data**. Chicago: Springer, 2006.

LIU, B. Sentiment Analysis and Opinion Mining. **Synthesis Lectures On Human Language Technologies**, Colorado, v. 5, n. 1, p. 1-167, 23 mai. 2012.

LOPES, T. J. P.; HIRATANI, G. Mineração de Opiniões e fatos aplicada à Análise de Investimentos. **Clube de Autores**, São Paulo, p. 73, 2008.

MARCHIORI, P. Z. A ciência e a gestão da informação: compatibilidades no espaço profissional. **Ci. Inf.**, v. 31, n. 2, p. 72-79, mai/ago 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12910.pdf>>. Acesso em: 23 mai. 2016.

MALHOTRA, N. **Pesquisa de marketing**: uma orientação aplicada. 4.ed. Porto Alegre: Bookman, 2006.

MARTELETO, R. M. Análise de redes sociais: aplicação nos estudos de transferência da informação. **Ci. Inf.**, Brasília, v. 30, n. 1, p. 71-81, jan. 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n1/a09v30n1>>. Acesso em: 06 jun. 2016.

MINISTÉRIO DA EDUCAÇÃO E CULTURA (MEC). **ENEM** - Apresentação. Disponível em: <<http://portal.mec.gov.br/enem-sp-2094708791>>. Acesso em 12 nov. 2016.

NOGUEIRA, E. D. A. **Análise de brand equity sob a perspectiva do consumidor nas mídias sociais por meio da mineração de opinião e análise de redes sociais**. Dissertação (Mestrado) - Curso de Programa de Pós- Graduação em Ciência, Gestão e Tecnologia da Informação, Universidade Federal do Paraná. Curitiba, p. 236. 2015.

OTLET, P. *Traité de documentation: le livre sur le livre, théorie et pratique*. **Mundaneum**, Bruxelles, 1934. Disponível em: <http://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf>. Acesso em: 23 abr. 2016.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends in Information Retrieval**, v. 2, n. 1-2, p. 1–135, 2008. Disponível em: <<http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>>. Acesso em: 6 abr. 2016.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up?: sentiment classification using machine learning techniques. **Proceedings of the ACL-02 conference on Empirical methods in natural language processing**, Stroudsburg, USA, v. 10, p. 79–86, 2002.

PINHEIRO, L. V. R. Gênese da Ciência da Informação ou sinais anunciadores da nova área. In: _____ **O campo da Ciência da Informação**: gênese, conexões e especificidades. João Pessoa, 2002. p. 61-86. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/17/1/LenaGeneseUFPB-2.pdf>>. Acesso em: 23 abr. 2016.

PINTO, C. A. D. S. et al. Vivendo e Aprendendo no Facebook: uma visão da usabilidade em redes sociais na sociedade do conhecimento. In: **V SIMPÓSIO NACIONAL DA ABCIBER**, Florianópolis, p. 1-15, 2011.

PORTAL BRASIL. Educação. **ENEM 2015 registra o menor número de faltas em sete anos**. 2015. Disponível em: <<http://www.brasil.gov.br/educacao/2015/10/enem-tem-25-5-de-abstencao-menor-taxa-desde-2009>>. Acesso em: 10 abr. 2016.

PRIMO, A. A cobertura e o debate público sobre os casos Madeleine e Isabella: encadeamento midiático de blogs, Twitter e mídia massiva. **Revista Galáxia**, São Paulo, p. 43-59, dez 2008.

RECUERO, R. **Redes sociais na internet**. Porto Alegre, 2009. 191 p.

RIBEIRO, E. Tecnologia e poder semiótico: escrever, hoje. **Texto livre: Linguagem e tecnologia**, v. 8, n. 1, 2005. Disponível em: <<http://www.periodicos.letras.ufmg.br/index.php/textolivres/article/download/8681/7701>>. Acesso em: 6 abr. 2016.

SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramZero, Revista de Ciência da Informação**, São Paulo, dezembro 1999. Disponível em: <www.ime.usp.br/~vwsetzer>. Acesso em: 23 abr. 2016.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistema de banco de dados**. 5. ed. Rio de Janeiro: Elsevier, 2006.

SILVA, N. G. R. D. **BestChoice: classificação de sentimento em ferramentas de expressão de opinião**. TCC (Graduação) - Curso de Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco. Recife. 2010.

VERCELLIS, C. **Business Intelligence: Data Mining and optimization for decision making**. Milão: John Wiley and Sons, 2009.

VON HAYEK, F.. The use of knowledge in society. **The American Economic Review**, Nashville, v. 35, n. 4, p. 519-530, September 1945. Disponível em: <http://www.kysq.org/docs/Hayek_45.pdf>. Acesso em: 23 abr. 2016.

WU, X.; KUMAR, V. The top ten algorithms in data mining. London: **CRC Press Book**, 2009.

5. Os textos base da questão estão diretamente envolvidos com a temática e auxiliam, de fato, na resolução das questões? *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

6. As quantidades de questões cobradas estão de acordo com o prazo disponível? *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

7. Você concorda que o ENEM é uma ferramenta de avaliação da qualidade do ensino médio? *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

8. Você considera justo que com apenas 450 pontos a certificação do ensino médio seja possível? *

Marcar apenas uma oval.

- Sim
 Não

9. Qual o seu sexo? *

Marcar apenas uma oval.

- Feminino
 Masculino

10. Em que faixa etária você está? *

Marcar apenas uma oval.

- 15 - 17 anos
 18 a 20 anos
 21 a 23 anos
 Tenho mais que 23 anos

11. Em que tipo de colégio que você estuda ou estudou? *

Marcar apenas uma oval.

- Escola pública
 Escola particular

APÊNDICE B – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO APLICADA AOS ALUNOS DOS CURSOS PREPARATÓRIOS

Questionário para coleta de opiniões sobre o Enem

Página 1

As informações aqui coletadas serão utilizadas para criar uma base de dados que passará por análise. Não serão divulgadas informações referentes à identidade dos alunos ou da instituição à qual o aluno pertence, sendo estas confidenciais. Esta pesquisa será utilizada como parte dos resultados de um trabalho de conclusão de curso de Gestão da Informação da UFPR.

1. Qual o seu gênero? *

- feminino
- masculino

2. Em que tipo de escola você estuda ou estudou no último ano do ensino médio? *

- escola pública
- escola particular

3. Quantos anos você tinha na última vez que fez o ENEM? *

anos

4. Em quanto tempo você se preparou para o ENEM? *

- menos de três meses
- entre quatro e seis meses
- entre sete meses e um ano
- mais de um ano
- mais de dois anos
- não me preparei (porque?)

5. Você já teve problemas com o sistema de inscrição? *

não

sim (qual?)

6. Você acredita que o preço cobrado para realizar a prova é justo? *

Atualmente o preço cobrado é de R\$68,00 reais, alunos do terceiro ano do ensino médio não pagam.

sim

não

7. Como você avalia o conteúdo cobrado na prova em relação ao seu conhecimento adquirido em sala de aula? *

ótimo

muito bom

bom

regular

insatisfatório

8. Você acredita que apenas com a instrução oferecida em sala aula é suficiente para a realização da prova? *

sim

não (porque?)

9. O número de questões está compatível com o tempo de prova? *

sim

não

10. Você concorda que o ENEM é uma ferramenta de avaliação da qualidade do ensino médio? *

sim

não (porque?)

11. Você concorda que a certificação do ensino médio seja possível com 450 pontos? *

sim

não

APÊNDICE C – QUESTIONÁRIO PARA COLETA DE OPINIÕES SOBRE O ENEM – VERSÃO APLICADA AOS ALUNOS DA UFPR

Questionário para coleta de opiniões sobre o Enem no Sociais Aplicadas

Página 1

Prezado(a) colaborador(a),

Meu nome é Ana Paula Fantin e estou realizando o trabalho de conclusão de curso no curso de Gestão da Informação da UFPR. O objetivo do meu trabalho é criar uma base de dados com as opiniões aqui coletadas. Não serão divulgadas informações referentes à identidade dos alunos, sendo estas confidenciais.

Para tal, estou encaminhando um questionário, que acredito não exigir mais do que 2 minutos para ser respondido.

Agradeço pela sua preciosa ajuda!

Qual o seu curso? *

Administração - UFPR

Economia - UFPR

Ciências Contábeis - UFPR

Gestão da Informação - UFPR

Você está atualmente matriculado na UFPR? *

sim

não

Página 2

1. Qual o seu gênero? *

feminino

masculino

2. Em que tipo de escola você estuda ou estudou no último ano do ensino médio? *

escola pública

escola particular

3. Quantos anos você tinha na última vez que fez o ENEM? *

anos

4. Em quanto tempo você se preparou para o ENEM? *

- menos de três meses
- entre quatro e seis meses
- entre sete meses e um ano
- mais de um ano
- mais de dois anos
- não me preparei (porque?)

5. Você já teve problemas com o sistema de inscrição? *

- não
- sim (qual?)

6. Você acredita que o preço cobrado para realizar a prova é justo? *

Atualmente o preço cobrado é de R\$68,00 reais, alunos do terceiro ano do ensino médio não pagam.

- sim
- não

7. Como você avalia o conteúdo cobrado na prova em relação ao seu conhecimento adquirido em sala de aula? *

- ótimo
- muito bom
- bom
- regular
- insatisfatório

8. Você acredita que apenas com a instrução oferecida em sala aula é suficiente para a realização da prova? *

- sim
- não (porque?)

9. O número de questões está compatível com o tempo de prova? *

- sim
 não

10. Você concorda que o ENEM é uma ferramenta de avaliação da qualidade do ensino médio? *

- sim
 não (porque?)

11. Você concorda que 450 pontos sejam suficientes para a obtenção da certificação do ensino médio? *

- sim
 não

12. Que nota você daria ao ENEM? *

Considere a nota variando de 0 - 10

Qual edição do ENEM você participou? *

por exemplo "ENEM 2015".

APÊNDICE D – BASE DE DADOS “OPINIÕES ENEM”

continua

ID	Frase	Rede social
OP_001	Deveria haver um exame nacional à exemplo do ENEM para o provimento de cargos Públicos !!!	Facebook
OP_002	Inadmissível um Instituto desse porte, com quase três meses de planejamento após o ENEM, não ter e dar suporte de qualidade ao site que move SETE MILHÕES DE PESSOAS EM TODO O BRASIL! Absurdo!	Facebook
OP_003	adianta muito liberar a nota incompleta...comprei um carro, mas comprei so os pneus, então nao comprei um carro né?!	Facebook
OP_004	Aí o site pede meu CPF minha sena e só libera se marcar o desenho escolhido aí ele pede pra marcar uma casa VC marca a porra da casa e aí diz que errei a marcação ????? AFF o site que precisa fazer uma prova né !	Facebook
OP_005	sabe que milhões de pessoas fizeram a merda da prova e mesmo assim insiste em deixar um site fuleiro pra suportar a quantidade de acessos. יריירירי	Facebook
OP_006	Para quer disseram que iram arrumar o site para ficar uma merda dessa, inep como sempre uma palhaçada.	Facebook
OP_007	Em saber q a pessoa estuda o ano todo pra fazer uma prova lixo, organizada por uma instituição lixo, pra entrar em uma universidade. Eita brasil	Facebook
OP_008	Gente quase morri hj cm essa prova vo aguarda para oq vinher amanha	Facebook
OP_009	Não estou conseguindo fazer a inscrição. Na metade dela apareceu: "SESSÃO EXPIRADA" Como assim? Eu estava lendo o enunciado (na inscrição) e na primeira linha dele sinalizou e fechou a minha inscrição!? Não faz sentido!!!!	Facebook
OP_010	continua "inscrição não confirmada, pagamento não efetuado" embora tenha pago no dia 26....e agora???	Facebook
OP_011	Fiz o pagamento dia 03/06 e até hoje consta que a inscrição não foi confirmada devido o não efetuoamento do pagamento. Já fazem 6 dias.... Como assim?	Facebook
OP_012	Como a página do enem pode publicar que as notas já podem ser conferidas quando na realidade do serão publicadas em março? Como uma prova assim pode ser usada pra vestibular? Caros organizadores do enem, se forem dar alguma informação aos participantes que esta seja correta por favor	Facebook
OP_013	Estão achando ruim por não conseguir ver a nota do ENEM? esperem até entrar na faculdade..... Tem professor que leciona aula desde a pré história e até hoje tem dificuldades pra lançar a nota em sistema.	Facebook
OP_014	O mínimo é 450,eu tirei no anterior 639 e não ganhei a bolsa de engenharia... Esse ano tô com 598	Facebook
OP_015	Eu tirei uma boa nota muito Obrigada senhor 600 pontos	Facebook
OP_016	ENEM = EXAME NACIONAL DO ENSINO MÉDIO , METADE DA PROVA TINHA NADA DOQ O PESSOAL ESTUDO NOS 3 ANOS , TOMEM MAIS VERGONHA !	Facebook
OP_017	Pela minha nota vai dar pra eu pegar um curso de entregador de água kkkkk	Facebook
OP_018	Aumentou o preço da inscrição e o sistema piorou.	Facebook
OP_019	e eu achando q tinha lacrado na redação :(Facebook
OP_020	O ENEM é uma palhaçada!	Facebook
OP_021	Caramba falto 120 pontos na redação pra eu concluir o ensino médio sera que tem algum jeito ou já era passei em todas as outras	Facebook

continua

ID	Frase	Rede social
OP_022	700 na redação e em matemática era tudo que eu queria e consegui! Haha \0/	Facebook
OP_023	sinceramente de saco cheio, passo o dia na frente do computador e não consigo passar do primeiro passo da inscrição porque fica mandando eu atualiza a página, que ocorreu uma FALHA INESPERADA. Que porcaria é essa? ENEM sempre com falhas se não é nas inscrições são nas provas!!!!	Facebook
OP_024	Acabei de me inscrever e não demorei nem 10min para finalizar minha inscrição !	Facebook
OP_025	Alguém conseguiu fazer a inscrição pelo celular? Eu sempre faço pelo celular, mas agora está pedindo pra baixar navegadores que JÁ USO, alguém com mesmo problema? ENEM sempre com problemas, que lixo, como se já não bastasse essa burocracia nojenta, 😞😞😞😞	Facebook
OP_026	O site tá fora do ar ! E agora ?!	Facebook
OP_027	Impossível fazer a inscrição, um questionário enorme para responder. E enquanto vc está lendo para responder da tempo expirado. 😞😞😞😞😞	Facebook
OP_028	Fiz com tranquilidade minha inscrição.	Facebook
OP_029	caraio... 65\$ vsf	Facebook
OP_030	solicitei carência, mas depois fui ver os critérios na lei e fui cancela-la e não têm a opção, como também não têm a opção de gerar boleto, e para piorar não têm nada confirmando a carência ou não. Não sei o que fazer, nem meus dados tou conseguindo editar. Isso aconteceu com vcs, como vou saber se foi aprovada?	Facebook
OP_031	Não estou conseguindo ver meu resultado de 2015. Preciso muito dele. O problema é que sempre que tento consulta-lo, a pagina redireciona para a inscrição do ENEM 2016	Facebook
OP_032	Não consigo nem iniciar a inscrição	Facebook
OP_033	Não consegui me inscrever e agora ja fecho a inscrição 😞 a pagina so deu erro toda vez que tentei é pra nao fazer mesmo	Facebook
OP_034	Mais alguém aí ta com problemas pra se inscrever? Não consigo, só aparece: Ocorreu uma falha inesperada. Por favor, atualize esta página e tente novamente	Facebook
OP_035	Pelo valor da inscrição acho que vc ja sai com o diploma na mão	Facebook
OP_036	E eu sou uma burra q gastei 4 meses estudando para ter a tenção de ficar 2 dias por varias horas quebrando a cabeça p fazer uma prova... P quando sair o resultado simplismente n posso usar pq errei a data de conclusão do ensino médio, q ja me formei mas coloquei q só vou formar ano q vem. Burrarraaaaaaaaaa	Facebook
OP_037	Já vou por meu currículo no Subway da minha cidade, aproveitar q abriu recentemente	Facebook
OP_038	Vi minha nota...e o Mc Donald não parece tão ruim.	Facebook
OP_039	Carneiro Que nota mais escrota foi essa que me deram. Desisto!	Facebook
OP_040	Vou processar o enem por danos morais, se eu quisesse ser revistada eu virava mulher de ladrão ou eu mesmo começava a roubar!	Facebook
OP_041	horrível mesmo. ENEM 2015 uma vergonha a organização	Facebook
OP_042	A unica coisa boa no enem foi o tema da redação..... Fiquei La hrs e hrs fui um dos três ultimos a sair da sala hj pra acertar 51 questoes de 180(segundo o gabarito do dia)... 😞😞😞👉👉	Facebook
OP_043	gente, prova de mtt tava fácil demais O.o	Facebook

		continua
ID	Frase	Rede social
OP_044	Matemática foi fácil	Facebook
OP_045	Brincadeira o que fizeram com o pessoal que guarda o sábado, fazer todos esperarem do meio dia até as dezenove horas, pra poder fazer a prova. Que bobeira. Ter que ficar sem fazer literalmente nada, até a prova começar. Que absurdo! E outra eu guardo o dia de sábado inteiro, e não só até dezenove horas e agora hein?! Palhaçada.	Facebook
OP_046	O que eu não vi é como proceder quanto a um recurso administrativo. Alguém sabe onde devo ir para entrar com um?	Facebook
OP_047	Opto por anulação do ENEM, pois deixará margem para que concursos com irregularidades façam o mesmo e ganhem na justiça...	Facebook
OP_048	sinceramente de saco cheio, passo o dia na frente do computador e não consigo passar do primeiro passo da inscrição porque fica mandando eu atualiza a página, que ocorreu uma FALHA INESPERADA. Que porcaria é essa? ENEM sempre com falhas se não é nas inscrições são nas provas!!!!	Facebook
OP_049	Alguém sabe como funciona o processo de isenção da taxa de inscrição? Quando fiz a inscrição, optei pela isenção do pagamento. Ai fica em período de análise, se não me engano são 48h, até obter uma resposta positiva ou negativa. Só que já se passou esse tempo e ainda não obtive resposta. Tentei ligar no 0800 e não consegui. No site do ENEM as informação sobre isso são poucas e ruins... Questão de requerer a isenção da taxa.. Eu já terminei o 3 ano, inclusive faço uma graduação, então tenho que declarar carência... Pesquisei em alguns lugares os requisitos para requerer a isenção, cada lugar fala uma coisa diferente, e no site do ENEM o que está escrito é pouco conclusivo. Alguém tem mais detalhes? Obrigada	Facebook
OP_050	Preciso de ajuda! Meu nome no CPF está "Thaina" com "i", e o certo seria com "y", gostaria de saber se tem algum problema, pois já fiz a inscrição e meu nome está errado, o que faço?	Facebook
OP_051	Que absurdo, tirei mais de 200 em matemática no ENEM. Se eu não soubesse contar não tinha tirado mais de 50%	Twitter
OP_052	tô vendo aqui usa a nota do enem então é um pouco mais fácil mas ainda assim	Twitter
OP_053	mulheres q so quer q o enem passe logo pra nao ter mais q ir pras aulas	Twitter
OP_054	se os professores soubessem o quanto deixa os alunos loco qnd começa a falar de enem faculdade cursos etc talvez nem comentassem	Twitter
OP_055	Tava lendo aqui meus tweets e tem tanto erro de concordância, ainda bem que isso aqui é twitter e não ENEM	Twitter
OP_056	qnd eu tiver que fazer enem eu vou forjar minha própria morte	Twitter
OP_057	Como se não bastasse o crush do enem	Twitter
OP_058	Só achei ruim a parte de que eu tenho q fazer ENEM esse ano ☹️👊	Twitter
OP_059	Minha nota no enem e meu curso normal não me permitem "me expressar"	Twitter
OP_060	Estudar pra porra ano que vem, porque tem ENEM ^{zzz}	Twitter
OP_061	Se Lula e Dilma conseguirem acertar 50% numa prova do ENEM, eu não falo mais dos dois.	Twitter
OP_062	Não sei, to loka devia ser uma das alternativas pra todas as questões do ENEM.	Twitter

		continua
ID	Frase	Rede social
OP_063	serio o enem realmente é ridiculo	Twitter
OP_064	seria meu sonho? ir bem no enem sem saber exatas	Twitter
OP_065	Mano tava vendo o caderno do enem, é tudo fácil menos matemática	Twitter
OP_066	Como q em 2016 ainda existem pessoas que não sabem a diferença de "mais" e "mas"? Eu realmente fico chocada! Depois reclamam da nota do ENEM	Twitter
OP_067	então remunera professor com salário descente pra na prova do enem ter chance pra nossa gente	Twitter
OP_068	até agora não achei ninguém que me disse que foi bem no enem	Twitter
OP_069	Prova que o Governo atual é funcional é o ENEM Sure. Ngm sabe como corrigem, como avaliam, nem q país eles vivem p considerar aquilo EM	Twitter
OP_070	sou tão difícil q to quase mudando meu nome pra enem	Twitter
OP_071	to refazendo a prova de matemática do enem desse ano que nao fiz por razões de demorei na redação e mano, tem muita questão fácil, caralho	Twitter
OP_072	O ENEM seria bem mais fácil se cada um fizesse uma prova que caísse só coisas que tinham a ver com a faculdade desejada né	Twitter
OP_073	Da onde que as questões do enem foram formuladas para pessoas que fizeram o ensino médio apenas ? ! Foram alem de questões de conceituais!	Twitter
OP_074	CARALHO QUE ORGULHO ESSE TEMA DO ENEM PRIMEIRA VEZ BRASIL FAZENDO ALGO CERTO (e justo) 🙄	Twitter
OP_075	A gente estuda gramática todos os anos da vida, pra chegar no enem ter só interpretação de texto	Twitter
OP_076	n q os vestibulares sejam fáceis mas pelo menos é só o conteúdo q a gente estuda ou sabe ou n sabe não fica de mimimi como o enem #ENEM2015	Twitter
OP_077	fiquei em 116 na colocação, e no enem eu nem conseguia uma colocação pq eu ia bem em historia/filosofia/sociologia e linguas e só	Twitter
OP_078	O enem esse ano ele cobrou questões com mais conteúdo do que interpretação de texto.. e isso fodeu. Em ciências da natureza fodeu bonito	Twitter
OP_079	Esse foi o pior ENEM da minha vida, e olha que já fiz muitos viu 😊	Twitter
OP_080	Eu achei esse enem o mais doido de todos. Deve tanta crítica/questionamentos sociais que achei até que fui que fiz as questões	Twitter
OP_081	achei que a prova do enem foi tão tranquila, porque assim eu que to no segundo ano resolvi todas as questões imagina quem tava no 3°	Twitter
OP_082	ENEM é uma prova de resistencia, ela te vence pelo cansaço	Twitter
OP_083	E aquela química do ENEM???? Estava destruidora	Twitter
OP_084	ENEM mais difícil a cada ano. Daqui a uns 3 anos a galera ta fazendo a prova do ita pra entrar na federal	Twitter
OP_085	Me sinto extremamente prejudicada hoje na prova do ENEM pelo motivo de: não ter carteira de canhoto	Twitter
OP_086	quando cheguei na parte de química no enem, eu esqueci até meu nome	Twitter
OP_087	@dilmabrdilma mulher a prova tava difícil demais tu se vingou bem da galera	Twitter
OP_088	Ou diminui as questões ou mais tempo de prova, quase não consegui completar o gabarito, querida! @dilmabr	Twitter

		conclusão
ID	Frase	Rede social
OP_089	@dilmabr como se a verdadeira qualidade da educação pudesse ser medida por uma prova exclusivamente lógico-matemática e linguística...	Twitter
OP_090	@dilmabr Hoje o ENEM é a forma mais justa para que todos tenham oportunidade	Twitter
OP_091	@dilmabr o ENEM foi o fiasco. Agora sem saber se mulher é mulher de fato? Palhaçada esquerdistas s escrúpulos.	Twitter
OP_092	Dano.ndução irreversível @dilmabr o ENEM é uma ótima idéia..mas quando vamos fzer a prova somos massacrados pela prova q parece ser para alunos de escolas privadas!	Twitter
OP_093	Ok, acho ótima a proposta de questões do ENEM. Mas não vem me dizer que é exame de Ensino Médio porque não tem NADA a ver com o método de EM	Twitter
OP_094	Espanhol no ENEMestava tão fácil, que eu me arrependi de ter escolhido inglês	Twitter
OP_095	No país onde mulheres são espancadas, estupradas, rebaixadas profissionalmente... Parabéns, ENEM, pelo tema da redação!	Twitter
OP_096	A sociedade precisa debater temas como a violência -física e verbal- contra a mulher e o ENEM forçou o diálogo que muitos evitam. Ameeeei	Twitter
OP_097	Até a assistente da minha sala tava falando que o ENEM cada ano fica mais difícil	Twitter
OP_098	Kkkkk! fui fazer a prova do ENEM 2015, das 14 que eu respondi eu acertei 10 Kkkkk eu to mto feliz	Twitter
OP_099	As vzs eu to aqui lendo as questões do enem de 2014 pra 2015 e as vzs eu fico em duvida se elas tao em portugues ou em arabico	Twitter
OP_100	meu deus do ceu aquela prova de natureza do enem 2015 foi obra do do demonio	Twitter
Total	100 sentenças coletadas	

FONTE: Dados da pesquisa (2016).