



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
IFCE CAMPUS ARACATI
COORDENADORIA DE CIÊNCIA DA COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

MARCÍLIO BEZERRA CARNEIRO

**SEMT - SOLUÇÃO CONSCIENTE DE EMOÇÃO BASEADA EM
MINERAÇÃO DE TEXTOS PARA PREVENÇÃO DA DEPRESSÃO
GESTACIONAL**

**ARACATI-CE
2020**

MARCÍLIO BEZERRA CARNEIRO

SEMT - SOLUÇÃO CONSCIENTE DE EMOÇÃO BASEADA EM MINERAÇÃO DE TEXTOS
PARA PREVENÇÃO DA DEPRESSÃO GESTACIONAL

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Bacharelado em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE - campus Aracati, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Mário Wedney de Lima Moreira

Aracati-CE
2020

MARCÍLIO BEZERRA CARNEIRO

SEMT - SOLUÇÃO CONSCIENTE DE EMOÇÃO BASEADA EM MINERAÇÃO DE TEXTOS
PARA PREVENÇÃO DA DEPRESSÃO GESTACIONAL

Trabalho de Conclusão de Curso (TCC)
apresentado ao curso de Bacharelado em
Ciência da Computação do Instituto Federal
de Educação, Ciência e Tecnologia do
Ceará - IFCE - campus Aracati, como re-
quisito parcial para obtenção do título de
Bacharel em Ciência da Computação.

Aprovada em 27/03/2020

BANCA EXAMINADORA

Prof. Dr. Mário Wedney de Lima Moreira (Orientador)
IFCE

Prof. Ma. Érica de Lima Gallindo
IFCE

Prof. Me. Silas Santiago Lopes Pereira
IFCE

DEDICATÓRIA

Aos meus pais e toda minha família que se fizeram presentes nos momentos de alegria e nos mais difíceis da minha vida. Admiração e respeito é o que sinto por todos vocês.

AGRADECIMENTOS

Agradeço a Deus por estar aqui tendo mais uma vitória na vida. O Senhor é a base de tudo. Aos meus filhos Davi Lucas e Ana Lívia, vocês são a minha maior motivação nessa vida e são as pessoas mais importantes para mim, amo vocês. Aos meus pais, Francisco Nunes e Marilene Bezerra, meus grandes exemplo na vida, agradeço por sempre me mostrarem o quanto a educação é tão importante em nossas vidas e agradeço por tanto esforço para que eu me tornasse a pessoa que sou hoje. Ao meu irmão, Marcelo Bezerra, por sempre estar ao meu lado.

A todos os professores, desde a base, que fizeram possível esse momento, principalmente ao meu avô Albismar Bezerra do Nascimento, que educou dezenas de pessoas em uma época tão difícil, obrigado pelas charadas difíceis que já me faziam usar a lógica. À minha avó Maria Zilma, que sempre me ajudou mais do que sempre pôde.

Ao professor Mário Wedney por aceitar dedicar seu valioso tempo na orientação deste trabalho, obrigado por toda atenção. À professora Carina Oliveira por sempre estar presente nos momentos que precisava.

Aos meus amigos que iniciaram uma longa e árdua jornada juntos, Djavam Moreira, Helioenai Mariano e Magno Braga obrigado por tantos momentos bons. A minha namorada Queila Valente, uma pessoa especial que está em minha vida e que não me deixa desistir nunca.

RESUMO

As emoções influenciam em todos os aspectos no comportamento humano, e todos esses aspectos vão modelando a vida de cada ser humano, isso impacta diretamente no modo de vida. Algumas doenças estão diretamente ligadas às emoções, entre elas a depressão que é uma das doenças com maiores impactos na sociedade. Diante dessa problemática o objetivo deste trabalho é apresentar a Solução consciente de Emoção baseada em Mineração de textos para prevenção da depressão gestacional (SEMT). O SEMT usa a mineração de texto para analisar textos extraídos a partir de gestantes com o objetivo de identificar seus sentimentos através de técnicas de processamento de linguagem natural e do algoritmo *naive bayes*. Como estudo de caso, é feito um questionário com gestantes para extrair os textos analisados. É analisado o desempenho do modelo através de métricas associadas a matriz de confusão e desempenho do modelo, quando submetido a uma nova base de dados coletada do questionário. Já com os resultados, o SEMT conseguiu ter bons desempenhos em todas as métricas e um excelente desempenho a classificar os novos atributos da base de dados coletada. Assim, os resultados obtidos pelo SEMT podem ser usados como um auxílio para profissionais da saúde que fazem o acompanhamento das gestantes.

Palavras-chaves: Emoções. Gestantes. Mineração de textos. Processamento de linguagem natural. *Naive bayes*.

ABSTRACT

Emotions influence all aspects of human behavior, and all these aspects shape the life of each human being, this directly impacts the way of life. Some diseases are directly linked to emotions, including depression, which is one of the diseases with the greatest impact on society. In view of this problem, the objective of this work is to present the Conscious Emotion Solution based on Mining texts for the prevention of gestational depression (SEMT). SEMT uses text mining to analyze texts extracted from pregnant women in order to identify their feelings through natural language processing techniques and the naive bayes algorithm. As a case study, a questionnaire is made with pregnant women to extract the analyzed texts. The model's performance is analyzed through metrics associated with the model's confusion and performance matrix, when submitted to a new database collected from the questionnaire. With the results, SEMT managed to have good performances in all metrics and an excellent performance to classify the new attributes of the collected database. Thus, the results obtained by SEMT can be used as an aid to health professionals who monitor the pregnant women.

Keywords: Emotions. Pregnant women. Text mining. Natural language processing. Naive bayes.

LISTA DE ILUSTRAÇÕES

Figura 1 – Linha do tempo da inteligência artificial.	17
Figura 2 – Classificação de textos por meio de aprendizado supervisionado. . .	20
Figura 3 – Processo de <i>Knowledge Discovery in Databases</i>	23
Figura 4 – Processamento de Linguagem Natural	24
Figura 5 – Gráficos de texto extraídos do Sobek.	26
Figura 6 – Arquitetura do EmoViz.	27
Figura 7 – Protótipo da ferramenta CroCA.	28
Figura 8 – Tela de registro do paciente.	30
Figura 9 – Etapas do Fluxo de Execução do SEMT.	32
Figura 10 – Divisão da Base de Dados.	33
Figura 11 – Classificação do texto	37
Figura 12 – Matriz de confusão do SEMT.	41
Figura 13 – Emoções	43
Figura 14 – Tela do questionário.	44
Figura 15 – Classificação dos Textos.	47
Figura 16 – Classificação por Emoção.	47

LISTA DE TABELAS

Tabela 1 – SEMT e trabalhos relacionados	31
Tabela 2 – divisão de dados em treinamento e teste.	34
Tabela 3 – Aplicação de <i>Stopwords</i>	35
Tabela 4 – Aplicação de <i>stemming</i>	35
Tabela 5 – Alegria x Tristeza.	36
Tabela 6 – Matriz de Confusão	38
Tabela 7 – Interpretação do coeficiente Kappa	40
Tabela 8 – Índice de desempenho do SEMT.	42
Tabela 9 – Índice Kappa do SEMT.	43
Tabela 10 –Validação do SEMT.	46

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
FACS	<i>Facial Action Coding System</i>
IA	Inteligência Artificial
IFCE	Instituto Federal de Educação e Tecnologia do Ceará.
KDD	<i>Knowledge Discovery in Database</i>
MD	Mineração de Dados
NLTK	<i>Natural Language Toolkit</i>
OMS	Organização Mundial da Saúde
PLN	Processamento de Linguagem Natural

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Motivação	15
1.2	Objetivos	15
1.2.1	Objetivo Geral	15
1.2.2	Objetivos Específicos	15
1.3	Organização do Trabalho	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Inteligência Artificial	16
2.2	Aprendizado de Máquina	18
2.2.1	Classificação	19
2.2.1.1	Classificação em Textos	19
2.2.1.2	Análise de Sentimentos	21
2.3	Mineração de Textos	22
2.4	Processamento de Linguagem Natural	23
2.5	O Algoritmo Naive Bayes	25
3	TRABALHOS RELACIONADOS	26
3.1	Using a Text Mining Tool to Support Text Summarization	26
3.2	EmoViz: Mining the World's Interest through Emotion Analysis	27
3.3	CroCA - Cromoterapia e Computação Afetiva: auxiliando os estados de ansiedade	28
3.4	Puerpério APP - uma Aplicação Móvel de Apoio à Puérpera com Depressão Pós - Parto baseada na Escala de Edimburgo	29
3.5	Descrição de um sistema de suporte ao diagnóstico de demência e Transtornos Mentais Relacionados	30
3.6	Comparativos da Proposta com os Trabalhos Relacionados	31
4	PROPOSTA	32
4.1	Coleta dos Dados	32
4.2	Base de Dados	33
4.3	Pré-Processamento dos Dados	34
4.3.1	<i>NLTK</i>	34
4.3.2	<i>Stopwords</i>	34
4.3.3	Stemming	35
4.4	Treinamento e Classificação dos Dados	36

4.5	Avaliação do Modelo Preditivo	37
4.5.1	Matriz de confusão e métricas associadas	37
4.5.2	Coeficiente Kappa	39
5	RESULTADOS	41
5.1	Desempenho do Modelo	41
5.2	Desempenho do Modelo na Classificação dos Textos	43
6	CONCLUSÃO	48
	REFERÊNCIAS	49

1 INTRODUÇÃO

As emoções modificam o comportamento humano. Sentimentos como raiva, alegria, tristeza, surpresa, decepção, entre outros, são algumas emoções que as pessoas têm diariamente. Estas emoções contribuem diretamente para o comportamento humano. Fatores como desemprego, violência, situação financeira, metas no trabalho, excesso de estudos entre outras são situações capazes de gerar um desequilíbrio emocional, ocasionando diversas doenças, a principal delas é a depressão. Segundo a Organização Mundial da Saúde (OMS), a depressão afeta 300 milhões de pessoas em todo o planeta, tornando assim a doença mais incapacitante até 2020. De acordo com (FREITAS-MAGALHÃES, 2013), todo o processo emocional desempenha um papel decisivo na conduta humana. A depressão é uma doença que impacta consideravelmente a vida do paciente e de toda sua família. Esta caracteriza-se por uma mudança grande no humor, um grau de tristeza elevado e falta de interesse em atividades cotidianas.

As mulheres estão mais vulneráveis à depressão, devido principalmente às questões hormonais. O estrogênio é um hormônio feminino que causa muitas alterações no corpo e no comportamento da mulher, chegando a altos índices durante o período pré e pós gestacional, fazendo com que a mulher fique muito sensível às mudanças em seu comportamento. As alterações deste hormônio podem causar insônia, dores de cabeça, irritabilidade e alteração no humor. Durante a gravidez, o estrogênio é responsável por constantes alterações no comportamento da mulher e, durante o parto, estas alterações são ainda maiores. Por isso, a importância em acompanhar as emoções das gestantes, tem como objetivo minimizar o risco de depressão gestacional, uma doença grave e cada vez mais comum entre as mulheres.

A depressão gestacional é um transtorno, que pode prejudicar a relação mãe-criança, por isso a importância do bem-estar físico e psicológico da mãe durante toda sua gestação e principalmente tanto no período que antecede o parto como no período pós-parto. É comum a mãe passar por um período de tristeza prolongado pós-parto, o chamado *baby blues*, devido à alteração hormonal. Existem inúmeros sintomas que podem indicar a depressão pós-parto. Segundo (SCHWENGBER; PICCININI, 2003), esses sintomas incluem irritabilidade, choro frequente, sentimentos de desamparo e desesperança, falta de energia e motivação, desinteresse sexual, transtornos alimentares e do sono, a sensação de ser incapaz de lidar com novas situações, bem como queixas psicossomáticas. Isso acontece devido a inúmeras mudanças em um período curto na vida da mulher, e vários fatores influenciam para a depressão gestacional, como a falta de apoio familiar e dificuldades financeiras, assim o profissional de saúde

deve estar atento ao paciente, que necessita, nestes casos, de um acompanhamento diário.

As emoções, além de expressadas entre relacionamento de pessoas, podem ser também dentro de determinados textos, imagens e sons. A análise de sentimento é o estudo destas emoções. Devido ao crescimento das redes sociais e os inúmeros dados gerados, a análise de sentimento vem sendo de grande utilidade para identificar emoções e reações das pessoas. De acordo com (AZEVEDO; BEHAR; REATEGUI, 2012), a mineração de textos pode ser definida como um processo intensivo de conhecimento no qual um usuário interage com uma grande quantidade de documentos utilizando ferramentas para análise do mesmo. O objetivo é extrair informações úteis a partir de coleções de documentos. Estas informações são identificadas através de padrões interessantes nos dados textuais não estruturados.

Sistemas e dispositivos inteligentes são capazes de fornecer importantes dados sobre as emoções humanas e vêm sendo muito utilizados na área da saúde, os dados são usados para monitorar pacientes com risco de desenvolver doenças de problemas mentais. Na computação afetiva estuda-se como os computadores podem reconhecer, modelar e responder às emoções humanas (dentre outros aspectos) e, desta forma, como podem expressá-las através de uma interface/interação computacional usando a análise de sentimento para identificar emoções. (NUNES, 2012).

Antes da computação afetiva identificar emoções e reações humanas através de sistemas e dispositivos, a psicologia em meados da década de 1970, através do psicólogo Paul Ekman já estudava o assunto, onde desenvolveu o *facial action coding system* (FACS). Segundo (FREITAS-MAGALHÃES, 2018), o FACS é um sistema de codificação facial que permite medir, com todo rigor científico, os movimentos musculares da face. Ekman determinou códigos para expressões faciais, sendo rotuladas inúmeras *action units*, que são ações faciais enumeradas. O sistema de Ekman influenciou a criação de diversos sistemas para o monitoramento da depressão e para medir dores em pessoas que não podiam expressar-se de forma verbal.

Com base no que foi apresentado, esse trabalho apresenta a solução inteligente consciente de emoção baseada em Mineração de Textos para prevenção da depressão gestacional (SEMT), com o objetivo de auxiliar o profissional de saúde a identificar emoções da paciente através de textos assim evitando e/ou diagnosticando de forma prematura a depressão gestacional. A ideia é que a solução possa identificar possíveis alterações de humor durante o período gestacional da paciente e que o profissional da área consiga mais informações sobre sua paciente, facilitando seu diagnóstico. A solução baseia-se nas teorias de Paul Ekman¹, relacionando alguns sentimentos que podem ser identificados através da mineração de texto.

¹ Psicólogo americano, e pioneiro no estudo de emoções

1.1 Motivação

A depressão é uma doença que muitas pessoas não levam tão a sério quanto deveriam, acham que aquela pessoa que apresenta um quadro de depressão pode sair a qualquer momento, é só querer, mas é bem mais difícil do que realmente parece, a depressão pode levar até a morte é por isso que essas pessoas precisam de um tratamento adequado. A motivação vem em saber que a prevenção é mais fácil do que a recuperação, pensando assim é proposto o SEMT (Solução consciente de Emoção baseada em Mineração de Textos para prevenção da depressão gestacional) essa solução é feita para auxiliar profissionais da saúde em gestantes um grupo de pessoas com um alto grau de risco de desenvolver a depressão.

1.2 Objetivos

1.2.1 *Objetivo Geral*

Desenvolver e avaliar uma solução baseada em mineração de texto para análise de sentimentos, a partir do uso de uma base de dados reais de gestantes.

1.2.2 *Objetivos Específicos*

- Desenvolver uma solução de análise de sentimentos que reconheça emoções em textos.
- Obter o desempenho do modelo superior a 80% em todas métricas avaliadas.
- Obter um desempenho satisfatório da solução proposta a partir de textos de gestantes.

1.3 Organização do Trabalho

O restante do trabalho está organizado da seguinte forma. O Capítulo 2 apresenta a fundamentação teórica, descrevendo os principais conceitos necessários para a pesquisa deste trabalho. O Capítulo 3 apresenta os trabalhos relacionados ao estudo, fazendo um comparativo entre eles. O Capítulo 4 apresenta a metodologia da pesquisa proposta, caracterizando todo o processo necessário para o desenvolvimento do modelo preditivo. O Capítulo 5 apresenta os resultados obtidos nos experimentos descritos no Capítulo anterior. Finalmente, o capítulo 6 apresenta as considerações finais do trabalho e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

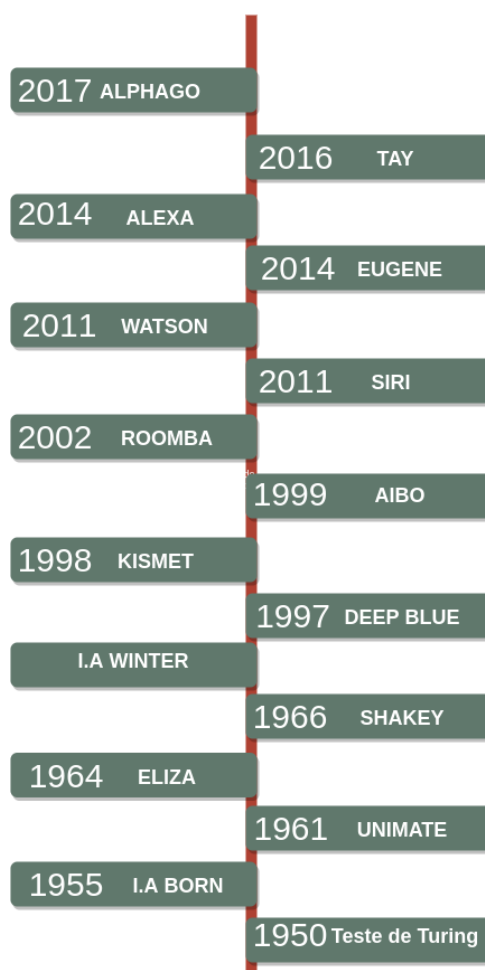
2.1 Inteligência Artificial

De acordo com (FRANCO, 2017) a inteligência artificial (IA) é a parte da ciência da computação voltada para o desenvolvimento de sistemas de computadores inteligentes, isto é, sistemas que exibem características que estão associadas à inteligência no comportamento humano, como compreensão da linguagem, aprendizado, raciocínio, resolução de problemas, entre outros. Ainda de acordo com (FRANCO, 2017), a IA vem sendo aplicada nas seguintes atividades:

- Biometria - reconhecimento facial, impressão digital e de voz;
- Fiscalização de trânsito - reconhecimento de placas de veículos infratores;
- Jogos - *Deep Blue*, determinação de comportamento de avatares;
- Diagnóstico médico - sistemas especialistas diagnosticam doenças com base em regras;
- Controle autônomo - veículos que se dirigem sem interferência humana;
- Robótica - robôs que auxiliam cirurgiões em microcirurgias;
- Pesquisa - motores de busca na internet fazem uso de técnicas diversas de IA para aumentarem sua eficiência e grau de acerto.

Segundo (NORVING; RUSSEL, 2014), os primeiros anos de IA foram repletos de sucessos, mas de uma forma limitada. Considerando-se os primitivos computadores, as ferramentas de programação da época e o fato de que apenas alguns anos antes os computadores eram vistos como objetos capazes de efetuar operações aritméticas e nada mais, causava surpresa o fato de um computador realizar qualquer atividade remotamente inteligente. De acordo com os autores, a IA é um dos campos mais recentes em ciências e engenharia. Seu desenvolvimento começou logo após a segunda guerra mundial, e o próprio nome foi cunhado em 1956. Juntamente com a biologia molecular, a IA é citada regularmente como "o campo em que eu mais gostaria de estar", por cientista de outras áreas. A Figura 1 apresenta uma linha do tempo que mostra algumas aplicações de destaque do uso da IA.

A Figura 1 mostra o quanto a IA, ao longo de quase 70 anos, obteve um grande avanço tecnológico desde o teste de *turing* em 1950 até o *alphago* em 2017.

Figura 1 – Linha do tempo da inteligência artificial.

Fonte: Elaborado pelo autor.

A tendência é que, a cada ano, aumente o número de aplicações com uso de IA, desde pequenas aplicações às aquelas de alto risco como, por exemplo, em saúde. A Figura 1 exemplifica estas aplicações, a saber:

- Teste de Turing (1950) - Alan Turing propôs um teste para a inteligência da máquina. Se a máquina induzir o humano a pensar que ela seria um humano então esta seria inteligente.
- IA *BORN* (1955) - O termo IA é usado pelo cientista da computação John McCarthy para descrever "a ciência e a engenharia de fabricação de máquinas inteligentes".
- *UNIMATE* (1961) - foi o primeiro robô a fazer funções humanas em trabalhos de linha de montagem.

- ELIZA (1964) - pioneira em conversação com humanos.
- SHAKEY (1966) - primeiro robô móvel capaz de raciocinar sobre suas ações.
- IA WINTER - foi um período frio da IA durante as décadas de 70 e 80, com pouco interesse e desenvolvimento. Somente na década de 90 a IA voltaria com força.
- DEEP BLUE (1997) - um computador da *International Business Machines Corporation (IBM)*, que venceu em uma partida de xadrez o atual campeão mundial, o russo Garry Kasparov.
- KISMET (1998) - foi um robô que podia reconhecer e simular emoções. Este fez parte de um experimento de computação afetiva.
- AIBO (1999) - foi o primeiro cão robô de estimação, criado pela empresa Sony. O robô adquire habilidades ao longo do tempo.
- ROMBA (2002) - foi o primeiro aspirador de pó inteligente produzido em massa pela empresa *iRobot* ele aprendia a limpar a casa.
- SIRI (2011) - um assistente virtual inteligente desenvolvido pela empresa *Apple* na qual possui uma interface de voz para integrar o *smartphone iPhone 4S*.
- WATSON (2011), um computador desenvolvido pela *IBM* que respondia perguntas. Conquistou o primeiro lugar em um programa de perguntas e respostas no valor de 1 milhão de dólares.
- EUGENE (2014), é um computador em forma de *chatbot* que passou no teste de Turing e convenceu 10 dos 30 juízes que era um garoto de 13 anos.
- ALEXA (2014), um assistente virtual desenvolvido pela empresa *Amazon* com interface de voz que pode concluir tarefas de compras.
- TAY (2016), um *chatbot* criado pela *Microsoft* para interagir nas redes sociais e se comportar como uma adolescente, quanto mais ele interagia mais inteligente ficava, no entanto foi desativado após gerar controvérsias após publicar mensagens de ódio nas redes sociais.
- ALPHAGO (2017), um computador da google que venceu o campeão mundial no jogo chinês *Go* no qual necessita de muita estratégia e raciocínio.

2.2 Aprendizado de Máquina

No final do século XX, o Aprendizado de Máquina (AM), *machine learning* em inglês, ganhava cada vez mais atenção. A grande quantidade de pesquisas sobre o

tema demonstra o interesse por essa área. Nesse período, Mitchell já definia a AM como a área de IA que objetiva desenvolver métodos computacionais sobre aprendizado bem como projetar sistemas capazes de aprender automaticamente (MITCHELL, 1999).

A AM aprende com a similaridade encontrada nos elementos, determinando padrões para um bom aprendizado. Segundo (TÓFOLI, 2014), o reconhecimento de padrões é a ciência que tem por objetivo a classificação de objetos em categorias ou classes. Desde os primórdios da computação implementar algoritmos apresenta-se como uma das mais intrigantes e desafiadoras tarefas da área da computação..

Existem entre outras três categorias bem comuns que usam técnicas de AM, entre elas estão a supervisionada, a semi-supervisionada e a não supervisionada. O aprendizado supervisionado funciona a partir de um conjunto de exemplos de treinamento, onde o algoritmo de aprendizado possa, com boa precisão, rotular novos exemplos não categorizados. Na aprendizagem semi-supervisionada apenas alguns dados são rotulados e na não supervisionada não existe um conjunto de dados rotulados.

2.2.1 Classificação

Formalmente, o objetivo da classificação supervisionada é induzir uma função F que seja capaz de mapear documentos rotulados para seus respectivos rótulos (ROSSI, 2016). A função F , também denominada de modelo de classificação ou classificador, relaciona os termos e suas frequências com cada uma das classes. O modelo de classificação induzido por meio de aprendizado indutivo supervisionado é então utilizado para prever a classe ou rotular novos dados. A indução da função F é também denominada de treinamento do modelo de classificação e o conjunto de documentos rotulados que são utilizados na indução do modelo são denominados de conjunto de treinamento.

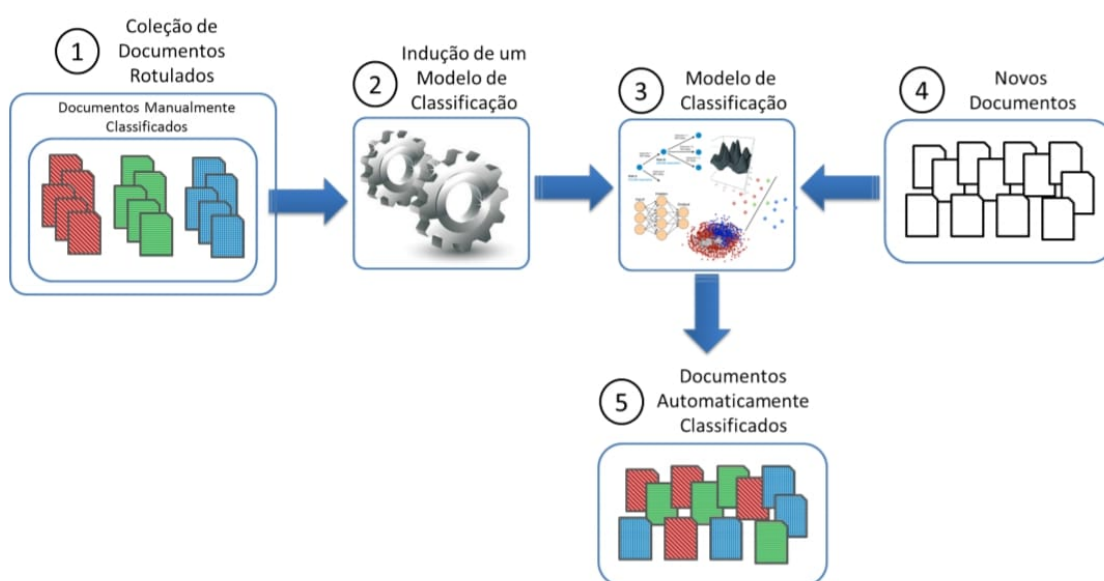
2.2.1.1 Classificação em Textos.

Segundo (NORVING; RUSSEL, 2014), a tarefa de classificação de texto, também conhecida como categorização, ou seja, dado algum tipo de texto, decide-se qual conjunto predefinido de classes pertencem a ele. A identificação da linguagem e a classificação do gênero são exemplos de classificação de texto, como também é a análise do sentimento (classificação de um filme ou revisão de produto como positivo ou negativo) e a descoberta de *spam* (classificação de uma mensagem de *e-mail* como *spam* ou não *spam*).

Atualmente, em meio a era digital, existe um tráfego imenso de textos na *Internet* que provêm de redes sociais, *e-mails*, artigos, entre outros. Estas informações estão sendo constantemente analisadas e compreendidas. A classificação de texto é usada para extrair emoções das frases, cabendo ao classificador entender as emoções inseridas no texto e fazer as relações dos atributos com suas respectivas classes a partir das características de cada palavra.

De acordo com (ROSSI, 2016), apesar de ainda haver a necessidade de intervenção humana por meio da rotulação dos documentos, essa atividade é menos custosa ou ainda mais fácil que a escrita de regras por especialistas de domínio/engenheiros de conhecimento. Além disso, os modelos de classificação construídos utilizando aprendizado indutivo supervisionado são mais fáceis de serem construídos e atualizados em comparação com a classificação manual ou a geração de regras por especialistas. Ainda de acordo com o autor, a classificação indutiva supervisionada de textos tem sido aplicada para diversas finalidades. Além disso, algoritmos desenvolvidos para a classificação indutiva supervisionada de textos também têm sido aplicados em outros domínios com características semelhantes ao domínio textual. Alguns exemplos de aplicações e do uso da classificação indutiva supervisionada de textos são filtragem e organização de notícias, organização e recuperação de documentos, recuperação de informação, classificação de *e-mails*, detecção de *spam* e resposta automática, mineração de opinião/análise de sentimentos, entre outras. A Figura 2 ilustra a classificação automática de textos por meio de aprendizado indutivo supervisionado.

Figura 2 – Classificação de textos por meio de aprendizado supervisionado.



Fonte: (ROSSI, 2016).

A figura 2 mostra cinco etapas do processo de aprendizado supervisionado em classificação de textos, são elas:

- Primeira etapa - coleção de documentos rotulados. É a base de dados a ser tratada;
- Segunda etapa - indução de um modelo de classificação. É a base de dados já rotulados (atributo-classe) usados no treinamento do modelo;
- Terceira etapa - modelo de classificação. São as técnicas de AM e o algoritmo utilizados para o tratamento e classificação dos documentos;
- Quarta etapa - novos documentos. São os documentos coletados e colocados para a predição do modelo para a classificação;
- Quinta etapa - documentos automaticamente classificados. São os documentos prontos já classificados em sua determinada classe pelo modelo.

2.2.1.2 Análise de Sentimentos

Segundo (MATURANA, 2009), as emoções não são o que correntemente chama-se de sentimento. Do ponto de vista biológico, o que se conota quando se fala de emoções são disposições corporais dinâmicas que definem os diferentes domínios de ação em que as pessoas se movem. Quando se muda de emoção, muda-se de domínio de ação. Na verdade, todos sabem isso na práxis da vida cotidiana, mas o negam, pois insiste-se que o que define a conduta das pessoas como humanas é elas serem racionais. Ao mesmo tempo, sabe-se que, quando as pessoas estão sob determinada emoção, há determinadas ações que podem ser feitas e outras que não se podem, e que se aceita como válidos certos argumentos que as pessoas não aceitariam sob outra emoção.

A mineração de emoções faz parte das técnicas de classificação, na qual é analisado um determinado texto e após serem aplicados algoritmos de AM pode-se compreender qual sentimento aquele texto possui. Geralmente, pesquisadores tomam por base as seis emoções que o psicólogo americano Paul Ekman determinou como base em seus estudos, são elas surpresa, raiva, tristeza, alegria, nojo e medo.

Tomadas de decisão são constantemente usadas para reverter uma determinada situação que esteja desfavorável. A mineração de emoções vem sendo bastante estudada para resolver situações como essas em inúmeras áreas de atuação, como no comércio por exemplo. Empresas usam a mineração de emoção para conhecer o nível de aceitação de sua marca ou de um determinado produto seu e tomar uma decisão rápida e eficaz. Outra área muito relevante em que esta técnica é utilizada é

área da saúde, principalmente relacionada com os transtornos mentais, em que busca entender de que forma os pacientes pensam e como isto influencia em seu comportamento.

2.3 Mineração de Textos

De acordo com (FELDMAN; SANGER, 2007), a mineração de texto é o processo de descobrir informações em grandes coleções de textos, identificando automaticamente padrões e relacionamentos interessantes. Esta é uma área de pesquisa relativamente nova, que recentemente despertou muito o interesse entre as comunidades de pesquisa e a indústria, principalmente devido a quantidade de informações disponíveis na Web. A mineração de texto é uma área de pesquisa altamente interdisciplinar, reunindo percepções de pesquisa das áreas de Mineração de Dados (MD), processamento de linguagem natural, AM e recuperação de informações. Em particular, a mineração de texto está intimamente relacionada à bem conhecida área de MD, que tem como alvo a extração de informações interessantes de registros de dados, embora a mineração de texto seja supostamente mais desafiadora, considerando que os dados de origem consistem em coleções não estruturados de documentos e banco de dados estruturados.

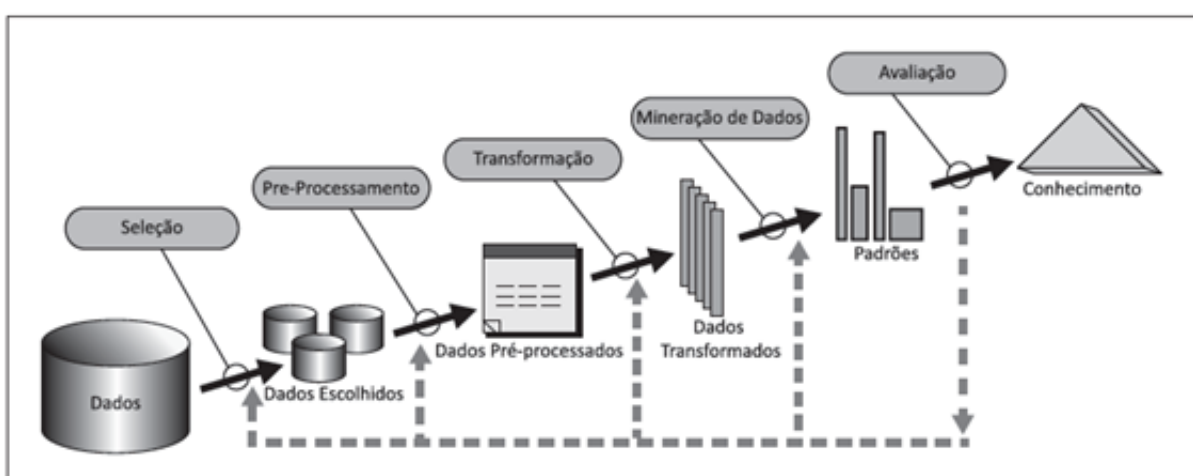
Um grande motivo de dificuldades em diversas aplicações de IA é a busca por uma autoaprendizagem de seus algoritmos. A mineração de texto consegue resolver esse problema, pois é usada para filtrar grandes quantidades de textos e extrair informações relevantes. Esta não se resume apenas a uma ferramenta de busca e sim a uma ferramenta que reconhece um padrão e uma tendência das palavras do texto. Assim, as informações relevantes tornam-se características buscadas pelos algoritmos de mineração de texto. Um exemplo comum do uso de mineração de texto trata-se de algoritmos usados para identificar se um determinado é *spam* ou não. A ideia básica é prever o e-mail através de padrões previamente conhecido pelo algoritmo.

O processo de mineração de texto engloba cinco grandes etapas. A Figura 3 apresenta o processo de *Knowledge Discovery in Databases* (KDD), que é um processo para descobrir informações de interesse em uma pesquisa de mineração de texto. A primeira etapa trata da seleção de documentos. As etapas são:

- Seleção - na primeira etapa é feita a coleta dos textos (dados) que serão processados.
- Pré-processamento - na segunda etapa é feita uma limpeza nos dados coletados preparando os dados para a etapa seguinte.

- Transformação - esta etapa é onde os dados serão analisados e processados de uma forma que a próxima etapa seja capaz de interpretá-los.
- Mineração - nessa etapa, os dados serão lidos e interpretados com a ajuda de inúmeros algoritmos de AM para extrair as informações necessárias.
- Classificação - na última etapa, onde ocorrem a avaliação e interpretação dos resultados, os dados serão analisados, filtrados e selecionados pelo usuário para ajudar com maior precisão uma tomada de decisão.

Figura 3 – Processo de *Knowledge Discovery in Databases*



Fonte: Disponível em: <https://danielteofilo.wordpress.com/2015/02/16/kdd-knowledge-discovery-in-database>. Acesso em 29-08-2019.

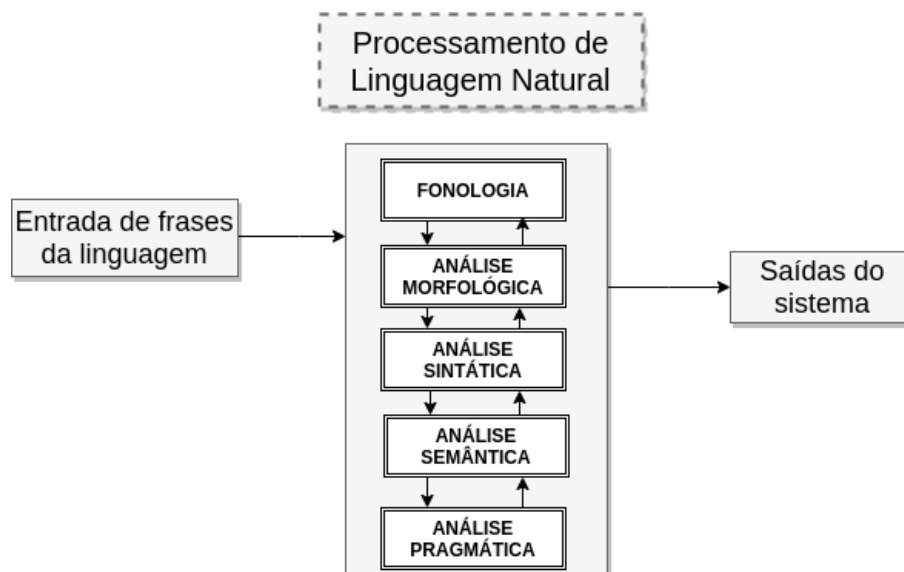
2.4 Processamento de Linguagem Natural

De acordo com (GUARNIER, 2018), o processamento de linguagem natural é uma área de pesquisa que envolve não somente IA, mas também várias outras áreas de conhecimento. Este método estuda a interação entre máquina-homem de forma que eles possam se comunicar com a linguagem natural do homem. Esse processamento está relacionado a três aspectos principais de comunicação em língua natural, que são o som, a estrutura e o significado. Na Figura 4, pode-se observar as fases de PLN a partir de uma frase.

Como observado na Figura 4, pode-se observar cinco fases:

- Fonologia - é definida como sendo o relacionamento dos sons com as palavras é usada no reconhecimento de voz;

Figura 4 – Processamento de Linguagem Natural



Fonte: Baseado em (GUARNIER, 2018).

- Análise morfológica - é responsável por analisar e busca classificar cada palavra de acordo com sua morfologia (adjetivos, substantivos e verbos) de maneira isolada;
- Análise sintática - visa analisar uma sequência de palavras a fim de analisar seu relacionamento e emprego na frase, buscando identificar sujeitos, predicados ou verbos;
- Análise semântica - busca com base nos resultados obtidos na análise sintática, realizar o mapeamento de sentenças visando atribuir seus significados;
- Análise pragmática - tem como objetivo verificar se o significado atribuído na análise semântica é o significado mais apropriado para o contexto atual (CAVALCANTE; BARBOSA, 2018).

Em PLN, existem inúmeras técnicas para o pré-processamentos de textos fazendo com que se melhore o desempenho de algoritmos de AM. Entre estas técnicas, destacam-se a *tokenização*, *stopwords* e a *stemming*.

De acordo com (AROSA, 2018), a primeira etapa na preparação de uma mensagem para utilização em algoritmo de AM é a separação por palavras. Este processo é chamado de *tokenização*, denominando-se cada palavra resultante como um *token*. A *tokenização* costuma incluir a separação de contração, como a transformação de “dela” em “de” e “ela”. *Stopwords* é uma técnica já implementada na biblioteca *Natural Language Toolkit* (NLTK) que facilita a remoção de palavras inúteis para a classificação de emoções no texto. A técnica de *stemming*, segundo (PASSINI; EBECKEN,

2012), tem o objetivo de retirar os sufixos e prefixos das palavras e encontrar a sua forma primitiva. Assim, as palavras no plural ou derivadas são reduzidas a um radical único, que é a sua raiz, simplificando a representação dos termos envolvidos no documento. Isto implica em uma única entrada nos índices, aumentando o desempenho do processo.

2.5 O Algoritmo Naive Bayes

De acordo com (DUARTE; MELO et al., 2016), o Algoritmo Naive Bayes é uma técnica de classificação baseada no Teorema de Bayes (teorema criado por Thomas Bayes, este teorema apresenta a probabilidade de um evento a partir de um conhecimento a priori que pode ter relação com o evento), com a suposição de independência entre os preditores. Em termos simples, um classificador Naive Bayes assume que a presença de uma característica particular em uma classe não está relacionada com a presença de qualquer outro recurso. Segundo (SOMBRA, 2018), o algoritmo Naive Bayes é considerado um dos algoritmos mais simples, porém completo, para classificação de dados. Seu conceito baseia-se na teoria das probabilidades que é considerada uma das principais fontes técnicas para classificação de bases de dados.

Ainda de acordo com (DUARTE; MELO et al., 2016), o Teorema de Bayes fornece uma forma de calcular a probabilidade de observações em grupos levando em considerações os dados o que é chamada de probabilidade posterior $P(c|x)$ a partir de $P(c)$, $P(x)$ e $P(x|c)$, conforme equação abaixo:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (2.1)$$

Onde,

- $P(c|x)$ é a probabilidade posterior da classe (c, alvo) dada preditor (x, atributos).
- $P(c)$ é a probabilidade original da classe.
- $P(x|c)$ é a probabilidade que representa a probabilidade do preditor dada a classe.
- $P(x)$ é a probabilidade original do preditor.

De acordo com (NORVING; RUSSEL, 2014), a principal vantagem do raciocínio probabilístico sobre o raciocínio lógico é o fato de que agentes podem tomar decisões racionais mesmo quando não existe informação o suficiente para se provar que uma ação funcionará.

3 TRABALHOS RELACIONADOS

3.1 Using a Text Mining Tool to Support Text Summarization

Em (REATEGUI; KLEMMANN; FINCO, 2012), é proposto um método específico de representação visual em forma de gráfico para auxílio a alunos em tarefas como a escrita de resumos e redações usando uma ferramenta de mineração de texto. A ferramenta Sobek, assim intitulada pelos autores, baseia-se em um desenvolvimento feito a partir de um algoritmo de distância, este algoritmo faz a extração de termos relevantes do texto analisado, conforme Figura 5.

Figura 5 – Gráficos de texto extraídos do Sobek.



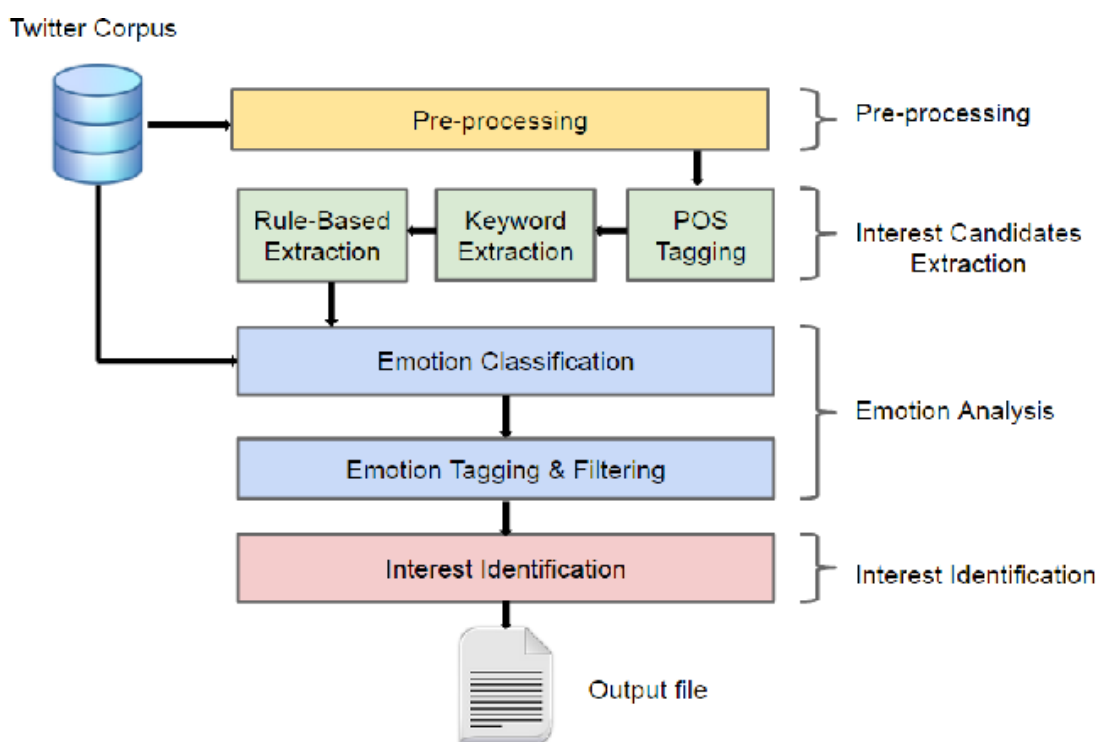
Fonte: (REATEGUI; KLEMMANN; FINCO, 2012).

Os autores dividiram o processo de escrita em três fases. Na primeira, denominada pré-escrita, o aluno lê o texto que será resumido por ele e identifica ideias necessárias para a escrita. Após essa primeira leitura, o aluno usa o Sobek para a extração de termos relevantes e relacionamentos do texto. Analisando o gráfico do texto, o aluno organiza suas ideias e faz uma releitura baseada nos termos extraídos e suas relações. Na fase seguinte, a fase de escrita, o aluno usa o gráfico de texto como comparativo, em relação ao resumo, para garantir que esteja de acordo com as ideias principais do texto. Na terceira e última fase, denominada de reescrita, o aluno revisa o texto já estruturado e escrito, facilitando a conclusão de seu trabalho. A efetividade do experimento foi verificada através de um estudo com vinte alunos do ensino médio entre 15 e 18 anos em um laboratório de informática. Segundo os resultados do experimento, os alunos usaram 61,6% dos termos extraídos do Sobek, mostrando sua relevância no auxílio do aluno na construção de resumos e redações através de mineração de texto.

3.2 EmoViz: Mining the World's Interest through Emotion Analysis

Em (SARAVIA; ARGUETA; CHEN, 2015), os autores propõem uma ferramenta para o entendimento de interesses e emoções dos usuários em proporção mundial. O sistema Emoviz é uma plataforma de visualização que consiste em identificar através das emoções um possível interesse ou tendência. Usuários do Emoviz podem inserir uma determinada palavra e verificar o interesse global em relação ao termo pesquisado. A Figura 6 mostra uma visão geral da arquitetura na qual a saída é usada para desenvolver a implementação do sistema EmoViz.

Figura 6 – Arquitetura do EmoViz.



Fonte: (SARAVIA; ARGUETA; CHEN, 2015).

A arquitetura, ilustrada anteriormente na Figura 6, é constituída por quatro fases: Na primeira fase, a de pré-processamento, são eliminadas palavras não relevantes. São extraídas apenas palavras de interesse oriundas do *Twitter*. Na segunda fase, de extração de interesse dos candidatos, são apresentadas técnicas de extração como a *ruled-based extraction*, *ketword extraction* e *POS tagging*, com o propósito de contribuir com a formação da nuvem de palavras do Emoviz. Na terceira fase, na qual é feita a análise de emoções, é rotulada a coleção de *tweets* às suas respectivas classes usando um algoritmo interno de classificação. Na última fase, a de classificação de interesse, na qual retorna uma saída com a respectiva classe de determinada palavra ou frase. A saída do conjunto de dados da arquitetura detalhada é utilizada

para implementar os componentes visuais do EmoViz, tais como mapa, gráficos e nuvem de palavras. Para a abordagem do problema, os autores utilizaram em torno de 6.000 *tweets* para identificar as emoções dos usuários em relação a cerimônia de inauguração do *iPhone 6*, em 2014.

3.3 CroCA - Cromoterapia e Computação Afetiva: auxiliando os estados de ansiedade

Em (MUNIZ; RODRIGUES; GUEDES, 2016), os autores propõem uma ferramenta Web, chamada CroCA. Essa ferramenta simula um diário *on-line* onde os usuários escrevem suas rotinas. Esses textos são processados e, caso tenha no texto traços de ansiedade, a interface da ferramenta CroCA é alterada em aspecto de cor, tendo uma resposta visual em azul para acalmar e o vermelho para provocar excitação. Os autores desenvolveram um *chatbot* capaz de retornar respostas adequadas ao interagir com o usuário. A Figura 7 exibe o protótipo da interface da ferramenta.

Figura 7 – Protótipo da ferramenta CroCA.



O protótipo da interface da ferramenta CroCA é exibido em um fundo azul claro. No topo, há um campo rotulado "Título" com um ícone de calendário e o texto "Meu dia". Abaixo dele, há um campo rotulado "Texto" com um ícone de mensagem e o texto "Eu me apavoro naquele momento, senti um aperto e uma apreensão terríveis.". Na base da interface, há dois botões: "Enviar" (em azul escuro) e "Cancelar" (em cinza).

Fonte: (MUNIZ; RODRIGUES; GUEDES, 2016).

A Figura 7 mostra o retorno da CroCA em resposta ao que foi expressado pelo usuário em forma de diário. Na situação apresentada, a ferramenta mostra uma interface azul, na qual tem a finalidade de acalmar. Para desenvolver a ferramenta os autores usaram procedimentos da computação afetiva. Em relação às cores para desenvolver comportamentos no usuário, os autores basearam-se em trabalhos científicos que ressaltam o efeito apresentado pelas cores no organismo.

3.4 Puerpério APP - uma Aplicação Móvel de Apoio à Puérpera com Depressão Pós - Parto baseada na Escala de Edimburgo

O trabalho apresentado em (BRITO, 2019) tem por objetivo desenvolver uma aplicação móvel que auxilie puérperas a fazer uma triagem com base na Escala de Edimburgo. A metodologia adotada para o desenvolvimento deste trabalho está dividida em quatro etapas: (i) pesquisa bibliográfica e proposta inicial (ii) estudo de caso (iii) desenvolvimento da aplicação e (iv) realização de testes. Espera-se que com o uso da aplicação móvel desenvolvida seja possível traçar um rastreio de depressão pós-parto do público alvo em questão.

Para o estudo de caso Brito *et al.* aplicaram um questionário a psicólogos com intuito de conhecer o contato deles com a puérpera e sua aceitação com o uso de aplicativos (BRITO, 2019). Ao analisar os dados, percebe-se que a maioria dos psicólogos atendem puérperas, sendo assim um estudo relevante no ponto de vista dos entrevistados. Ainda com base no questionário, 80% dos entrevistados concordaram plenamente no desenvolvimento do aplicativo.

O aplicativo Puerpério App, assim intitulado pelo autor do trabalho, contém seis telas, sendo elas:

- Tela Principal - essa tela descreve a principal função do aplicativo;
- Tela Questionário - nessa tela é apresentado 10 perguntas de múltipla escolha para o usuário responder;
- Tela de Resultado - nessa tela é retornado para o usuário o resultado das perguntas da tela anterior;
- Tela Histórico de Resultados - essa tela mostra todo o histórico e seus respectivos resultados em uma determinada data;
- Tela de Sobre - esta tela possui função de descrição informativa sobre depressão pós parto e da Escala de Edimburgo;
- Tela Psicólogos - nessa tela os psicólogos têm acesso aos endereços das pacientes e suas respectivas unidade de saúde.

Para avaliar o desempenho do Puerpério APP foi feito outro questionário, em que oito mães e dois psicólogos avaliaram positivamente o aplicativo.

3.5 Descrição de um sistema de suporte ao diagnóstico de demência e Transtornos Mentais Relacionados

Seixas *et al.* descreve um sistema de suporte à decisão clínica para auxiliar médicos no diagnóstico da demência e outras doenças relacionadas (SEIXAS *et al.*, 2017). O sistema permite ao médico cadastrar os registros clínicos do paciente no momento do atendimento, avaliando o diagnóstico positivo ou negativo para a demência, doença de Alzheimer e transtorno cognitivo leve e respectivo fator de certeza.

De acordo com os autores, para a avaliação das respostas do modelo de decisão, a proposta foi apresentar em uma tela que permite ao médico discordar de algum resultado apresentado pelo modelo de decisão e ainda justificar o porquê da discordância. O médico pode discordar do resultado do diagnóstico, dos registros clínicos mais relevantes que levaram ao diagnóstico ou dos registros que levaram a uma condição mais assertiva do diagnóstico.

A Figura 8 apresenta registros de um determinado paciente com a probabilidade de certeza e um *feedback* por parte do médico.

Figura 8 – Tela de registro do paciente.

The image shows two side-by-side screenshots of a mobile application interface. The left screenshot displays a diagnostic result: 'Diagnóstico positivo.' with a 'Fator de certeza: Baixo: 66.31%'. It lists 'Relação de registros clínicos que levaram a este diagnóstico: Clinical dementia rating scale, Gênero.' and 'Registros clínicos ainda não observados que poderiam confirmar o diagnóstico: Depressão, Berg balance scale, Verbal fluency test score.' At the bottom, there are two buttons: 'Concordo com o diagnóstico' (with a thumbs up icon) and 'Discordo do diagnóstico' (with a thumbs down icon). The right screenshot shows a feedback section with three items, each with a checkbox: 'O resultado do diagnóstico para Demência pode estar incorreto.', 'A lista de registros de saúde relevantes que levaram ao diagnóstico para Demência pode estar incorreta.', and 'A lista de registros de saúde relevantes ainda não observados para o diagnóstico de Demência pode estar incorreta.' Below these is a text input field labeled 'Justificativa:*' and a bottom navigation bar with standard Android icons.

Fonte: (SEIXAS *et al.*, 2017).

3.6 Comparativos da Proposta com os Trabalhos Relacionados

A Tabela 1 apresenta semelhanças e diferenças entre a proposta desta pesquisa com os trabalhos relacionados.

Tabela 1 – SEMT e trabalhos relacionados

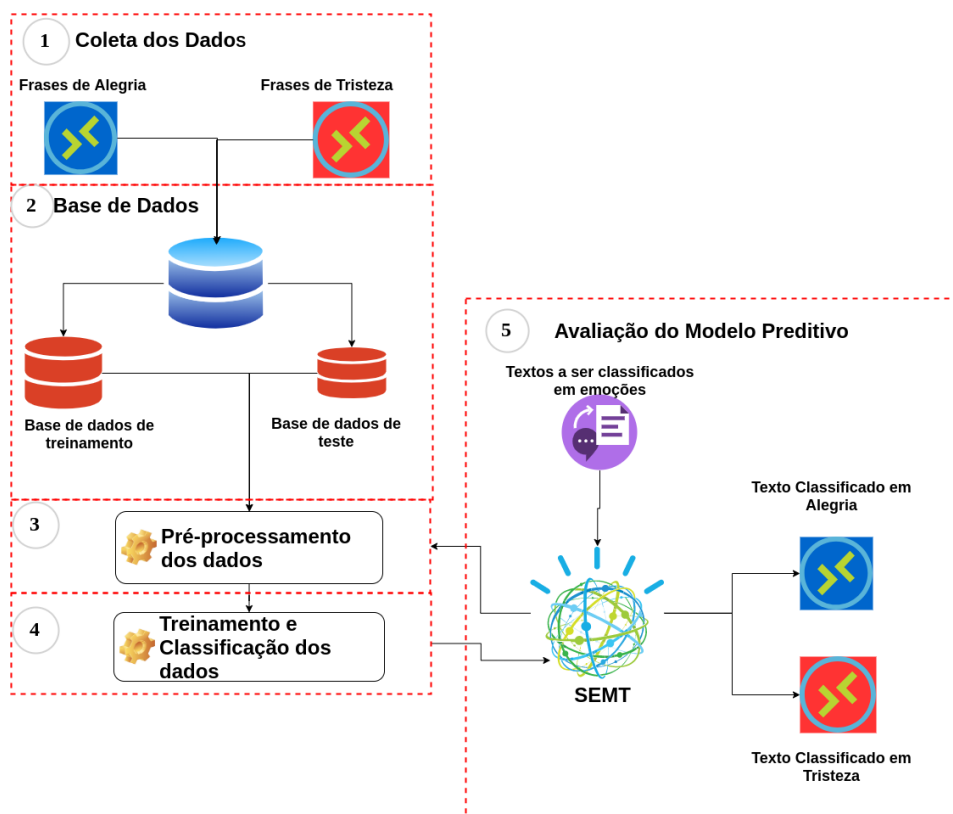
Trabalhos	Português	Análise de emoções	Dados reais	Probabilidade	saúde
Sobek		✓	✓		
EmoViz		✓	✓		
CroCA	✓	✓	✓		✓
Puerpério App	✓		✓		✓
Seixas (2017)	✓			✓	✓
SEMT	✓	✓	✓	✓	✓

Elaborado pelo autor.

4 PROPOSTA

A figura 9 mostra as etapas do fluxo de execução do modelo proposto o SEMT, desde a primeira etapa ao coletar as frases, passando pela segunda etapa ao construir a base de dados e dividindo-a em base de treinamento e teste. Em seguida, na terceira etapa são usadas técnicas de AM para o pré-processamento dos dados. Na etapa seguinte é exemplificado o treinamento dos dados. Na quinta etapa mostra-se como os dados são classificados e, finalmente, a última etapa mostra a atuação do SEMT na classificação desses textos com relação às emoções de alegria e tristeza.

Figura 9 – Etapas do Fluxo de Execução do SEMT.



Fonte: Elaborado pelo autor.

4.1 Coleta dos Dados

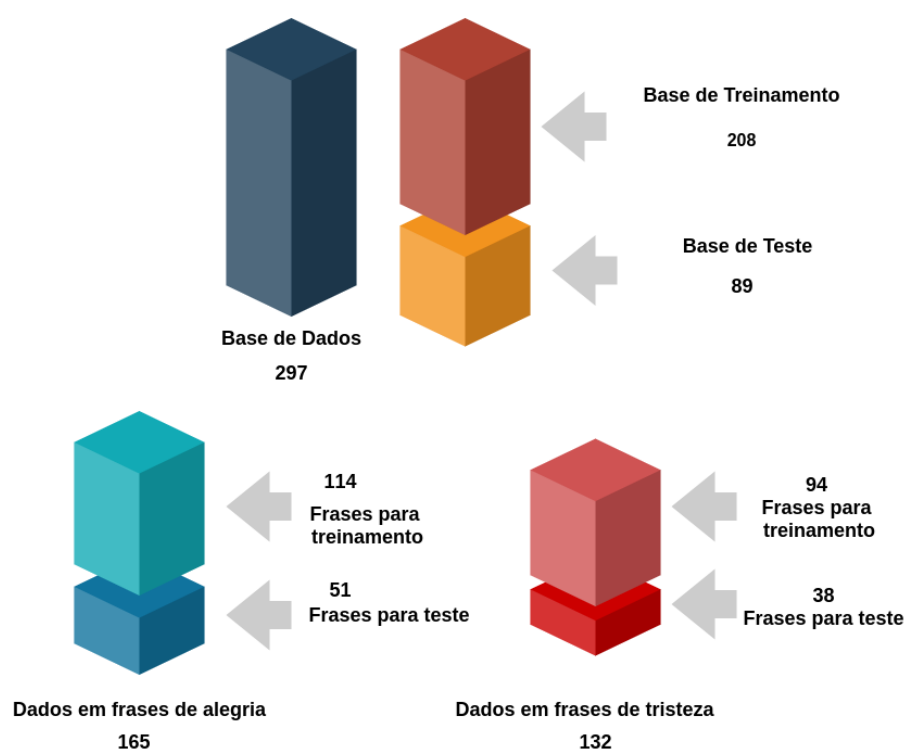
Na coleta dos dados para a construção do *dataset* foram escolhidas duas emoções entre as seis emoções primárias estudadas por *Paul Ekman*. As emoções usadas foram a alegria e a tristeza. Essas emoções são bem distintas e é esperado um alto índice de acerto por parte do SEMT em relação a sua classificação preditiva. Para a construção das frases dessas emoções foram ouvidas mulheres gestantes, mães,

técnicas de enfermagem e um aluno do último semestre do curso de psicologia da Universidade Potiguar (RN), reunindo assim um total de onze pessoas entrevistadas, que serviram de base para a construção das frases de alegria e tristeza.

4.2 Base de Dados

A base de dados foi construída em um arquivo de texto contendo 297 frases, das quais 165 frases são pertencentes à classe de alegria e 132 frases são pertencentes à classe de tristeza. A base de dados foi dividida em outras duas bases, a saber, a base de treinamento e a base de teste, como pode ser observado na Figura 10.

Figura 10 – Divisão da Base de Dados.



Fonte: Elaborado pelo autor.

Considerando a literatura, a divisão mais comum é que o conjunto de dados seja dividido em uma proporção de 70% para treinamento e 30% para testes. A Tabela 2 descreve o número de frases usadas para treinamento e teste do *dataset*.

A decisão de optar por essa proporção está relacionado ao número de registros da base de dados um número considerado baixo para um *dataset*. Caso este número fosse maior poderia-se ter usado uma proporção de 80% e 20% ou até 90% e 10% para treino e teste respectivamente, no entanto neste caso específico, ficaria um número de testes muito baixo e poderia comprometer o desempenho do modelo.

Tabela 2 – divisão de dados em treinamento e teste.

Base	Treinamento	Teste
Base completa	208	89
Base alegria	114	51
Base tristeza	94	38

Fonte: Elaborado pelo autor.

4.3 Pré-Processamento dos Dados

Nesta etapa, os dados são processados através de técnicas de AM. Neste momento é exigido um bom planejamento e processamento para obter um desempenho favorável no processo final.

4.3.1 NLTK

Para o uso de técnicas necessárias no pré-processamento dos dados é utilizada uma ferramenta para a solução proposta a biblioteca *Natural Language Toolkit* (NLTK). Segundo (BARBOSA et al., 2017), o NLTK é uma plataforma usada para desenvolver programas em linguagem *python* que trabalham com dados de linguagem humana para aplicações em PLN. Segundo os autores, o NLTK define uma infraestrutura que pode ser usada para construir programas de PLN em *python*; fornece classes básicas para representar dados relevantes para o PLN; interfaces padrão para executar tarefas como tokenização, *part-of-speech*, análise sintática e classificação de texto; e implementações padrões para cada tarefa que poderão ser combinadas para resolver problemas complexos.

4.3.2 Stopwords

As *stopwords* são palavras que não são relevantes para a compreensão das emoções no texto. Alguns exemplos entre outros são: as, de, para, com, entre, foi. Segundo (MARTINS, 2003), a eliminação de *stopwords* é um método aplicado exclusivamente a textos. Termos (palavras) que são preposições, artigos, conjunções, podem ser consideradas irrelevantes para a tarefa independentemente da tarefa a ser realizada. Porém, os termos que são *stopwords* são dependentes da língua. As *stopwords* são muito comuns e ocorrem com alta frequência na língua considerada. Dessa forma, o fato de eliminá-las do vetor de atributos faz com que haja uma drástica redução no conjunto de atributos. A tabela 3 apresenta como uma frase ficaria ao aplicar a técnica de *stopwords*.

Para obter as palavras, como vistas na tabela acima, usa-se o comando:

Tabela 3 – Aplicação de *Stopwords*.

	Frase
Com <i>stopwords</i>	Somos felizes temos muito amor.
Sem <i>stopwords</i>	felizes amor.

Fonte: elaborado pelo autor.

```
stopwordsnltk = nltk.corpus.stopwords.words('portuguese')
```

Depois remove-se as *stopwords* ao aplicar a função abaixo:

```
def removestopwords(texto):
    frases = []
    for(palavras, emocao) in texto:
        semstop = [p for p in palavras.split()
                   if p not in stopwordsnltk]
        frases.append((semstop, emocao))
    return frases
```

4.3.3 Stemming

Outra técnica de PLN usada foi a *stemming*. Esta técnica que permite eliminar os prefixos e sufixos diminui consideravelmente os espaços representativos dos dados. A Tabela 4 apresenta o uso da aplicação da técnica de *stemming*.

Tabela 4 – Aplicação de *stemming*.

	Frase
Sem <i>stemming</i>	feliz, amoroso, felicidade, triste.
Com <i>stemming</i>	feli, am, feli, trist.

Fonte: Elaborado pelo o autor.

A técnica de *stemming* também pode ser encontrada no *NLTK* e pode ser aplicada da seguinte forma:

```
def aplicastemmer(texto):
    stemmer = nltk.stem.RSLPStemmer()
    frasesestemming = []
    for (palavras, emocao) in texto:
        cste = [str(stemmer.stem(p))
                for p in palavras.split() if p not in stopwordsnltk]
        frasesestemming.append((cste, emocao))
```

```
return frasessestemming
```

4.4 Treinamento e Classificação dos Dados

Uma ferramenta importante nessa pesquisa é o algoritmo *Naive Bayes*, que realizará a tarefa de treinamento e classificação dos dados. Segundo Evangelista e Padilha, o classificador Naive Bayes é um método simples e frequentemente utilizado para o aprendizado supervisionado, fornecendo uma maneira flexível para lidar com qualquer número de atributos ou classes. Além disso, pequenas quantidades de dados ruins não atrapalham muito os resultados (EVANGELISTA; PADILHA, 2014).

O *Naive Bayes*, a partir da base de dados, cria uma tabela de probabilidade por onde ele aprende a determinar a classe de um atributo. ou seja, se um novo registro desconhecido é submetido à tabela o Naive Bayes vai retornar a probabilidade a qual classe o novo registro pertence. A Tabela 5 mostra como fica a frase pré-processada, que está pronta a iniciar o treinamento pelo algoritmo Naive Bayes.

Tabela 5 – Alegria x Tristeza.

Frase original	Frase pré-processada	classe
Minha vida está ótima, estou feliz	vi otim feli	Alegria
Estou me sentindo feliz	sent feli	Alegria
Estou com medo de ficar doente	med fi doen	Tristeza
isso me deixa com medo	deix med	Tristeza

Fonte: Elaborado pelo autor.

Com o uso do *NLTK*, é construída a tabela de probabilidades que o *Naive Bayes* usará para o treinamento e sua aprendizagem para novos atributos. Com o comando a seguir forma-se a tabela:

```
classificador = nltk.NaiveBayesClassifier.train(basecomptreinamento)
```

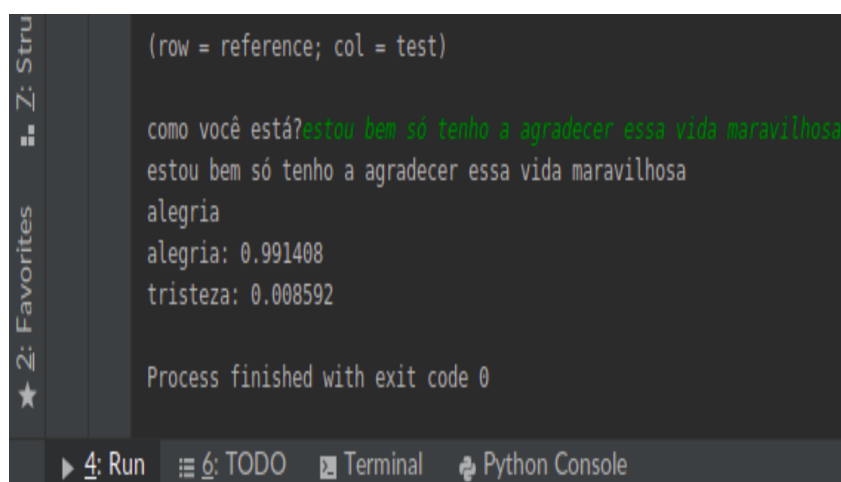
A partir desse ponto, o algoritmo já tem a informação de qual a probabilidade de cada palavra em relação a sua determinada classe e já consegue prever novos atributos desconhecidos a uma classe. O *Naive Bayes* já pode realizar a classificação de uma frase. Para que isso ocorra, é preciso fazer o seguinte comando:

```
print(classificador.classify(novo))
```

Este é o comando em que é feita a classificação dos textos em alegria ou tristeza, usando o algoritmo Naive Bayes e a função `classify`. O texto classificado com

sua respectiva emoção é determinada através de probabilidades presentes no classificador Naive Bayes. Ainda usando técnicas do *NLTK*, pode ser gerada a probabilidade da classe que foi predita como correta. A Figura 11 mostra a classificação do texto feita pelo algoritmo e a probabilidade da classe.

Figura 11 – Classificação do texto

A screenshot of a Python console window. The window title is "Z: Stru". The console output shows: "(row = reference; col = test)", "como você está?estou bem só tenho a agradecer essa vida maravilhosa", "estou bem só tenho a agradecer essa vida maravilhosa", "alegria", "alegria: 0.991408", "tristeza: 0.008592", and "Process finished with exit code 0". The console interface includes a sidebar with "Favorites" and "Z: Stru" tabs, and a bottom bar with "Run", "TODO", "Terminal", and "Python Console" buttons.

```
(row = reference; col = test)

como você está?estou bem só tenho a agradecer essa vida maravilhosa
estou bem só tenho a agradecer essa vida maravilhosa
alegria
alegria: 0.991408
tristeza: 0.008592

Process finished with exit code 0
```

Fonte:Elaborado pelo autor

Na Figura 11 o algoritmo recebe a frase: "estou bem (sic) só tenho a agradecer essa vida maravilhosa", como parâmetro e retorna como resposta que a frase pertence a classe de alegria com uma probabilidade de 99.55%, e com uma probabilidade de 0.45% de pertencer à classe de tristeza.

4.5 Avaliação do Modelo Preditivo

A avaliação do modelo preditivo é feita a partir de métricas comuns usadas na literatura. Estas métricas mostram acertos e erros para um determinado conjunto de dados. As métricas usadas como ferramentas de avaliação do modelo são detalhadas a seguir.

4.5.1 Matriz de confusão e métricas associadas

Na área de AM, e especificamente em classificação, é utilizada a matriz de confusão como uma forma de visualização do desempenho de determinado algoritmo, verificando-se se o modelo preditivo não está confundindo nenhuma classe.

Neste trabalho, o problema de classificação consiste em classificar um novo texto em alegria ou tristeza de forma correta. A tabela 6 apresenta a matriz de confusão, da qual são obtidas métricas de análise de desempenho para avaliação do

modelo preditivo. Para usar as métricas na qual são associadas a matriz de confusão devem ser analisados os dados que a matriz de confusão fornece, a saber, verdadeiros positivos (VP), verdadeiros negativos (VN), falsos negativos (FN) e os falsos positivos (FP).

Tabela 6 – Matriz de Confusão

	Classe Predita (Alegria)	Classe Predita (Tristeza)
Classe real Alegria	VP	FN
Classe real Tristeza	FP	VN

Fonte: Baseado em (DUARTE; MELO et al., 2016).

Existem quatro resultados possíveis na matriz de confusão, são eles:

- verdadeiros positivos (VP) - número de exemplos positivos classificados corretamente, no caso específico do trabalho são as frases de alegrias que são classificadas corretamente como alegria;
- verdadeiros negativos (VN) - número de exemplos negativos classificados corretamente, nesse trabalho são as frases de tristezas classificadas corretamente como tristeza;
- falsos negativos (FN) - número de exemplos positivos que são classificados incorretamente, ou seja no caso específico neste trabalho são frases de alegria que são classificadas como tristeza;
- falsos positivos (FP) - número de exemplos negativos que são classificadas incorretamente, neste trabalho são os exemplos de frases de tristezas que são classificadas como alegria.

Para analisar o desempenho são utilizadas quatro métricas: *Precisão*, *Recall*, *F-Score* e *Acurácia*. Segue abaixo cada uma dessas métricas:

- *Precisão* - esta métrica é a medida para analisar o desempenho de classificação de registros considerados positivos. É definida como o número de VP dividido pela soma de VP mais FP:

$$Precisao = \frac{VP}{VP + FP} \quad (4.1)$$

Neste caso quanto maior a precisão, maior o acerto de verdadeiros positivos.

- *Recall* - nessa métrica é analisado a taxa de VP, ou seja, o quanto o classificador prevê de forma correta os VP, assim como na *Precisão*, mas com registros de

FN. Esta métrica é definida como o número de VP dividido pela soma de VP mais FN:

$$Recall = \frac{VP}{VP + FN} \quad (4.2)$$

Um percentual baixo nessa métrica indica que o classificador classifica muitos registros positivos como FN.

- *F-Score* - esta métrica representa a média harmônica das duas anteriores de modo a chegar em um número que indica a qualidade geral do modelo. Este indicador trabalha bem até com base de dados que possuem classes desproporcionais. Esta métrica é definida como o produto entre a Precisão e o *Recall* dividido pela soma das mesmas multiplicada por dois.

$$F - Score = 2 \cdot \frac{Precisao \cdot Recall}{Precisao + Recall} \quad (4.3)$$

Com essa medida podemos dizer a performance do classificador com apenas um indicador através dessa média harmônica.

- Acurácia, essa métrica representa a taxa de acerto de todo o modelo, é definida como a soma dos VP e os VN dividido pela soma de todos os outros registros:

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.4)$$

O melhor resultado para um algoritmo, em relação aos indicadores da matriz de confusão, é apresentar 100% de acurácia. Nesse caso, apenas a diagonal principal da matriz estaria preenchida.

4.5.2 Coeficiente Kappa

O coeficiente *Kappa*, de acordo com (DUARTE; MELO et al., 2016), foi proposto por Cohen em 1960. É uma medida robusta ponderada que leva em conta os acertos e erros de acordo com a matriz de confusão. Esse coeficiente é expresso por:

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (4.5)$$

$$P_0 = \frac{\sum_{i=1}^M (n_d)}{N} \quad (4.6)$$

$$P_e = \frac{\sum_{i=1}^M (n_{lc})}{N^2} \quad (4.7)$$

Onde n_d é o número total da diagonal principal da matriz de confusão; n_l é total da linha i da matriz e n_c é total da coluna j da mesma matriz; M é total de classes da matriz e N é o número total de decisões presente na matriz.

Segundo Duarte *et al.*, este coeficiente é comumente utilizado para termos de avaliação de treinamento. A partir dele pode-se distinguir em que situação se encontra a concordância das avaliações em relação a um sistema de referência. A Tabela 9 apresenta o grau de concordância em relação ao coeficiente Kappa.

Tabela 7 – Interpretação do coeficiente Kappa

Coeficiente Kappa	Grau de concordância
< 0.00	péssimo
0.00 - 0.20	ruim
0.21 - 0.40	fraco
0.41 - 0.60	regular
0.61 - 0.80	boa
0.81 - 0.99	excelente
1.00	perfeito

Fonte: Baseado em (DUARTE; MELO *et al.*, 2016).

O Capítulo 6 apresenta os resultados do modelo preditivo e os resultados da classificação dos novos registros inserido na solução SEMT.

5 RESULTADOS

Neste capítulo são apresentados os resultados de desempenho do modelo preditivo através das métricas associadas a matriz de confusão e os resultados obtidos pelo modelo ao classificar emoções através de textos em uma nova base de dados.

5.1 Desempenho do Modelo

Para obter os resultados de desempenho do SEMT foi gerada a matriz de confusão, na figura 12, pode ser observado os indicativos para análise das métricas associadas a matriz de confusão e seus impactos no desempenho do SEMT.

Figura 12 – Matriz de confusão do SEMT.

	t	
	a	r
	l	i
	e	s
	g	t
	r	e
	i	z
	a	a
-----+-----+		
alegria	<47>	3
tristeza	9	<29>

Fonte :Elaborado pelo autor.

A matriz de confusão foi construída através do *NLTK*. Com a matriz obtida foi possível calcular os índices de desempenho do SEMT usando as métricas discutidas no capítulo anterior.

- Precisão: substituindo os dados na fórmula dessa métrica obtemos o seguinte valor:

$$Precisao = \frac{47}{47 + 9} \quad (5.1)$$

Tendo como resultado o valor de 0.8392.

- *Recall*: O resultado obtido dessa métrica foi:

$$Recall = \frac{47}{47 + 3} \quad (5.2)$$

O modelo obteve um *Recall* no valor de 0.94.

- *F-Score*: como vimos no capítulo anterior essa métrica é a média harmônica das duas anteriores, foi obtido o seguinte desempenho:

$$F - Score = 2 \cdot \frac{0.8392 \cdot 0.94}{0.8392 + 0.94} \quad (5.3)$$

O modelo obteve nessa métrica um valor de 0.8866.

- Acurácia: foi obtida no modelo a seguinte acurácia:

$$Acuracia = \frac{47 + 29}{47 + 29 + 9 + 3} \quad (5.4)$$

O SEMT apresentou um valor de 0.8636.

- Coeficiente Kappa, que apresentou o seguinte resultado:
Como visto no capítulo anterior precisamos antes calcular P_0 e P_e :

$$P_0 = \frac{47 + 29}{47 + 29 + 9 + 3} = 0.8636 \quad (5.5)$$

$$P_e = \frac{106 + 70}{176^2} = 0.0056 \quad (5.6)$$

Obtendo P_0 e P_e , pode-se usar a fórmula (4.5) e chegar ao resultado para o coeficiente.

$$K = \frac{0.8636 - 0.0056}{1 - 0.0056} = 0.8628 \quad (5.7)$$

A tabela 8 apresenta o desempenho do SEMT a partir das métricas da matriz de confusão. São apresentados os resultado em forma do valor obtido pelas fórmulas e pelo seu valor percentual. Os resultados foram considerados satisfatório para as medidas de desempenho.

Tabela 8 – Índice de desempenho do SEMT.

Métricas	Valor	valor(%)
Precisão	0.8392	83.92
<i>Recall</i>	0.94	94.00
<i>F1-Score</i>	0.8866	88.66
Acurácia	0.8636	88.36

Fonte: Elaborado pelo autor.

O resultado do SEMT em relação ao coeficiente Kappa está descrito na tabela 9, como visto no capítulo anterior. Foi obtido um resultado de 0.8626, como

pode ser observado em destaque. O modelo alcançou um desempenho com grau de concordância "excelente".

Tabela 9 – Índice Kappa do SEMT.

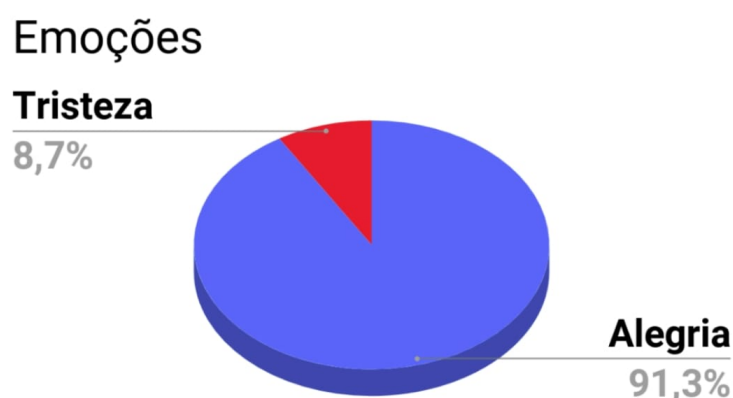
Coeficiente <i>Kappa</i>	Coeficiente obtido	Grau de concordância
<0.00	-	péssimo
0.00 - 0.20	-	ruim
0.21 - 0.40	-	fraca
0.41 - 0.60	-	regular
0.61 - 0.80	-	bom
0.81 - 0.99	0.8628	excelente
1.00	-	perfeito

Fonte: Elaborado pelo autor.

5.2 Desempenho do Modelo na Classificação dos Textos

Após o excelente resultado obtido pelo modelo SEMT para análise de sentimentos, foi aplicado um questionário entre os dias 02 e 06 de março de 2020 pelo site *SurveyMonkey*, no qual foi solicitada a participação de mulheres gestantes pelo aplicativo *WhatsApp*. Foram obtidas 23 respostas, das quais 21 tiveram resposta de alegria e 2 de tristeza. A Figura 13 mostra o percentual desses valores.

Figura 13 – Emoções



Fonte: Elaborado pelo autor.

O questionário contém apenas duas perguntas e não é preciso identificação por parte do usuário, com isso, as gestantes ficam à vontade para se expressarem, em forma de texto, o que realmente estão sentindo. O objetivo é identificar qual dos sentimentos, alegria ou tristeza, a gestante está sentindo no momento. É possível ter acesso ao questionário através de link: <https://pt.surveymonkey.com/r/>

DMC8LS6. A Figura 14 mostra a tela do aplicativo com as perguntas que as gestantes tinham acesso.

Figura 14 – Tela do questionário.


Emoções na gestação

1. Qual seu sentimento durante sua gestação?

Alegria

Tristeza

2. Descreva o por que de está com esse sentimento.



Fonte: Elaborado pelo autor.

Para validar o SEMT foram coletadas as 23 respostas e submetidas ao modelo com o objetivo de classificar as respostas na classe (emoção) correta. A seguir, apresenta-se as 23 respostas coletadas pelo questionário:

- Resposta 1 - *"Por está feliz".*
- Resposta 2 - *"Sempre foi meu sonho e agora eu realizei".*
- Resposta 3 - *"Sempre quis ser mãe, mas a notícia foi uma surpresa pq não esperava, sentir um turbilhão de sensações maravilhosas".*
- Resposta 4 - *"Porque minha gestação sempre foi planejada ,eu há queria já muito tempo , então depois de quase 1 ano tentando deus me presenteou com o maior presente das nossas vidas ,meu filho . E eu desconheço sentimento maior e puro que possa existir e o sentimento da maternidade !!! Feliz demais".*
- Resposta 5 - *"Porque saber que está gerando um filho é uma dádiva de Deus, e mim sinto muito honrada por isso, sei que meu filho só trará felicidade, já está trazendo desde que soube da notícia.".*

- Resposta 6 - *"Porque sempre quis ser mãe"*.
- Resposta 7 - *"Porque eu e meu marido sempre quisemos ter um filho"*.
- Resposta 8 - *"Porque filho é presente do Deus divino"*.
- Resposta 9 - *"Porque eu ainda n consigo me imaginar grávida, tô confusa"*.
- Resposta 10 - *"Porque é meu sonho ser mãe"*.
- Resposta 11 - *"Porque minha família toda quer muito essa criança"*.
- Resposta 12 - *"Porque eu tava na espera antes de saber, todo mês fazia o teste pra saber se tava grávida kk"*.
- Resposta 13 - *"Porque todo dia eu pedia a Deus por esse momento"*.
- Resposta 14 - *"Meu planejamento de meses deu certo kk"*.
- Resposta 15 - *"Eu e meu marido estávamos tentando pra trazer mais alegria pra casa"*.
- Resposta 16 - *"Eu sempre quis ser mãe"*.
- Resposta 17 - *"Porque é muito bom uma criança"*.
- Resposta 18 - *"Por esta esperando uma princesa linda em saber q e parte de mim"*.
- Resposta 19 - *"Porque a minha filha é um presente de Deus em minha vida"*.
- Resposta 20 - *"Porque é uma experiência única"*.
- Resposta 21 - *"Por que estou com medo devido minha idade"*.
- Resposta 22 - *"Estou me sentindo muito feliz acho que a minha família agora está completa"*.
- Resposta 23 - *"Porque ?era o meu maior sonho ser mãe hoje sinto uma alegria ,e um privilégio q nem um sentimento pode definir"*.

A emoção presente em cada frase já era conhecida por este estudo. Assim, o SEMT deve classificar a mesma emoção que as gestantes responderam. Na tabela 10 são descritas as respostas de cada gestante e sua respectiva emoção, a emoção classificada pelo SEMT e a probabilidade da frase pertencer à classe correta. Com classificações bastantes satisfatórias o SEMT obteve probabilidades bastante altas, chegando a classificar 16 das 23 respostas com uma probabilidade superior a 80% e tendo como classificação errada apenas a resposta 12.

Tabela 10 – Validação do SEMT.

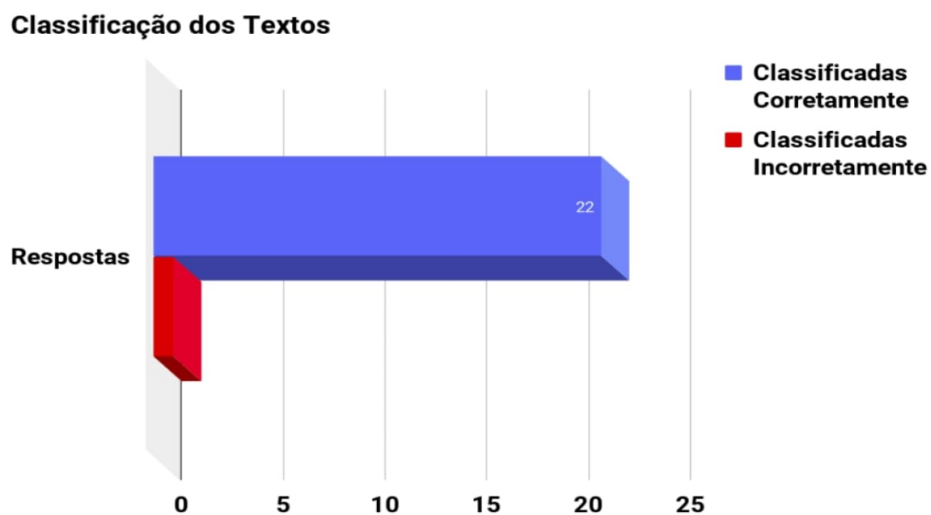
	Emoção real	Emoção classificada	%Alegria	%Tristeza
Resposta 01	Alegria	Alegria	87.80	12.20
Resposta 02	Alegria	Alegria	90.00	10.00
Resposta 03	Alegria	Alegria	99.90	00.10
Resposta 04	Alegria	Alegria	99.20	00.80
Resposta 05	Alegria	Alegria	88.00	12.00
Resposta 06	Alegria	Alegria	65.00	35.00
Resposta 07	Alegria	Alegria	84.00	16.00
Resposta 08	Alegria	Alegria	56.00	44.00
Resposta 09	Tristeza	Tristeza	00.90	99.10
Resposta 10	Alegria	Alegria	88.00	12.00
Resposta 11	Alegria	Alegria	83.00	17.00
Resposta 12	Alegria	Tristeza	15.50	84.50
Resposta 13	Alegria	Alegria	67.00	33.00
Resposta 14	Alegria	Alegria	93.00	07.00
Resposta 15	Alegria	Alegria	82.60	17.40
Resposta 16	Alegria	Alegria	65.40	34.60
Resposta 17	Alegria	Alegria	86.00	14.00
Resposta 18	Alegria	Alegria	72.00	28.00
Resposta 19	Alegria	Alegria	99.30	00.70
Resposta 20	Alegria	Alegria	56.00	44.00
Resposta 21	Tristeza	Tristeza	35.70	64.30
Resposta 22	Alegria	Alegria	99.40	00.60
Resposta 23	Alegria	Alegria	99.70	00.30

Fonte: Elaborado pelo autor.

O modelo SEMT conseguiu obter um resultado favorável, classificando os dados na classe correta. Foram 22 classificações corretas de 23 possíveis, conforme Figura 14.

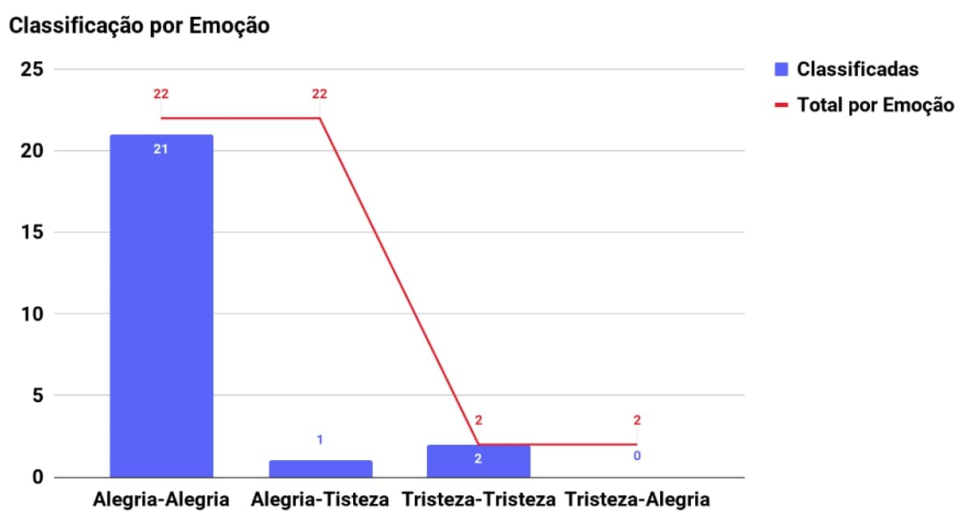
O SEMT também demonstrou ter um excelente desempenho quando analisado por classe. A classe tristeza apresentou um número muito baixo de respostas, o que dificultou a análise de sua precisão. Porém, com os dados obtidos o SEMT classificou corretamente todas as respostas da classe tristeza. A classe alegria teve um número bem maior de respostas e obteve uma classificação de quase 100% ao classificar 21 das 22 respostas direcionada a essa classe. Como podemos observar na Figura 16:

Figura 15 – Classificação dos Textos.



Fonte: Elaborado pelo autor.

Figura 16 – Classificação por Emoção.



Fonte: Elaborado pelo autor.

6 CONCLUSÃO

A mineração de textos é definida como uma área da MD que tem como principal objetivo, construir métodos para explorar um conjunto de dados textuais e extrair informações importantes. Neste trabalho são exploradas técnicas de mineração de textos e de AM em dados textuais de gestantes para realizar a prevenção da depressão gestacional.

Através do desenvolvimento deste trabalho, foi possível validar um modelo de classificação de emoções em textos, técnica esta na qual tende a crescer seu uso na medicina. O objetivo é fornecer aos profissionais de saúde mais uma ferramenta para auxiliar o diagnóstico.

Analisando os resultados de desempenho do SEMT, foram obtidos excelentes resultados através das métricas associadas à matriz de confusão. Foram obtidos os seguintes resultados: Precisão (83.92%), *Recall* (94.00%), F-Score (88.66%), Acurácia (86.36%) e coeficiente Kappa (0.8628), resultados estes que validaram o uso do SEMT para a análise de sentimentos nos textos quando submetido a uma nova base de dados. O desempenho do SEMT, em relação a uma nova base de dados foi muito satisfatória, classificando 22 dos 23 dados de forma correta.

Como trabalhos futuros, espera-se adicionar outras emoções e aumentar a base de dados, aprimorar o modelo preditivo utilizando novas técnicas de AM, construir uma aplicação *Web* para tornar o SEMT uma ferramenta de uso diário da gestante e para o profissional da saúde acompanhar melhor determinando período da gestação.

REFERÊNCIAS

- AROSA, B. V. *Análise de sentimentos em textos curtos baseada em processamento de linguagem natural e aprendizado de máquina*. 68 p. — Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, 2018. Citado na página 24.
- AZEVEDO, B. F.; BEHAR, P. A.; REATEGUI, E. B. Análise das mensagens de fóruns de discussão através de um software para mineração de textos. In: *Simpósio Brasileiro de Informática na Educação (SBIE)*. Rio de Janeiro, RJ: Sociedade Brasileira de Computação (SBC), 2012. p. 20–29. Citado na página 14.
- BARBOSA, J. L. et al. Introdução ao processamento de linguagem natural usando python. In: *III Escola Regional de Informática do Piauí (ERIP)*. Picos, PI: Sociedade Brasileira de Computação (SBC), 2017. p. 336–360. Citado na página 34.
- BRITO, N. P. *Puerpério APP-uma aplicação móvel de apoio à puerpéras com depressão pós-parto baseada na Escala de Edimburgo*. 66 p. — Universidade Federal do Amazonas (UFAM), Itacoatiara, AM, 2019. Citado na página 29.
- CAVALCANTE, P. E. C.; BARBOSA, Y. A. *Um dataset para análise de sentimentos na língua portuguesa*. 12 p. — Universidade Federal da Paraíba (UFPB), Mamanguape, PB, 2018. Citado na página 24.
- DUARTE, L. B.; MELO, I. R. et al. Análise do algoritmo naive bayes na classificação de amostras do banco de dados hepatite. In: *Congresso Nacional de Pesquisa e Ensino em Ciências (CONAPESC)*. Campina Grande, PB: [s.n.], 2016. Citado 4 vezes nas páginas 25, 38, 39 e 40.
- EVANGELISTA, T. R.; PADILHA, T. P. Monitoramento de posts sobre empresas de e-commerce em redes sociais utilizando análise de sentimentos. In: SBC. *III Brazilian Workshop on Social Network Analysis and Mining (BRASNAM)*. Brasília, DF, 2014. p. 152–163. Citado na página 36.
- FELDMAN, R.; SANGER, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge, United Kingdom: Cambridge University Press, 2007. Citado na página 22.
- FRANCO, C. R. *Inteligência Artificial*. Macapá, AP: Editora e Distribuidora Educacional, 2017. Citado na página 16.
- FREITAS-MAGALHÃES, A. *A psicologia das emoções - O fascínio do rosto humano*. Lisboa, Portugal: Leya, 2013. Citado na página 13.
- FREITAS-MAGALHÃES, A. *Facial Action Coding System 3.0 - Manual de Codificação Científica da Face Humana - FM BabyFACS*. Lisboa, Portugal: Leya, 2018. Citado na página 14.
- GUARNIER, T. S. *Processamento de Linguagem Natural (PLN)*. São Mateus, ES: Universidade Federal do Espírito Santo (UFES), 2018. Citado 2 vezes nas páginas 23 e 24.

- MARTINS, C. A. *Uma abordagem para pré-processamento de dados textuais em algoritmos de aprendizado*. 208 p. Tese (Doutorado) — Universidade de São Paulo (USP), São Carlos, SP, 2003. Citado na página 34.
- MATURANA, H. R. *Emoções e linguagem na educação e na política*. [S.l.]: Editora UFMG, 2009. Citado na página 21.
- MITCHELL, T. M. Machine learning and data mining. *Communications of the ACM*, ACM, New York, NY, USA, v. 42, n. 11, p. 30–36, 1999. Citado na página 19.
- MUNIZ, R. S.; RODRIGUES, R. G.; GUEDES, G. P. Croca - cromoterapia e computação afetiva: auxiliando os estados de ansiedade. In: SBC. *XXII Simpósio Brasileiro de Sistemas Multimídia e Web (WEBMEDIA)*. Teresina, PI, 2016. p. 145–148. Citado na página 28.
- NORVING, P.; RUSSEL, S. *Inteligência Artificial*. Upper Saddle River, NJ, EUA: Prentice Hall, 2014. Citado 3 vezes nas páginas 16, 19 e 25.
- NUNES, M. A. Computação afetiva personalizando interfaces, interações e recomendações de produtos, serviços e pessoas em ambientes computacionais. Editora UFS, São Cristóvão, SE, p. 115–151, 2012. Citado na página 14.
- PASSINI, M.; EBECKEN, N. F. *Mineração de textos para organização de documentos em centrais de atendimento*. 130 p. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, 2012. Citado na página 25.
- REATEGUI, E.; KLEMANN, M.; FINCO, M. D. Using a text mining tool to support text summarization. In: IEEE. *12th International Conference on Advanced Learning Technologies (ICALT)*. Rome, Italy, 2012. p. 607–609. Citado na página 26.
- ROSSI, R. G. *Classificação automática de textos por meio de aprendizado de máquina baseado em redes*. 285 p. Tese (Doutorado) — Universidade de São Paulo (USP), São Carlos, SP, 2016. Citado 2 vezes nas páginas 19 e 20.
- SARAVIA, E.; ARGUETA, C.; CHEN, Y.-S. Emoviz: Mining the world's interest through emotion analysis. In: IEEE/ACM. *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. Paris, France, 2015. p. 753–756. Citado na página 27.
- SCHWENGBER, D. D.; PICCININI, C. A. O impacto da depressão pós parto para a interação mãe-bebê. *Estudos de Psicologia*, SciELO Brasil, v. 8, n. 3, p. 403–411, 2003. Citado na página 13.
- SEIXAS, F. L. et al. Descrição de um sistema de suporte ao diagnóstico de demência e transtornos mentais relacionados. *Journal of Health Informatics*, v. 9, n. 3, 2017. Citado na página 30.
- SOMBRA, T. R. *Reconhecimento de padrões em rede social científica: aplicação do algoritmo Naive Bayes para classificação de papers no Mendeley*. 198 p. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, 2018. Citado na página 25.

TÓFOLI, R. J. *Casa inteligente—sistema de automação residencial*. 74 p. — Instituto Municipal de Ensino Superior de Assis (IMESA), Assis, SP, 2014. Citado na página [19](#).