



**INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE FEDERAL DE ALAGOAS
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

SAMUEL FEITOSA BATISTA

**ANÁLISE DE SENTIMENTOS EM NARRATIVAS DE GESTORES
EDUCACIONAIS COM ÊNFASE NO IDEB**

MACEIÓ

2021

SAMUEL FEITOSA BATISTA

**ANÁLISE DE SENTIMENTOS EM NARRATIVAS DE GESTORES
EDUCACIONAIS COM ÊNFASE NO IDEB**

Projeto do Trabalho de Conclusão de Curso
– TCC, apresentado ao Curso de Sistemas de
Informação da Universidade Federal de
Alagoas - UFAL, como requisito para
obtenção do grau de Bacharel em Sistemas de
Informação.

Orientador(a): Prof. Dr. Olival de Gusmão
Freitas Júnior

Coorientador: Prof. MSc. Wanderson Rubian
Martins Rodrigues

MACEIÓ

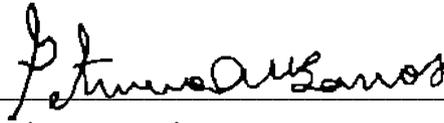
2021

SAMUEL FEITOSA BATISTA

**ANÁLISE DE SENTIMENTOS EM NARRATIVAS DE GESTORES
EDUCACIONAIS COM ÊNFASE NO IDEB**

Este Trabalho de Conclusão de Curso (TCC) foi julgado adequado para obtenção do Título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Instituto de Computação da Universidade Federal de Alagoas.

Maceió, 25 de novembro de 2021.



Prof. PETRÚCIO ANTÔNIO MEDEIROS BARROS, Msc.
Coordenador do Curso de Sistemas de Informação

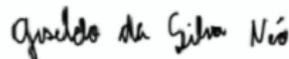
Banca Examinadora:



Prof. OLIVAL DE GUSMÃO FREITAS JÚNIOR, Dr.
Orientador



Prof. WANDERSON RUBIAN MARTINS RODRIGUES, MSc.
Coorientador



Prof. GISELDO NEO, MSc.

IFAL



Prof. BRUNO ALMEIDA PIMENTEL, Dr.
UFAL

EPÍGRAFE

“Nosso conhecimento nasce da dúvida e se alimenta da incerteza. Precisamos aprender a viver no repouso do movimento e na segurança da incerteza”.

Hilton Japiassú

AGRADECIMENTOS

Agradeço primeiramente a Deus, Autor e consumidor da minha fé. A minha família, em especial minha querida mãe, razão de minha existência, meu maior apoio e por sempre acreditar em mim. Gratidão ao meu orientador por confiar em mim, pela paciência em vários momentos de dúvidas e aos grandes ensinamentos. Aos amigos que a Universidade me proporcionou ao longo desse período. A cada um que me ajudou chegar até aqui.

RESUMO

No presente trabalho foi analisado a taxa de aprovação de alunos nas escolas públicas municipais de Maceió. O trabalho tem como objetivo principal analisar de que forma a opinião dos gestores escolares se correlaciona com essa taxa de aprovação através da Análise de Sentimentos. A Taxa de aprovação e o índice de evasão escolar comprovado pelo censo da educação básica publicado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) nos últimos anos, foi objeto de motivação da pesquisa. Este trabalho apresenta uma abordagem de análise de sentimentos que começa na coleta de opiniões através de um questionário aplicado a 12 gestores de escolas públicas municipais de Maceió. A pesquisa caracteriza-se como exploratória com abordagem quantitativa. Após a coleta de dados fazendo uso de técnicas de análises estatísticas a hipótese levantada foi testada e verificou-se uma correlação linear negativa para os dados apresentados, resultando numa inversão acerca da taxa de aprovação e sentimento do gestor.

PALAVRAS-CHAVE: Análise de Sentimentos. Estatística Descritiva. Taxa de Aprovação.

ABSTRACT

In the present work, the approval rate of students in the municipal public schools of Maceió was analyzed. The main objective of the work is to analyze how the opinion of school managers correlates with this approval rate through Sentiment Analysis. The approval rate and the school dropout rate proven by the basic education census published by the National Institute of Educational Studies and Research Anísio Teixeira (INEP) in recent years, was the object of motivation for the research. This work presents a sentiment analysis approach that begins with the collection of opinions through a questionnaire applied to 12 managers of municipal public schools in Maceió. The research is characterized as exploratory with a quantitative approach. After collecting data using statistical analysis techniques, the hypothesis was tested and a negative linear correlation was found for the data presented, resulting in an inversion of the approval rate and the manager's feelings.

KEYWORDS: Sentiment Analysis. Descriptive Statistics. Approval Rate.

LISTA DE FIGURAS

Figura 1 – Fluxograma de Aprendizagem de Máquina.....	16
Figura 2 – Indicador de fluxo.....	20
Figura 3 – Fluxograma do processo do trabalho.....	21
Figura 4 – Diagrama da Regressão Linear.....	28

LISTA DE GRÁFICOS

Gráfico 1 – Evolução do IDEB.....	18
Gráfico 2 – Média do município em relação à média do país.....	19
Gráfico 3 – Taxa de aprovação no ensino fundamental.....	20

LISTA DE QUADROS

Quadro 1 – Discrições das Variáveis.....	24
--	----

LISTA DE TABELAS

Tabela 1 – Resultados dos Sentimentos	25
Tabela 2 – Sentimento x Taxa de aprovação 2015.....	25
Tabela 3 – Descritiva dos dados.....	26
Tabela 4 – Correlação.....	27
Tabela 5 – Resultado da Regressão.....	28

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Justificativa.....	13
1.2 Hipótese.....	14
1.3 Objetivos.....	14
1.3.1 Objetivo Geral.....	14
1.3.2 Objetivos Específicos	14
2 FUNDAMENTAÇÃO TEÓRICA.....	15
2.1 Análise de Sentimento	15
2.2 IDEB.....	17
2.3 Taxa de Aprovação	19
2.3.1 Taxa de Aprovação no Ensino Fundamental	20
3 METODOLOGIA.....	21
3.1 Estatística Descritiva Utilizada Nesta Pesquisa.....	21
3.2 Amostra.....	22
3.3 Ferramentas Utilizadas.....	23
3.4 Pré-Processamento.....	23
3.5 Variáveis	24
4 RESULTADOS E DISCUSSÃO.....	25
4.1 Análise Descritiva.....	26
4.1.1 Taxa de Aprovação de 2015	26
4.1.2 Sentimento	27
4.2 Correlação.....	27
4.3 Regressão Linear.....	27
5 CONSIDERAÇÕES FINAIS.....	30
5.1 Sugestão Para Trabalhos Futuros.....	30
REFERÊNCIAS.....	31

1 INTRODUÇÃO

Com o avanço do mundo digital, cresceu também a busca por informações através de dados coletados, algo que é de suma importância para obter melhores resultados. Dessa maneira surge a Análise de Sentimento, também conhecida como Mineração de Opinião, visando tratar e analisar automaticamente os sentimentos presentes nesses dados. A Análise de Sentimento consiste em classificar documentos de texto, baseando-se no sentimento que estão presentes neles (PANG; LEE, 2008), buscando extrair informações para classificar uma sentença, ou um conjunto de sentenças, como positivas, negativas ou neutras. Essa classificação é realizada automaticamente buscando extrair informações, criando conhecimento estruturado que pode ser utilizado por um sistema computacional (FELDMAN, 2013).

Para Britto e Pacífico (2019), o principal objetivo da Análise de Sentimento é buscar por informações úteis, tais como: conhecimentos, críticas e opiniões sobre determinado tema. Na área educacional, esse termo Análise de Sentimento tem um papel de proporcionar um ensino personalizado, tentando mitigar aspectos que influenciam negativamente a motivação e o rendimento dos alunos (MAC KIM; CALVO, 2010). Em 2007, com o propósito de medir a qualidade do aprendizado e estabelecer metas para melhoria do ensino, o Ministério da Educação e Cultura (MEC), através do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), criou o Índice de Desenvolvimento da Educação Básica (IDEB).

A partir dos anos de 1990 a Política Educacional brasileira ampliou, de forma acentuada, o acesso à escola. Segundo Bomeny (2001), no início dos anos 2000, mais de 90% da população na faixa etária de 7 a 14 anos já tinha alcançado atendimento escolar, porém a qualidade da educação não acompanhou o aumento da taxa de acesso. Sendo assim, a melhoria da qualidade se constitui em um grande desafio para a atual política da educação básica do país.

1.1 Justificativa

A Análise de Sentimento vem se tornando uma área de suma importância para pesquisas científicas e comerciais, pois busca extrair sentimentos e opiniões contidas nos textos. Assim, este trabalho propõe, através de técnicas da Análise de Sentimento, coletar informações através de opiniões de 12 gestores de escolas da rede municipal de ensino, destacando a polaridade dos sentimentos extraídos nos textos, tendo como base a pesquisa iniciada em um trabalho de

mestrado realizado por Rodrigues (2020), no qual, foi necessária a aplicação de questionários para compor a base de análise.

A questão de pesquisa que orienta este estudo é: o sentimento (positivo, negativo ou neutro) dos gestores, capturados em entrevista, tem impacto na taxa de aprovação dos alunos das escolas municipais.

1.2 Hipótese

A hipótese levantada nesta pesquisa é: os gestores que utilizam emoções mais positivas têm maiores taxas de aprovação em suas escolas.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral desta pesquisa é avaliar a influência do sentimento do gestor, coletado em entrevista, na taxa de aprovação dos alunos das 12 escolas municipais em que os gestores foram entrevistados.

1.3.2 Objetivos Específicos

Os objetivos específicos são:

- a) Selecionar as ferramentas de Análise de Sentimento;
- b) Identificar os sentimentos demonstrados pelos gestores, por meio da narrativa de uma entrevista realizada *in loco*;
- c) Analisar o sentimento oriundo dessa entrevista;
- d) Avaliar a influência na taxa de aprovação dos alunos.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção aborda conceitos acerca da origem da Análise de Sentimentos, sua evolução e principais contribuições para o campo da subárea da Inteligência Artificial. Apresenta também os conceitos básicos a respeito da formação do Índice de Desenvolvimento da Educação Básica (IDEB) e sua importância.

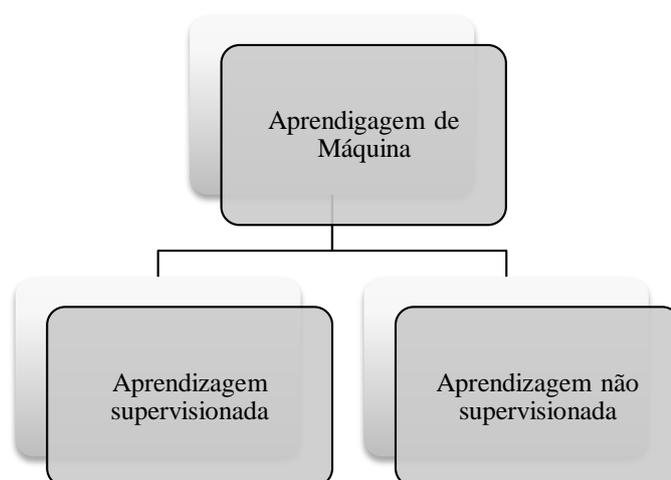
2.1 Análise de Sentimento

A Análise de Sentimento é um ramo do processamento de linguagem natural, aplicado com o objetivo de extrair, classificar e analisar opiniões em relação aos temas presentes em consideráveis volumes de dados textuais (YOON; ELHADAD; BAKKEN, 2013). Essa ferramenta objetiva classificar textos, não por conteúdos, mas pelo sentimento ou opinião, os quais estão contidos em uma base de dados (DI CARO; GRELLA, 2013). A mesma técnica é capaz de analisar opiniões, sentimentos, emoções em um texto falado ou escrito.

Ainda no século XX, surgiram alguns trabalhos relacionados com Análise de Sentimentos, com interpretações de metáforas, subjetividade, adjetivos e pontos de vista. Entretanto, o primórdio dessa subárea da Inteligência Artificial avançou no século XXI. Observa-se o crescimento da Análise de Sentimento, de acordo com os interesses simultâneos de pesquisadores, estudantes, buscando resultados concretos desse assunto. Em 2003, é possível encontrar alguns trabalhos que já mencionavam a Análise de Sentimento, como, por exemplo, o de (NASUKAWA; YI, 2003).

Essa técnica, que tem como função básica classificar automaticamente bases de dados contendo opiniões polarizadas (positivas, negativas ou neutras), auxiliada pela tecnologia da informação, diminui significativamente a necessidade de sondar elevadas quantidades de dados para colher opiniões (YU; DUAN; CAO, 2013).

Assim, a Análise de Sentimento ou Mineração de Opinião estuda as opiniões e emoções expressas em textos por meio de técnicas da Inteligência Artificial e na identificação automática da polaridade de opiniões (LIU, 2012). Na atualidade, a técnica de Análise de Sentimento é aplicada sob duas abordagens principais: aprendizagem supervisionada e aprendizagem não supervisionada (PANG; LEE, 2008). No fluxograma a seguir, podemos ver as duas abordagens citadas.

Figura 1 - Fluxograma de Aprendizagem de Máquina

Fonte: Elaborado pelo autor (2021)

Para Nascimento (2018), a técnica de rotular o modelo para a aprendizagem de máquina é denominada de aprendizagem supervisionada, porque o algoritmo de classificação precisa do modelo de dados rotulados anteriormente para ser treinado para as previsões. Segundo Becker e Tumitan (2013), uma das grandes limitações no uso de aprendizado supervisionado para definição de polaridade é a necessidade de dados rotulados para treino. O desempenho desses métodos é afetado não somente pela quantidade, mas também pela qualidade dos dados de treino disponíveis.

Já na aprendizagem de máquina não supervisionada, não há uma clareza do que será previsto, sendo que é necessário identificar as semelhanças nos dados, sem um objetivo específico. Nesses algoritmos, o rótulo da classe de cada amostra do treinamento não é conhecido, e o número, ou conjunto de classes, a ser treinado pode não ser conhecido a priori do mesmo, ou seja, não é baseado em um conjunto de treinamento (CORCOVIA; ALVES, 2019). Assim, a aprendizagem de máquina não supervisionada procura achar padrões escondidos nos dados sem respostas rotuladas.

A vantagem da aprendizagem supervisionada em relação à aprendizagem não supervisionada está no fato de que, quando as amostras de dados são classificadas, incorporam-se, ao classificador, palavras ou sentenças pertencentes ao mesmo domínio e contexto da base de dados, que será submetida à classificação automatizada, proporcionando maior precisão nos resultados (LIU; ZHANG, 2012). Por essa razão, na maioria das pesquisas, a abordagem da aprendizagem supervisionada prevalece (KONTOPOULOS et al., 2013).

2.2 IDEB

O Índice de Desenvolvimento da Educação Básica (IDEB) foi criado pelo Governo Federal por meio do Decreto nº 6094 de 24 de abril de 2007. É um índice que analisa o nível da educação básica do Brasil. Foi uma das primeiras iniciativas adotadas para medir a qualidade do aprendizado da educação no país. Visando alcançar resultados elevados, foram traçadas metas.

As metas são diferenciadas para todos: cada unidade, rede e escola, e são apresentadas bienalmente de 2007 a 2021. A última edição foi a de 2019. Logo, os estados, municípios e escolas deverão melhorar seus índices e contribuir, em conjunto, para que o Brasil alcance a meta de 6,0 pontos em 2022. O objetivo é de que cada instância evolua de forma a contribuir para que o Brasil, como um todo, atinja o patamar educacional da média dos países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). O IDEB tem o propósito de monitorar e acompanhar a qualidade das escolas de ensino fundamental, com ênfase atualmente no 5º e 9º das escolas de ensino fundamental e na 3ª série do ensino médio. Portanto, sempre no final de um ciclo da educação básica, esta é avaliada.

O IDEB visa medir a qualidade do ensino oferecido na educação e é calculado a partir de dois componentes básicos: a taxa de rendimento escolar (aprovação) e as médias de desempenho obtidas nos exames do Sistema de Avaliação da Educação Básica (SAEB). Os índices de aprovação são obtidos a partir do Censo Escolar, que é realizado anualmente.

No final do mês de junho de 2018, o MEC anunciou uma série de mudanças nas avaliações padronizadas que aplica regularmente. Todas elas serão unificadas sob o nome de SAEB. Assim, as denominações Avaliação Nacional da Alfabetização (ANA) e a Avaliação Nacional do Rendimento Escolar (ANRESC), mais conhecida como Prova Brasil, deixam de ser usadas.

O IDEB é um dos indicadores do SAEB, ou melhor, um indicador estatístico, que permite definir metas e acompanhar a qualidade do ensino básico no país, fornecendo informações referentes ao desempenho de cada escola brasileira de educação básica. A concepção básica de qualidade de ensino desse indicador é de que “o aluno aprenda e passe de ano” (FRANCO; ALVES; BONAMINO, 2007). Assim, teríamos um sistema de ensino ideal, no qual as crianças têm acesso, não apresentam distorção idade-série devido às sucessivas reprovações, não abandonam a escola e, ao final, aprendem.

Para Paz e Raphael (2010), as avaliações que compõem o SAEB são testes para diagnóstico, em larga escala, desenvolvidas pelo MEC, as quais têm o objetivo de avaliar a

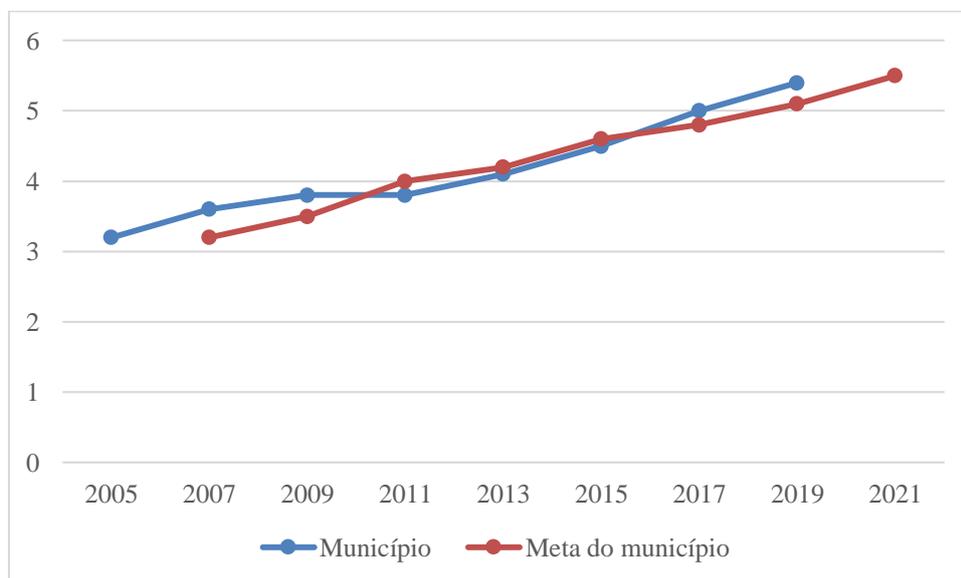
qualidade do ensino oferecido e ofertado pelo sistema educacional brasileiro a partir de testes padronizados e questionários socioeconômicos.

O IDEB apresenta-se como um dos instrumentos utilizados pelos gestores da política educacional brasileira, com o firme propósito da melhoria na qualidade da educação. Atualmente, é o principal índice capaz de aferir a qualidade do ensino nas escolas e a principal base para tomadas de decisão.

O IDEB varia de zero até 10 pontos em uma escala de qualidade, e quanto maior a nota, melhor o desempenho dos alunos e maior a regularidade no fluxo escolar. Nesse caso, o resultado é apresentado em pontos em uma escala de proficiência (Escala SAEB), posicionado e distribuído em quatro níveis: Insuficiente, Básico, Proficiente e Avançado.

O IDEB da rede pública de ensino municipal de Maceió cresceu, porém ainda não atingiu a meta de 6.0 pontos. A evolução do município em referência à meta é destacada no gráfico a seguir.

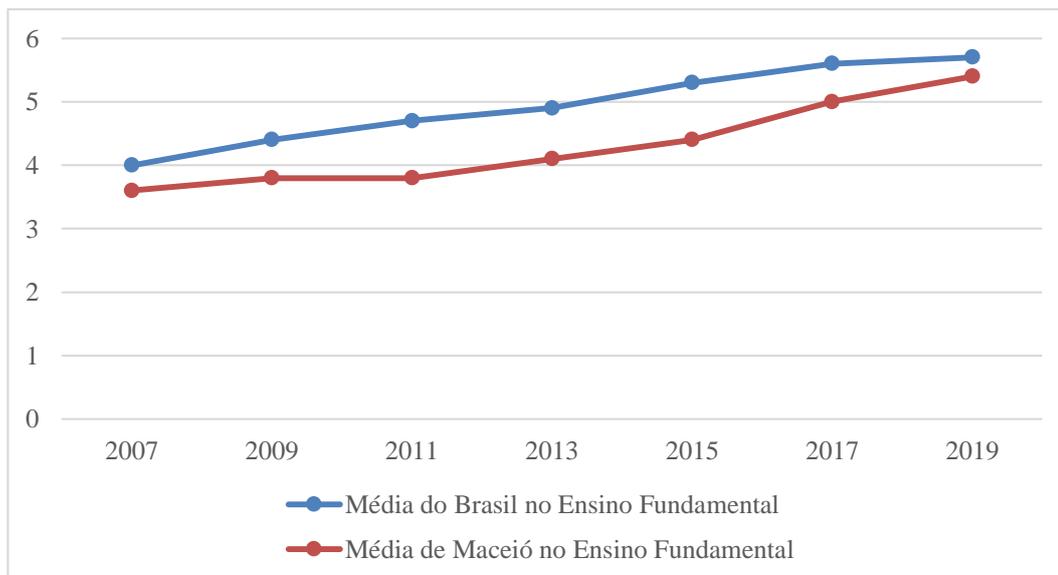
Gráfico 1 – Evolução do IDEB



Fonte: Dados do IDEB/INEP (2019)

De acordo com a edição do IDEB 2019, o município de Maceió ultrapassou a meta estabelecida, atingido 5.4 pontos, em que a meta estabelecida era de 5,1 pontos. Apesar de Maceió ultrapassar a meta proposta, ao comparamos com a média nacional nos anos iniciais do ensino fundamental, podemos perceber, no gráfico a seguir, que Maceió ainda tem uma média menor que a nacional.

Gráfico 2 - Média do município em relação à média do país



Fonte: Adaptado do QEDU (2021)

Apesar de Maceió está cumprindo sua meta, existe uma expectativa para o presente ano de 2021, não apenas nesse município, mas em todo o país devido à pandemia causada pelo Covid-19. Em razão do isolamento social e a suspensão das aulas presenciais, em especial na rede municipal, foi necessário haver mudanças drásticas no cenário escolar. O grande desafio foi fazer com que a “educação” chegasse às casas dos alunos de forma remota, de uma hora para outra, voltando o olhar de todos para o papel do computador e da informática na vida de professores e alunos (CARRARO; OSTEMBERG; DOS SANTOS, 2020). Toda a classe de profissionais da educação foi afetada. E não só os professores, mas os alunos e até as próprias instituições de ensino tiveram que se adaptar à nova realidade.

A pandemia afetou diretamente também a educação, podendo impactar o resultado do IDEB 2021, visto que o Brasil é um país de diferentes classes sociais, muitos alunos da classe baixa têm difícil acesso aos dispositivos móveis, o que pode também gerar evasão escolar. Segundo Carneiro e Vidal (2020), é evidente que a educação sofre as consequências dessa desigualdade social, impactando os direitos do cidadão ao acesso à educação.

2.3 Taxa de Aprovação

Segundo o INEP, a taxa de aprovação indica a porcentagem de alunos que, ao final do ano letivo, alcançaram os critérios mínimos para a conclusão satisfatória da etapa de ensino na qual se encontravam.

$$\text{Taxa de aprovação} = [\text{APR} / (\text{APR} + \text{REP} + \text{ABA})] \times 100.$$

APR – Número de matrículas aprovadas;
 REP – Número de matrículas reprovadas;
 ABA – Número de matrículas que deixaram de frequentar.

Figura 2 - Indicador de fluxo

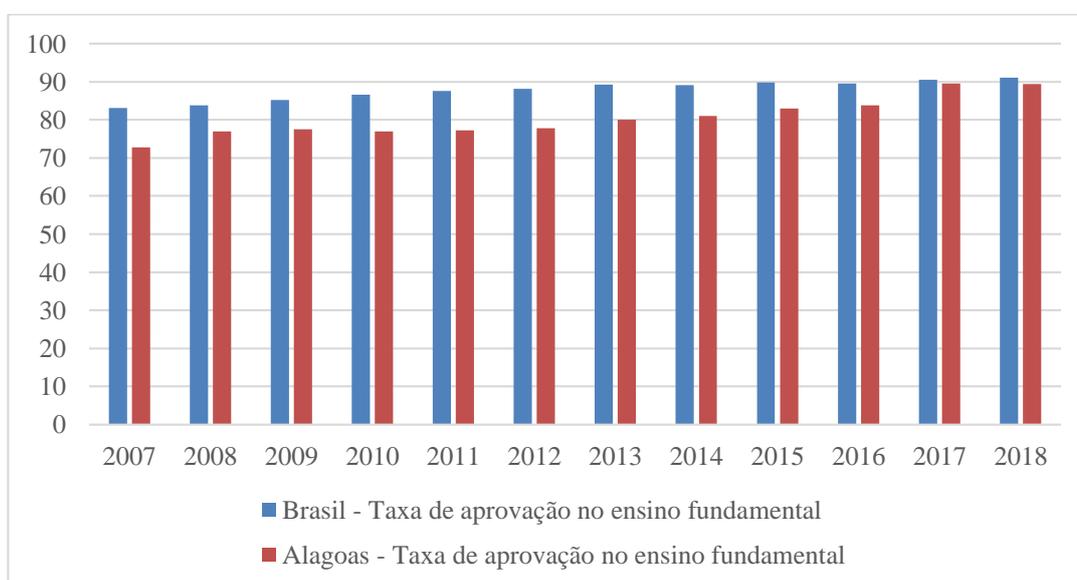


Fonte: Dados do IDEB/INEP (2019)

2.3.1 Taxa de Aprovação no Ensino Fundamental

A taxa de aprovação no ensino fundamental indica o percentual de alunos que frequentaram o ensino fundamental e progrediram para a série seguinte no final do ano letivo. Representa a condição do aluno que frequentou o ensino fundamental na mesma unidade escolar, durante o andamento do ano letivo, e foi aprovado no final desse mesmo ano. De acordo com o gráfico 3, veremos a taxa de aprovação do estado em relação ao país.

Gráfico 3 - Taxa de aprovação do ensino fundamental



Fonte: Adaptado do MEC (2021)

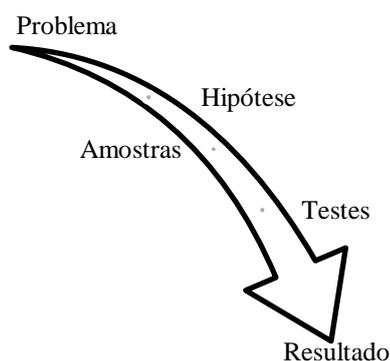
3 METODOLOGIA

A Análise de Sentimento é amplamente usada em aplicações de voz do cliente (VOC), como na análise de respostas de um questionário ou comentários em uma avaliação. Nesta pesquisa, as técnicas de Análise de Sentimento foram usadas para coletar e classificar os sentimentos e emoções detectados nos registros de áudios gravados, que foram transformados em textos. Esta seção apresenta os passos percorridos durante este trabalho, os quais foram necessários para a sua consecução.

3.1 Estatística Descritiva Utilizada Nesta Pesquisa

Devido ao problema de reprovação e evasão escolar nas escolas municipais, surge a suposição, ou seja, a hipótese de que gestores com sentimentos positivos têm taxa de aprovação maior. Neste trabalho, foi utilizado o método de pesquisa exploratória, com destaque na abordagem quantitativa. Podemos ver como o processo foi aplicado no trabalho na figura abaixo.

Figura 3 - Fluxograma do processo do trabalho



Fonte: Elaborado pelo autor (2021)

No presente trabalho foram usadas amostras, que foram as entrevistas dos gestores. Uma amostra é um subconjunto de um conjunto maior de dados. Os estatísticos chamam esse conjunto maior de população (BRUCE, A.; BRUCE, P. 2019). É preciso garantir que essas amostras sejam obtidas por processos adequados (GUIMARÃES, 2020).

Também foi analisado o desvio padrão, que é uma medida de dispersão, e o seu valor reflete a variabilidade das observações em relação à média, que visa indicar o grau de dispersão

de uma distribuição de dados (LUNET; SEVERO; BARROS, 2006). Quando se coleta o desvio e este é um resultado baixo, isso indica que está perto da média. Mas quando o resultado obtido é alto, indica que o resultado está distante. Para calcular é necessário, primeiramente, saber a variância em relação à média.

De acordo com Queiroz (2020), a correlação e regressão são métodos estatísticos que buscam entender a relação de acordo com os dados amostrais para saber se e como duas ou mais variáveis quantitativas estão relacionadas entre si. Quando existe uma relação, de alguma forma, entre duas variáveis podemos dizer que existe uma correlação entre elas. Dessa maneira, quando existe uma variação de valores em uma variável, isso causará mudança na outra variável. Logo, é gerado um diagrama de dispersão, o qual analisa esses dados correlacionais, no intuito de conferir se existe correlação entre X e Y, que são valores das duas variáveis dentro da representação gráfica.

A relação entre duas variáveis quantitativas é preferencialmente expressa por um coeficiente de correlação, os quais variam entre -1 e +1, quando mais próximo for desse e não dependem das unidades de medida das variáveis (SOUSA, 2019). Logo, quando o coeficiente for <0 , o coeficiente será negativo, e quando for >0 , será positivo. Para Canhanga (2020), se o coeficiente de correlação estiver próximo de zero, o seu valor, significa que o modelo linear não é adequado para descrever a relação entre as variáveis dependente e independente da situação.

A regressão linear é um processo que busca, por meio de uma equação, estimar a condicional da variável, visando à relação estatística. Se for necessário, ela apenas analisa a relação que ambas possuem, calcular o coeficiente já resolveria. Queiroz (2020), define o Modelo de Regressão Linear Simples provê uma equação que descreve o comportamento de uma variável em função do comportamento da outra.

3.2 Amostra

Para compor as amostras, foram selecionadas doze escolas do município de Maceió com as correspondentes taxas de aprovação do ano de 2015. Para coletar os dados, foram escolhidas doze escolas da rede pública, de forma aleatória, na capital metropolitana do estado.

Posteriormente, foram entrevistados os gestores de cada escola selecionada. Os registros de áudios gravados foram convertidos em textos e, em seguida, o texto foi traduzido para o inglês e determinado o índice de sentimento demonstrado para as doze entrevistas.

3.3 Ferramentas Utilizadas

Para traduzir os textos necessários, foi utilizado o tradutor *DeepL*, que é um serviço on-line de fácil acesso e com mecanismos interessantes, como, por exemplo, ao clicar em qualquer palavra, ele oferece sugestões de palavras semelhantes, visando oferecer um entendimento melhor do texto, sem perder o sentido do texto alterado, em qualquer idioma selecionado.

Para entender o sentimento demonstrado na linguagem natural, foi usado o *Python*, que é uma linguagem de programação de alto nível, linguagem muito conhecida e versátil. Nesse contexto, foi utilizado o algoritmo *Pattern Analyser*, que foi escolhido devido a sua velocidade e disponibilidade de uso em comparação com outros algoritmos de Análise de Sentimento (BENEVENUTO; RIBEIRO, 2015). Esse método de classificação utiliza um léxico estendido a partir do dicionário *WordNet3* (BENEVENUTO; RIBEIRO, 2015), chamado *Duch sentiment lexicon* (DE SMEDT; DAELEMANS, 2012). Esse dicionário possui aproximadamente 5.750 adjetivos anotados com seus respectivos indicadores de polaridade e foi treinado com mais de 1000 anotações manuais, os demais adjetivos foram anotados com o auxílio da técnica não supervisionada K-NN (DE SMEDT; DAELEMANS, 2012). Dessa forma, as narrativas dos gestores foram analisadas para obter os sentimentos contidos nos textos.

Além disso, foram usadas algumas outras ferramentas para comparação de sentimento, a saber: a plataforma on-line da IBM *Watson*, que disponibiliza serviços de Inteligência Artificial, pela qual foram extraídos os sentimentos dos textos coletados. Também foi utilizado o *Got It*, que também é uma ferramenta on-line, para obter os sentimentos. A diferença dessas ferramentas para a coleta dos sentimentos é que, quando usamos o algoritmo que é avaliado internamente, é como um teste de caixa-branca, em que o algoritmo é analisado, e, assim, há o acesso ao código, para efetuar ligações às bibliotecas, correções etc. Já as demais ferramentas utilizam modelo de caixa-preta, em que não há um conhecimento interno, pois foi obtido apenas o resultado e não se considera o comportamento interno que gerou o resultado do sentimento analisado. Esse tipo de teste desconsidera completamente o código, linguagem e estrutura da aplicação, o que é diferente do teste de caixa-branca (ALVES et al., 2019).

3.4 Pré-Processamento

O pré-processamento é a etapa que consiste na limpeza dos dados. Houve o tratamento nas narrativas dos diretores, ocorreram à limpeza das preposições que havia no texto, como também palavras repetitivas e duplos espaços que estavam presentes nas entrevistas. Todo esse

tratamento dos dados visa à correção de possíveis erros ortográficos, abreviaturas, *stopwords* e gírias para que não venham a interferir no resultado.

As *stopwords* são traduzidas como “palavras irrelevantes”, pois são palavras (como por exemplo, artigos, pronomes e preposições) que não vão alterar o resultado, por não trazerem relevância no tocante ao sentido, ou seja, não tem função ou papel de palavra-chave em uma sentença. Com a eliminação dessas palavras, espera-se simplificar a ideia principal da mensagem para fins de classificação do texto contido na mensagem (DRAGUT et al., 2009). No *Python*, existem pacotes que auxiliam na remoção dessas *stopwords*.

Os textos foram submetidos a uma ferramenta de análise de sentimentos utilizando a linguagem de programação *Python*, cuja biblioteca usada foi o *TextBlob* devido à sua simplicidade e velocidade. O algoritmo utilizado foi o *Pattern Analyser*. Sabendo-se que a biblioteca só suporta texto em inglês, os textos foram convertidos fazendo uso do tradutor *DeepL*. Além de ser analisado em *Python*, utilizamos outras duas plataformas, que são a *Watson* e a *Got It*, para obter os sentimentos e ver a relação entre ambos.

No processo da tradução das entrevistas, foi analisado o texto visando encontrar ironias, gírias locais, para que não venham a interferir no resultado da pesquisa, pois pode haver diferentes resultados por causa dos idiomas e gírias nacionais, que o tradutor não irá distinguir.

3.5 Variáveis

A variável explicada (variável dependente) é a taxa de aprovação dos alunos das escolas do ano de 2015. A variável explicativa é o sentimento (variável independente) coletado na entrevista. Podemos analisar no quadro proposto as descrições das variáveis usadas na pesquisa.

Quadro 1 – Discrições das Variáveis

Tipo	Nome	Domínio	Descrição	Sinal Esperado	Origem
Explicada	Taxa de aprovação	Número Real Intervalo: 0-10	Indicador da taxa de aprovação dos alunos do ano de 2015		Retirado do site www.qedu.org.br
Explicativa	Sentimento	Número Real Intervalo: -1, 0, 1.	Sentimento coletado na entrevista	Positivo	Transcrição das entrevistas realizadas na escola

Fonte: Elaborado pelo autor (2021)

4 RESULTADOS E DISCUSSÃO

Na tabela abaixo, observa-se o resultado das análises de sentimentos no *Got it*, *Pattern Analyser* e *Watson*. Diante das três análises, foi notória a diferença entre elas. Assim sendo, usaremos o sentimento resultante do *Pattern Analyser*, pois temos acesso interno ao algoritmo e conseguimos entender todo o processo. Nas demais análises, temos apenas os resultados.

Tabela 1 – Resultados dos Sentimentos

Escola	<i>Got It</i>	<i>Pattern Analyser</i>	<i>Watson</i>
1	0,21	0.085	0,46
2	0,24	0.059	0,39
3	0,31	0.085	-0,35
4	0,24	0.118	0,34
5	-0,17	0.027	-0,26
6	0,29	-0.017	0,6
7	0,18	0.080	0,43
8	0,24	0.045	0,47
9	0,2	0.074	-0,39
10	0,3	0.153	0,44
11	-0,03	0.068	0,6
12	0,13	0.014	-0,29

Fonte: Elaborado pelo autor (2021)

Os dados obtidos da variável explicativa sentimento (x) e da variável explicada taxa de aprovação (y) foram organizados na tabela abaixo.

Tabela 2 – Sentimento x Taxa de aprovação de 2015

(continua)

Escola	Sentimento(x)	Taxa de aprovação 2015(y)
1	0.084820	0.81
2	0.058806	0.77
3	0.085288	0.89
4	0.117654	0.92
5	0.027232	0.87
6	-0.016667	0.87
7	0.079904	0.8
8	0.045141	0.93
9	0.074225	0.64

Tabela 2 – Sentimento x Taxa de aprovação de 2015

(conclusão)		
Escola	Sentimento(x)	Taxa de aprovação 2015(y)
10	0.152891	0.55
11	0.068095	0.88
12	0.013636	0.9

Fonte: Elaborado pelo autor (2021)

De acordo com a hipótese formulada, os gestores que utilizaram emoções mais positivas têm taxa de aprovação maior em suas escolas. No entanto, é preciso verificar se os dados coletados confirmam ou refutam essa hipótese.

4.1 Análise Descritiva

Com os dados obtidos da taxa de aprovação e sentimentos dos gestores, foram obtidos os seguintes resultados que podem ser observados na tabela a seguir: a média dos dados e desvio padrão são relativos às variáveis das taxas de aprovação e sentimento. Também na mesma tabela foi inserido o valor máximo e mínimo de cada variável e, diante disso, podemos ver se a média se aproxima mais do máximo ou do mínimo.

Tabela 3 – Descritiva dos dados

Nome da Variável	Nº de observações	Média	Desvio Padrão	Min	Máx
Taxa de aprovação	12	0.820	0.117	0.55	0.93
Sentimento	12	0.065	0.045	-0.017	0.153

Fonte: Elaborado pelo autor (2021)

4.1.1 Taxa de Aprovação de 2015

De acordo com a tabela 3, que descreveu os dados, a média da taxa de aprovação das escolas pesquisadas corresponde a 0,82, ou em termos percentuais 82%, ou seja, de cada 100 alunos apenas 18 são reprovados. Ao mesmo tempo em que o desvio padrão corresponde a 0,11, observa-se que a taxa de aprovação se desvia de 11% da média obtida. O valor mínimo da taxa

de aprovação da pior escola foi de 55%, ou seja, quase metade dos alunos foi reprovada. Por outro lado, a escola de Maceió com maior taxa de aprovação foi de 93%, isso significa que a cada 100 alunos, apenas 7 foram reprovados.

4.1.2 Sentimento

Como consta na tabela 3, a média da variável sentimento na amostra coletada é de 0,065. Nesses dados, houve a presença do sentimento negativo, sendo o seu valor de -0,017, enquanto o maior foi de 0,153. Porém, houve uma variação média de 0,045 (Desvio Padrão) entre os dados coletados da variável sentimento.

4.2 Correlação

Para Triola (1999), o Coeficiente de Correlação Linear amostral r deve estar sempre compreendido entre -1 e +1, sendo que valores próximos de -1 e +1 e valores próximos de zero demonstram maior correlação. Pode-se concluir que há ausência de correlação entre as duas variáveis x e y que foram analisadas, veja na tabela a seguir.

Tabela 4 – Correlação

	Taxa de aprovação	Sentimento
Taxa de aprovação	1.000000	-0.529223
Sentimento	-0.529223	1.000000

Fonte: Elaborado pelo autor (2021)

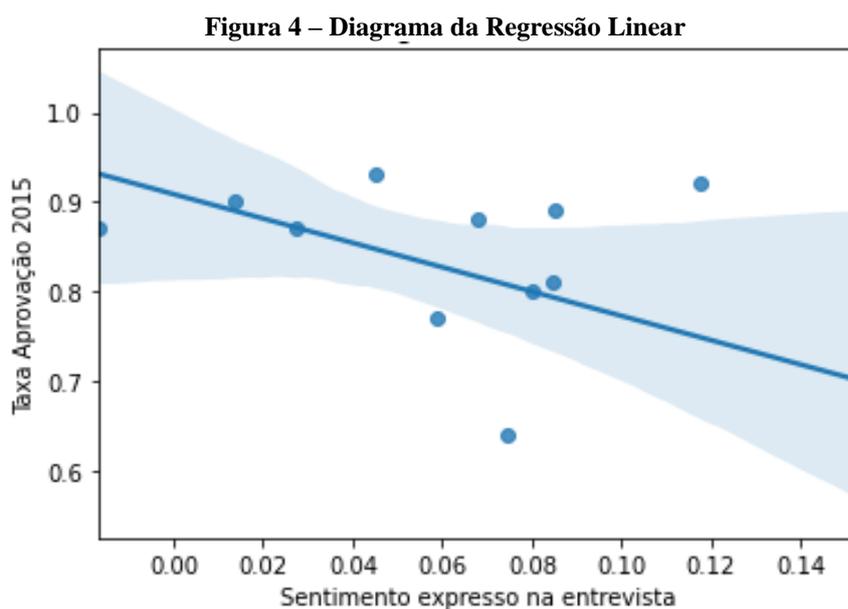
Assim, existe uma correlação média inversa entre o sentimento e a taxa de aprovação de -0.5.

4.3 Regressão Linear

A Regressão Linear Simples ela busca estimar um modelo, cujo valor da variável dependente Y é predito pela variável independente X . Logo denomina-se Regressão Linear Simples pelo fato de haver apenas uma variável independente e se considerar uma tendência linear entre Y e X . A equação abaixo é chamada equação de Regressão Linear Simples:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Em que β_0 representa a interseção, β_1 a inclinação da reta e ε o erro aleatório. Segundo Minitab (2017), essa regressão é responsável por avaliar a relação linear entre duas variáveis, sendo uma a resposta (y) e a outra preditora (x). Pela relação entre elas é possível chegar a um valor mais exato da resposta (y), levando-se em consideração o valor da preditora (x). Essa regressão torna possível encontrar a “linha reta” que melhor se ajusta aos dados, a qual pode ser utilizada para avaliar o comportamento da variável resposta, quando há mudanças na variável preditora, como se vê na figura 4.



Fonte: Elaborado pelo autor (2021)

Na tabela a seguir, é apresentado o resultado da regressão linear obtida.

Tabela 5 - Resultado da Regressão

(continua)

Resultado 1				Resultado 2			
Dep. Variable:		y		R-squared:		0.280	
Model:		OLS		Adj. R-squared:		0.208	
Method:		Least Squares		F-statistic:		3.890	
Date:		Tue, 22 Jun 2021		Prob (F-statistic):		0.0768	
Time:		14:49:46		Log-Likelihood:		11.225	
No. Observations:		12		AIC:		-18.45	
Df Residuals:		10		BIC:		-17.48	
Df Model:		1					
Covariance Type:		nonrobust					
	coef	std err	t	P> t	[0.025	0.975]	
const	0.9088	0.054	16.687	0.000	0.787	1.030	

Tabela 5 - Resultado da Regressão

(conclusão)

Resultado 1			Resultado 2			
x1	-1.3595	0.689	-1.972	0.077	-2.895	0.176
Omnibus:		0.116	Durbin-Watson:			1.771
Prob(Omnibus):		0.944	Jarque-Bera (JB):			0.239
Skew:		-0.180	Prob(JB):			0.887
Kurtosis:		2.409	Cond. No.			23.1

Fonte: Elaborado pelo autor (2021)

O modelo proposto nesta pesquisa é:

$$\text{Taxa} = \beta_0 + \beta_1 \text{ sentimento} + \varepsilon$$

Depois da realização da regressão linear, foram calculados os coeficientes β_0 e β_1 , resultando em 0,90 e -1,35, respectivamente. Assim, estabelece-se a equação com os coeficientes:

$$\text{Taxa} = 0,90 - 1,35 \text{ sentimento} + \varepsilon$$

O R^2 calculado foi de 0,28 ou 28%. Este valor evidencia que 28% das variações na Taxa de Variação são explicadas pelas variações na variável sentimento. Os demais percentuais são explicados por alterações em outras variáveis.

Conforme evidencia o modelo de regressão, rejeita-se a hipótese nula de não haver efeito do sentimento sobre a taxa de aprovação. Como o coeficiente do efeito do sentimento na taxa de aprovação foi negativo, isto é, -1,35 (p-valor = 0,07), conclui-se que existe, em média, uma relação inversa entre o sentimento demonstrado nas entrevistas dos diretores e a taxa de aprovação da escola, para todas as escolas das amostras no ano de 2015, pois a cada unidade que cresce o sentimento demonstrado nas entrevistas, decresce, em média, 13,5% a taxa de aprovação. Conforme a análise do R^2 a porcentagem de variação na resposta que é explicada pelo modelo proposto foi de 28%.

5 CONSIDERAÇÕES FINAIS

A hipótese levantada foi testada com uso das ferramentas estatísticas: Correlação, Diagrama de Dispersão e Regressão Linear Simples. Utilizou-se a técnica de Correlação, com o objetivo de analisar o grau de relacionamento entre as variáveis analisados. No entanto, o Coeficiente de Correlação Linear obtido foi um número real negativo, o que evidenciou a existência de uma relação linear inversa entre as variáveis consideradas (Sentimento e Taxa de Aprovação do ano 2015).

A equação de Regressão Linear Simples mostra que valores da variável taxa de aprovação podem ser estimados a partir de valores atribuídos a variável sentimento. Essa mesma equação mostra que para o aumento de uma unidade positiva no sentimento do gestor, a taxa de aprovação diminui 1,7% com 10% de significância, o que contraria nossa hipótese inicialmente levantada.

5.1 Sugestão Para Trabalhos Futuros

Para trabalhos futuros, pode-se utilizar uma base de entrevistas maior que 12, não apenas com gestores, mas também com alunos, funcionários e outras variáveis no modelo para identificação das emoções. Além disso, outra contribuição possível reside numa análise do F-Index das transcrições.

REFERÊNCIAS

- ALVES, Arthur Moreira et al. **Comparativo de ferramentas de automatização de testes: Selenium IDE e Selenium WebDriver**. 2019.
- BECKER, K.; TUMITAN, D. **Introdução à mineração de opiniões: Conceitos, aplicações e desafios**. In: Lectures of the 28th Brazilian Symposium on Databases. [S.l.: s.n.], 2013.
- BENEVENUTO, Fabrício; RIBEIRO, Filipe; ARAÚJO, Matheus. Métodos para análise de sentimentos em mídias sociais. **Sociedade Brasileira de Computação**, 2015.
- BOMENY, H. **Os Intelectuais da Educação**. Jorge Zahar Editor, Rio de Janeiro, 2001.
- BRASIL. Ministério da Educação. *O Plano de Ações Financiáveis do PDE - Escola*. Disponível em: http://portal.mec.gov.br/seb/arquivos/pdf/paf_pde_esc.pdf. Acesso em: 20 jan. 2021.
- BRITTO, Larissa; PACÍFICO, Luciano. **Análise de sentimentos para revisões de aplicativos mobile em português brasileiro**. In: Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional. SBC, 2019. p. 1080-1090.
- BRUCE, Andrew; BRUCE, Peter. **Estatística Prática para Cientistas de Dados**. Alta Books, 2019.
- CARNEIRO, Claudio; VIDAL, Osvaldo Veloso. DIREITO À EDUCAÇÃO E A PANDEMIA DO COVID-19. **Ciências Jurídicas e Sociais-IURJ**, v. 1, n. 1, p. 7-18, 2020.
- CANHANGA, Pascoal Epanidi. **Modelo matemático para previsão de vendas: Regressão Linear Simples**. Cadernos do IME-Série Matemática, n. 14, p. 71-81, 2020.
- CARRARO, Marcia Regina Simpioni; OSTEMBERG, Eber; DOS SANTOS, Pricila Kohls. As tecnologias digitais na educação e nos processos educativos durante a pandemia do COVID-19. **Educação Por Escrito**, v. 11, n. 2, p. e38859-e38859, 2020.
- CORCOVIA, Lucas Oukus; ALVES, Renato Santos. APRENDIZAGEM DE MÁQUINA E MINERAÇÃO DE DADOS: avaliação de métodos de aprendizagem. **Revista Interface Tecnológica**, v. 16, n. 1, p. 90-101, 2019.
- DEEPL. **DeepL Tradutor**, 2021. Página Inicial. Disponível em: <<https://www.deepl.com/pt-BR/translator>>. Acesso em: 03 de mar. de 2021.
- DE SMEDT, Tom; DAELEMANS, Walter. Pattern for python. **The Journal of Machine Learning Research**, v. 13, n. 1, p. 2063-2067, 2012.
- DI CARO, L; GRELLA, M. (2013). **Sentiment analysis via dependency parsing**. *Computer Standards & Interfaces*, 35(5), 442-453.
- DRAGUT, E. et al. **Stop word and related problems in web interface integration**. *Proc. VLDB Endow.*, VLDB Endowment, v. 2, n. 1, p. 349–360, ago. 2009. ISSN 2150-8097.

FELDMAN, R. **Techniques and Applications for Sentiment Analysis**. Commun. ACM, v. 56, n. 4, p. 82–89, abr. 2013.

FRANCO, C.; ALVES, F. BONAMINO, A. **Qualidade do ensino fundamental: políticas, suas possibilidades, seus limites**. Educação & Sociedade, São Paulo, v. 28, n. 100, especial, p. 989-1014, out. 2007.

GOTIT, **Sentimento**, 2021. Página Inicial. Disponível em: <<https://gotit.ai/pt-br/Home/Sentimenr>>. Acesso em: 08 de abr. de 2021.

GUIMARÃES, Paulo Ricardo B. **Estatística e pesquisa de opinião**. Departamento de Estatística-Universidade Federal do Paraná, UFPR, 2020.

IDEB. Índice de Desenvolvimento da Educação Básica. Disponível em: <http://ideb.inep.gov.br/>>. Acesso em: 19 de fev. de 2021.

KONTOPOULOS, et al.(2013). **Ontology-based sentiment analysis of Twitter posts**. *Expert Systems with Applications*, 40(10), 4065-4074.

LIU, B. **Sentiment analysis and opinion mining**, Morgan and Claypool Publishers. 2012.

LIU, B; & ZHANG, L. (2012). **A survey of opinion mining and sentiment analysis**. In C. C. Aggarwal; & C. X., Zhai (Ed.), *Mining text data* (Cap. 13, pp. 415-463). Chicago: Springer.

LUNET, Nuno; SEVERO, Milton; BARROS, Henrique. **Desvio padrão ou erro padrão**. Arquivos de Medicina, v. 20, n. 1-2, p. 55-59, 2006.

MAC KIM, Sunghwan; CALVO, Rafael A. **Sentiment Analysis in Student Experiences of Learning**. In: EDM. 2010. p. 111-120.

MINITAB. (2017). **Tipos de análise de regressão**. Recuperado em 19, setembro, 2017, de <https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/types-of-regression-analyses/>

NASCIMENTO, André; REIS, Marcelo. **Análise de Sentimento de Insegurança Através do Twitter**. In: Anais do XVIII Escola Regional de Computação Bahia, Alagoas e Sergipe. SBC, 2018. p. 199-208.

NASUKAWA, T.; YI, J. (2003). **Sentiment analysis: capturing favorability using natural language processing**. In the Proceedings of the 2nd International Conference on Knowledge Capture, pp. 70-77.

QUEIROZ, Matheus Vinícius Francelino. **CORRELAÇÃO LINEAR E REGRESSÃO LINEAR SIMPLES NO CONTEÚDO DE MATEMÁTICA DO ENSINO MÉDIO**. 2020. Tese de Doutorado. UNIVERSIDADE FEDERAL DE CAMPINA GRANDE.

PANG, B.; LEE, L. (2008, janeiro). **Opinion mining and sentiment analysis**. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

PAZ, F. M.; RAPHAEL, H. S. **O IDEB e a qualidade da educação no Ensino Fundamental: fundamentos, problemas e primeiras análises comparativas.** Revista OMNIA HUMANAS, v. 3, n.1, 8-30, janeiro/julho de 2010.

QEDU. **Aprendizado Dos Alunos:** Maceió. Disponível em: <<https://www.qedu.org.br/cidade/5405-maceio/aprendizado%20dos%20alunos:%20Macei%C3%B3>>. Acesso em: 14 de mar. de 2021.

RODRIGUES, Wanderson Rubian Martins. **Uma abordagem humano computacional para identificação dos fatores que influenciam o IDEB na rede de ensino básico na visão de gestores educacionais** / Wanderson Rubian Martins Rodrigues. 2020

SOUSA, Áurea. Coeficiente de correlação de Pearson e coeficiente de correlação de Spearman: o que medem e em que situações devem ser utilizados?. **Correio dos Açores**, p. 19-19, 2019.

TEXTBLOB. **TextBlob: Simplified Text Processing**, 2021. Disponível em: <<https://textblob.readthedocs.io/en/dev/>>. Acesso em: 03 de abr. de 2021.

TRIOLA, M. F. (1999). **Introdução à estatística.** Rio de Janeiro: Livros Técnicos Científicos,1999.

WATSON, ibm. 2021. **Natural Language Understanding.** Disponível em: <<https://natural-language-understanding-demo.ng.bluemix.net/>>. Acesso em: 07 de abr. de 2021.

YOON, S; ELHADAD, N.; BAKKEN, S. (2013). **A practical approach for content mining of tweets.** *American Journal of Preventive Medicine*, 45(1), 122-129.

YU, Y; DUAN, W & CAO, Q. (2013). **The impact of social and conventional media on firm equity value: a sentiment analysis approach.** *Decision Support Systems*, 55(4), 919-926.