UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

DEPARTAMENTO DE ESTATÍSTICA E INFORMÁTICA

PROGRAMA DE PÓS–GRADUAÇÃO EM INFORMÁTICA APLICADA

ARTHUR DIEGO DE GODOY BARBOSA

ANÁLISE DO IMPACTO DA TRADUÇÃO PARA A CLASSIFICAÇÃO AUTOMATIZADA DE TEXTOS EDUCACIONAIS

RECIFE – PE

2021

ARTHUR DIEGO DE GODOY BARBOSA

ANÁLISE DO IMPACTO DA TRADUÇÃO PARA A CLASSIFICAÇÃO AUTOMATIZADA DE TEXTOS EDUCACIONAIS

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Informática Aplicada do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco, como parte dos requisitos necessários para obtenção do grau de Mestre em Informática Aplicada.

ORIENTADOR: Prof. Dr. Rafael Ferreira Leite de Mello

RECIFE – PE

2021

ARTHUR DIEGO DE GODOY BARBOSA

ANÁLISE DO IMPACTO DA TRADUÇÃO PARA A CLASSIFICAÇÃO AUTOMATIZADA DE TEXTOS EDUCACIONAIS

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Informática Aplicada do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco, como parte dos requisitos necessários para obtenção do grau de Mestre em Informática Aplicada.

Aprovada em: 31 de maio de 2021.

BANCA EXAMINADORA

Prof. Dr. Rafael Ferreira Leite de Mello (Orientador)
Universidade Federal Rural de Pernambuco - UFRPE
Departamento de Informática

Prof. Dr. Pericles Barbosa Cunha de Miranda
Universidade Federal Rural de Pernambuco - UFRPE
Departamento de Computação

Profa. Dra. Ellen Polliana Ramos Souza
Universidade Federal Rural de Pernambuco - UFRPE
Unidade Acadêmica de Serra Talhada - UAST

Aos meus pais, irmãs, familiares, amigos, orientador, professores e, principalmente, a Deus, por me apoiarem e acreditarem em mim, renovando minhas forças para superar as dificuldades.

Agradecimentos

Primeiramente quero agradecer a Jeová Deus pelo privilégio de viver essa maravilhosa oportunidade com as devidas graças e forças.

Meu agradecimento ao meu orientador Prof. Dr. Rafael Ferreira, que me acompanhou antes mesmo de ingressar no mestrado, se mostrando professor exemplar, prestativo e paciente, proporcionando uma equipe prestativa (AI Box) e instruções adequadas para seguir conhecimento auto-regulatório e conseguir sucesso nesse trabalho.

A meus pais Jackson Barbosa e Mita Godoy pela criação exemplar e dedicação que me proporcionaram. As minhas irmãs Mirella de Godoy Barbosa e Maisa Godoy pelo conforto e apoio que me deram. Aos familiares que sempre me admiraram e incentivaram, em especial: vó Rosália e vô Chico, vó Dô e vô Madeira, Cida, Mila, Léo e Marina, Deda, Edna, Renan e Renata.

Aos amigos que torceram por mim, em especial aos que não consigo imaginar minha longa passagem pelo mestrado em Recife sem lembrar de pessoas maravilhosas que me acolheram longe de casa: Ana Jéssica, Alexandre Duarte, Diego George, Leandro Silva, João Carlos e Vinícius Goes.

Aos companheiros de jornada do curso, principalmente André Marverik, Anderson Cavalcanti, Elaine Marques, Paulo César Marques, Amarildo, Josival, Pedro, Gustavo, Halisson e Ameliara. Agradeço também aos envolvidos no curso de Sistemas de Informação da UFRPE Serra Talhada, em especial Ellen Polliana, Natan Albuquerque, Danilo Costa, Berg, Sérgio Paiva, Carlos Batista, Marcelo Yuri e Glauber Pires.

Deem graças por todas as coisas. Essa é a vontade de Deus para vocês em Cristo Jesus.

(1 Tessalonicenses 5:18)

Resumo

Com a grande adoção de Ambiente Virtuais de Aprendizagem (AVA) as interações educacionais passaram a gerar grandes quantidades de dados sobre cursos e alunos, tornando em alguns casos humanamente impossível um acompanhamento adequado pelos professores. Assim, vários estudos utilizam análise automática de textos educacionais para melhorar a aprendizagem. Entre os exemplos de textos analisados na literatura estão o fórum de discussão e mensagens de *feedback*. Contudo, a análise textual fica limitada ao idioma disponível em ferramentas para extração de características. Este trabalho relata os resultados de um estudo que avaliou a eficácia do emprego de métodos de tradução automática de texto para classificação de textos educacionais. Especificamente, examinou-se mensagens de fóruns de discussão *Online*, 1.500 em português e 1747 em inglês, e mensagens de *feedback* para respostas a atividades em AVAs, 1000 em português e 1056 em inglês. Visando comparar os resultados em diferentes idiomas, a proposta foi avaliar classificadores treinados nos conjuntos de dados antes e depois da aplicação da tradução do texto. Os modelos de classificação em inglês apresentaram, com os textos originais e traduzidos, resultados (acurácia e coeficiente $kappa$) semelhantes aos dos estudos relatados na literatura. Por outro lado, a tradução para o português na maioria dos casos levou a uma diminuição no desempenho. Isso indica a viabilidade geral da abordagem proposta ao converter o texto para o inglês. Além disso, este estudo destacou a importância de diferentes características e categorias para a classificação, e as limitações dos recursos para o português como justificativas dos resultados obtidos. Como o estudo utilizou dois tradutores, o Google Translate e Microsoft Azure, também verificou-se a similaridade do desempenho entre eles.

Palavras-chave: Tradução de textos, Presença Cognitiva, *Feedback* Educacional, Classificação Automatizada, Ambiente Virtual de Aprendizagem, Fórum de Discussão *Online*.

Abstract

With the broad adoption of Learning Management Systems (LMS), the educational interactions they started to generate large amounts of data about courses and students, making in some cases humanly impossible an adequate monitoring by teachers. Therefore, several studies use automatic analysis of educational texts to improve learning. Online discussion and feedback messages are among the most explored textual resources in educational settings. However, a textual analysis is limited to the language available in feature extraction tools. This paper reports the results of a study that evaluated the effectiveness of using automatic text translation methods for classifying educational texts. Specifically, we evaluated the classification of messages from online discussion forums, 1500 in Portuguese and 1747 in English, and feedback from response to LMS's activities, 1000 in Portuguese and 1056 in English, before and after applying the text translation. Aiming to compare the results in different languages, the proposal was to evaluate classifiers trained in the data sets before and after the application of the text translation. The classification models in English presented, with the original and translated texts, results (accuracy and coefficient *kappa*) similar to those of the studies reported in the literature. On the other hand, the translation into Portuguese in most cases led to a decrease in performance. This indicates the overall feasibility of the proposed approach when converting text to English. Furthermore, this study highlighted the importance of different characteristics and categories for the classification, and the limitations of resources for Portuguese as justification for the results obtained. As the study used two translators, Google Translate and Microsoft Azure, the similarity of performance between them was also verified.

Keywords: Text Translation, Cognitive Presence, Educational Feedback, Automated Classification, Learning Management System, Online Discussion Forums.

Lista de Figuras

Lista de tabelas

Lista de Siglas

Sumário

1 Introdução

O processo de ensino e aprendizagem tem base na comunicação humana, entre aluno e professor. Ao longo dos anos, as inovações tecnológicas relevantes que produziram efeitos transformadores nas relações humanas, foram exploradas para o método de transmissão de conhecimento, desde os livros até elementos modernos como computadores e internet (SPITZBERG, 2006). Assim, o uso da Comunicação Mediada por Computador (CMC), do inglês *Computer-Mediated Communication*, diversificou os meios pelos quais a comunicação é feita, não somente por *sites*, mas por aplicativos disponíveis em *desktops, laptops* e *smartphones* (WALTHER et al., 2015).

Com o crescente aumento de acesso à essas tecnologias, que as pessoas contam para a maioria das suas atividades diárias, criou-se uma linha de pensamento para atender as necessidades acadêmicas através de portais *online* que garantam segurança, confiabilidade e precisão das informações, que sejam de fácil acesso e exploração por parte dos alunos ganhando destaque no ensino superior e diversas pesquisas do contexto comunicativo (ADZHARUDDIN; LING, 2013; TOLMIE; BOYLE, 2000; VANDERGRIFF, 2013) .

Uma das formas de CMC é a educação à distância, a qual considerada por Aretio (1994) como uma forma de diálogo entre o professor e os alunos, localizados em espaços separados, para facilitar a aprendizagem de uma forma independente e colaborativa, mediados por algum documento ou outra forma de tecnologia. Esse é um exemplo de um Ambiente Virtual de Aprendizagem (AVA), do inglês Virtual Learning Environment, uma variação do termo *Learning Management System* (LMS) (MCGILL; KLOBAS, 2009). Os AVAs consistem em plataformas pedagógicas, que promovem interações humanas, disponibiliza conteúdos de aprendizagem, avaliação de apoio e avanço da aprendizagem tradicional na escola ou no ensino superior (ADZHARUDDIN; LING, 2013; COATES et al., 2005). As formas convencionais de ensino, como aulas expositivas, leituras de textos e atividades, ainda predominam no cenário geral como as principais formas de interação entre alunos e professores, mas nos AVAs, visando promover suas finalidades específicas para o meio virtual, foram implementadas ferramentas como: chat, enquetes, atividades, fórum de discussão online, *wiki* e outras formas (FERREIRA-MELLO et al., 2019).

Muitas dessas ferramentas têm um retorno automatizado, por tratar de dados e situações quantitativas e objetivas, como respostas a questões de múltipla escolha, status de aprovação por

nota ou falta. Mas na interação direta entre professores e alunos, há ferramentas que permitem mensagens textuais complexas, que demandam tempo e conhecimento para a interpretação e continuidade da interação, entre elas se encontram os fóruns de discussão online e as avaliações de atividades.

Dentro desse contexto, os fóruns de discussão, são espaços de interação colaborativos dentro das AVA, nos quais os alunos e professores lançam, respondem e debatem mensagens sobre um determinado tema (Syerov et al., 2016). Por não ser item obrigatório do processo geral de avaliação, a interação pelos fóruns ocorre de maneira espontânea, e acaba por levantar assuntos e dúvidas que não ficaram bem esclarecidas durante a aula (Fu et al., 2017).

Essa espontaneidade e colaboratividade buscando analisar e desenvolver experiências de aprendizagem, cria uma relação comunitária entre os envolvidos, caracterizando uma Comunidade de Investigação (COI), do inglês *Community of Inquiry* (TOLMIE; BOYLE, 2000), um dos modelos pedagógicos tradicionais mais pesquisados e validados envolvendo educação à distância, por enfatizar a natureza social da aprendizagem *online* . Segundo GARRISON et al. (1999) esse modelo de aprendizagem ocorre através da interação entre três elementos essenciais: (i) *presença cognitiva*, que captura o desenvolvimento das habilidades de pensamento crítico e profundo dos alunos (GARRISON et al., 2001); (ii) *presença social*, que mede a capacidade de humanizar as relações entre os participantes de uma discussão (RICHARDSON et al., 2016); e (iii) *presença de ensino*, que foca o papel dos professores antes (design do curso) e durante (facilitação e instrução direta) a discussão (ANDERSON et al., 2001b).

Já o envio de atividades, permite identificar outra parte importante no processo de ensino aprendizagem, o *feedback*(COATES et al., 2005), uma resposta adequada dada pelo professor, auxilia os alunos a identificar áreas de melhoria de seus conhecimentos e habilidades, e buscar melhores formas de aprendizagem (SADLER, 1989). Para os os professores, o *feedback* permite identificar as necessidades dos alunos e assim adaptar suas estratégias e métodos de ensino, para melhor atender essas necessidades e garantir um impacto positivo na aprendizagem do conteúdo a ser passado. (LANGER, 2011; NICOL; MACFARLANE-DICK, 2006).

Para fornecer uma experiência completa de aprendizagem do assunto, cada aluno deverá ser amparado com respostas que complementam as necessidades que lhe foram identificadas. Mas para isso, é preciso detectar nas respostas, tanto o tipo de *feedback*, quanto o nível de presença cognitiva, o que demanda uma análise complexa de textos educacionais.

1.1 Problema de Pesquisa

O número crescente de alunos matriculados no ensino online aumenta os desafios dos professores para fornecer *feedback* informativo de alta qualidade, e monitorar o andamento dos fóruns de discussão, para auxiliar os alunos com suas necessidades no processo de aprendizagem. Percebendo essa sobrecarga do professor, pesquisas recentes foram desenvolvidas demonstrando abordagens que automatizam classificação de textos educacionais, para identificar tanto os níveis de *feedback*, quanto para presença cognitiva (Marin et al., 2017; Cheniti Belcadhi, 2016).

Anteriormente foram desenvolvidas e testadas algumas abordagens para análise de presença cognitiva que encontraram alguns problemas como: formas de análise por um meio não-automatizado, como a leitura dos debates e análises estatísticas básicas oferecidas pela plataforma para administradores e moderadores do fórum (Fu et al., 2017), algo que se demonstrou muito complexo e demorado para ser implantado em tempo real, além de outros meios que demonstraram ser muito invasivos (GARRISON et al., 2001; ARBAUGH et al., 2008). Tentando evitar esses obstáculos, outras abordagens focaram em desenvolver classificadores automáticos usando recursos extraídos de dados transcritos das mensagens dos fóruns, tanto em inglês como em português (KOVANOVIĆ et al., 2016; NETO et al., 2018).

Muitos desses trabalhos de automatizar a análise textual para gerar a classificação de *feedback* e presença cognitiva, utilizam as ferramentas Coh-Metrix[1] e LIWC [2] para geraração de features a serem utilizadas pelo classificador. Essas ferramentas originalmente foram extraídas para o idioma Inglês, e vêm recebendo adaptações para outros idiomas, como o Coh-Metrix-Port [3] e uma versão para português do LWIC pela PortLex [4]. Porém essas adaptações ainda estão em fase inicial de desenvolvimento, o que ocasiona indisponibilidade de 18 recursos, o que representa cerca de 11% de diferença em relação à ferramenta original para o idioma Inglês.

O estudo de BANEA et al. (2008), caracteriza a tradução automática como alternativa viável para a construção de recursos e ferramentas de análise de subjetividade em textos de outro idioma. A pesquisa de AMINI et al. (2009) apontou que a tradução automática pode aumentar o desempenho de classificadores quando há poucos dados rotulados disponíveis para treinamento. Alguns estudos já adotam uma abordagem em construir modelos para um idioma destino e utilizá-los para classificar documentos de outros idiomas traduzidos para o idioma destino (SHANAHAN

[1] https://cohmetrix.com/
[2] http://liwc.wpengine.com/
[3] http://143.107.183.175:22680/
[4] http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc

et al., 2004), e adotar métodos de tradução de texto antes de desenvolver o classificador (SHI et al., 2010). Também encontram-se abordagens de classificação usando recursos padrões de conjunto de palavras e contagem de palavras com base no processo psicológico (XU et al., 2016; EVANS; ACEVES, 2016).

1.2 Perguntas de Pesquisa

Conforme será apresentado no Capítulo 3, há uma diversidade de métodos e ferramentas para codificação automática de mensagens de discussão online de acordo com os níveis da presença cognitiva, e codificação automática de respostas a atividades em plataformas de ensino online de acordo com os níveis de *feedback*, sendo as duas formas aplicadas em diferentes idiomas. Entretanto, algumas línguas não possuem os recursos linguísticos necessários para aplicar a metodologia proposta pelos classificadores de última geração. Portanto, nossa primeira pergunta de pesquisa é:

> **PERGUNTA DE PESQUISA 1 (PP1):**
> *Qual a eficácia da adoção de métodos de tradução automática de textos para diferentes idiomas no processo de classificação automática de textos educacionais para níveis de feedback educacional e presença cognitiva?*

A pergunta de pesquisa 1 tem como objetivo analisar até que ponto a precisão da classificação automática pode ser alcançada quando métodos de tradução de texto são usados. Também pretende-se fornecer informações adicionais sobre as características mais relevantes para os textos originais e traduzidos. Esta análise pode fornecer *insights* sobre o impacto da limitação dos recursos linguísticos para idiomas diferentes do Inglês. Portanto, a segunda pergunta de pesquisa foi:

> **PERGUNTA DE PESQUISA 2 (PP2):**
> *Quais características(features) melhor predizem os níveis de presença cognitiva e de feedback educacional para mensagens de transcrição online originais e traduzidas?*

A pergunta de pesquisa 2 tem como finalidade compreender a importância das características obtidas pela codificação de cada idioma no processo de classificação, e verificar a variação dessa importância quando o texto é traduzido. A análise textual é um processo complexo,

na qual diferentes ferramentas de tradução apresentam resultados com o mesmo sentido, mas com palavras diferentes, levando assim a terceira pergunta de pesquisa:

> **PERGUNTA DE PESQUISA 3 (PP3):**
> *Há divergências significativas entre ferramentas de tradução automática para extração de recursos/características que tenham impacto relevante na classificação?*

A pergunta de pesquisa 3 tem como finalidade comparar o desempenho obtido por diferentes ferramentas de tradução no processo de classificação, e esclarecer se uma determinada ferramenta é mais apropriada que outra nesse processo.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral da presente pesquisa tem como finalidade demonstrar se o uso da tradução de textos para o idioma Inglês, mantém ou melhora o desempenho de classificadores automatizados de textos educacionais para identificação dos níveis de *feedback* educacional e presença cognitiva.

1.3.2 Objetivo Específico

Para atingir o objetivo geral foram definidos os seguintes objetivos específicos:

- Criar um algoritmo para tradução de textos do Inglês para o Português, e do Português para o Inglês.
- Gerar características de textos educacionais em diferentes idiomas através de ferramentas automatizadas de representação linguística.
- Desenvolver um modelo para classificação de níveis de *feedback* e presença cognitiva em textos educacionais.
- Avaliar o modelo desenvolvido com bases de treinamento e teste, devidamente anotadas.
- Ordenar as *features* por maior impacto no modelo de classificação.

1.4 Organização do Trabalho

Este trabalho está dividido nos seguintes capítulos: Capítulo 2 que apresentam as fundamentações teóricas do trabalho com os conceitos de COI, *feedback* educacional, mineração de texto, algoritmo de classificação e tradução de textos. No Capítulo 3 são citados os principais trabalhos relacionados a este estudo, referentes à classificação automática dos níveis de *feedback* e presença cognitiva, e as limitações dos trabalhos. No Capítulo 4 são detalhadas as etapas da metodologia desde a descrição e tradução do *corpus*, extração de características, até a construção e avaliação do modelo proposto. No Capítulo 5 são apresentados os resultados e discussões da pesquisa. Por fim, no capítulo 6 são feitas as considerações finais sobre a pesquisa, juntamente com os artigos submetidos, limitações da pesquisa e propostas de trabalhos futuros.

2 Embasamento Teórico

Neste capítulo, apresentam-se alguns dos conceitos básicos que embasaram o desenvolvimento da presente dissertação.

A Seção 1 elucida sobre as presenças do Modelo de Comunidade de Investigação (COI). A Seção 2.1.1 explica sobre os níveis de Feedback. A Seção 2.3 apresenta os principais métodos e ferramentas para Mineração de Texto. A próxima seção explicará os métodos estatísticos de Avaliação e interpretação de Resultados. A seção seguinte demonstra a importância e eficácia da Tradução de Textos. Por fim, a última seção do capítulo, demonstra as bibliotecas *Python* que auxiliaram no trabalho.

2.1 Modelo de Comunidade de Investigação

Dentre vários modelos pedagógicos conceituais de análise de aprendizagem em fóruns de discussão, um dos que se destaca é o modelo de COI, focado no processo social de construção conjunta e colaborativa do conhecimento em ambientes de comunicação assíncrona, baseado em textos (GARRISON et al., 1999). Esse modelo surgiu de uma pesquisa para examinar a qualidade no processo de aprendizagem e seus resultados, por meio do ensino baseado em CMC, em textos extraídos de ambientes virtuais de aprendizagem de nível superior. Para os autores, a educação virtual permite diferentes tipos de interação entre os principais componentes do processo educacional: professores e alunos, permitindo uma integração complementar a ponto de desenvolver capacidades cognitivas mais complexas, com ganhos de aprendizagem de maneira organizada, com metas constantes de exploração e participação.

Em uma definição mais simplista, o COI serve como modelo teórico adotado para orientar as diretrizes fundamentais da pesquisa e prática que facilitam o processo de aprendizagem *online*(TOLMIE; BOYLE, 2000). Essas diretrizes são estipuladas como as três dimensões imprescindíveis para que haja uma experiência educacional de sucesso (Figura 1): a presença social, a presença de ensino e a presença cognitiva.

Nas próximas subseções, serão detalhadas cada uma das dimensões e suas particularidades como fases e indicadores que caracterizam as presenças.

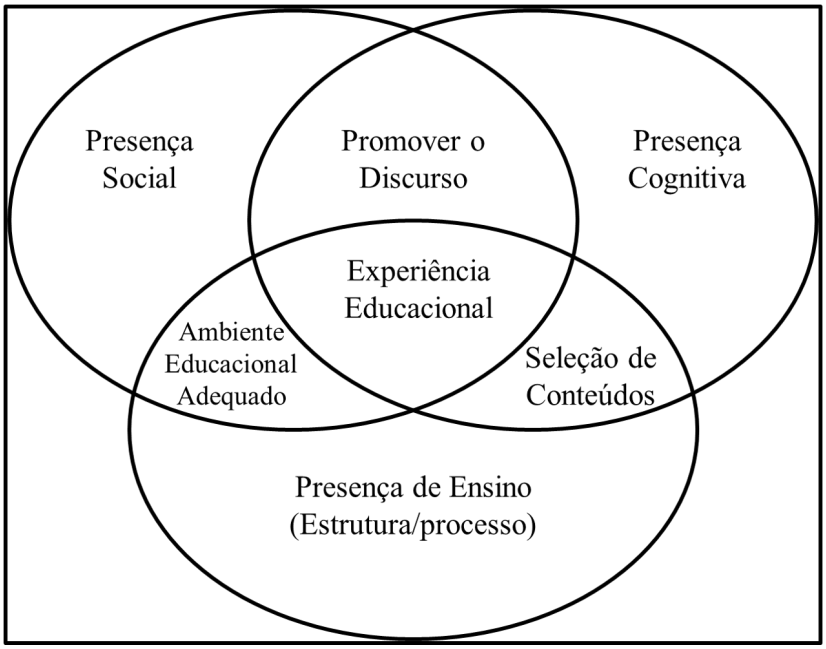

Figura 1 – Modelo de uma comunidade de investigação
(adaptado de GARRISON et al.(1999))

2.1.1 Presença Cognitiva

A presença cognitiva se baseia em uma abordagem de investigação prática, desenvolvendo pensamento crítico através de experiências, envolvendo uma relação interativa e recíproca entre o mundo pessoal e compartilhado (GARRISON et al., 1999). Assim, os estudantes por meio de uma análise, podem construir e confirmar o significado de determinados contextos através de discussões e reflexões (GARRISON et al., 2001; ARAUJO; NETO, 2013).

Adaptado das fases reflexivas da investigação prática, foram estabelecidos dois eixos para conceber o ciclo que compõem a experiência (GARRISON et al., 2001). O primeiro eixo é integrado pela concepção das idéias relacionadas à percepção da consciência. Já o segundo eixo é integrado pela ação prática em alusão à necessidade de uma aplicabilidade obtida pelo questionamento (deliberação). Os quatro quadrantes de intersecção entre os dois eixos refletem a sequência lógica da investigação prática correspondente aos indicadores de presença cognitiva, são eles (Figura 2):

- **Evento desencadeador (*Triggering Event*):** quadrante que se encontra entre o eixo da ação prática e o eixo da percepção da consciência. Em um ambiente propício, no qual vários integrantes compartilham informações e dúvidas, algum elemento, estando insatisfeito por ter a percepção de um problema ou questionamento e tomar a ação de buscar a solução ou melhoria do que foi questionado. Assim, o elemento que desencadeia a interação sobre

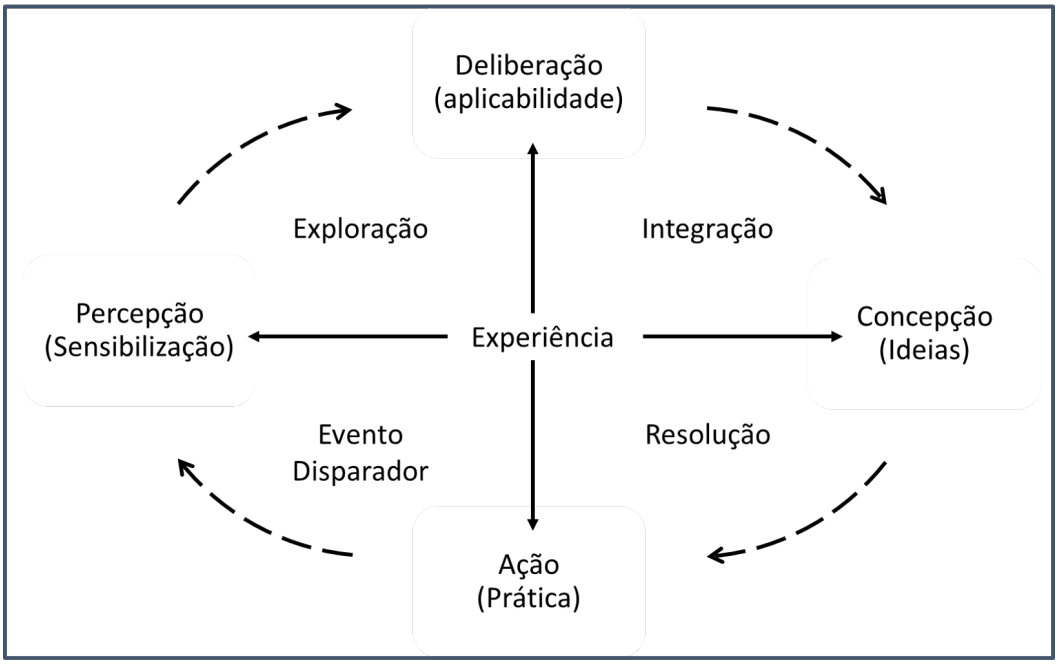

Figura 2 – Modelo Prático de Investigação
(adaptado de GARRISON; ARBAUGH(2007b))

um determinado assunto.

- **Exploração (*Exploration*):** quadrante que se encontra entre o eixo da percepção da consciência e o eixo da Deliberação(aplicabilidade). Uma fase da análise de experiências em particular que buscam esclarecimentos que possam ajudar a entender a situação ou o problema desencadeado para apontar algumas soluções válidas e aplicáveis à sua realidade.
- **Integração (*Integration*):** quadrante que se encontra entre o eixo da Deliberação(aplicabilidade) e o eixo da Concepção(Ideias). Com as informações obtidas pela fase da exploração, várias hipóteses e possíveis soluções são levantadas. Assim é necessário uma reflexão que busque uma solução eficiente por descartar ideias muito custosas, ou combinar ideias simples para a solução. Assim, visa integrar as ideias e conceitos mais relevantes e relacionados em uma hipótese mais coerente.
- **Resolução (*Resolution*):** quadrante que se encontra entre o eixo da Concepção (ideias), e o eixo da Ação. Nessa fase são discutidas as melhores ideias obtidas para entrar em consenso entre os integrantes, de qual será aplicada ou estabelecida como a resolução do que foi proposto pelo evento desencadeador.

Esses quadrantes da investigação prática na presença cognitiva, conceituam indicadores expressivos para modelo de análise e classificação confiável, permitindo uma codificação objetiva e consistente de mensagens, os quais são itens mais comuns em fóruns de discussão educacional em ambientes de ensino.

2.1.2 Presença Social

Para as características sociais da comunicação, Mehrabian (1968) resume algumas medidas significativas de atitudes de um comunicador e o diálogo com o seu destinatário.São como indicadores, pistas de comunicação que aumentam a proximidade e interação não-verbal, que aumentam a estimulação não verbal dos interlocutores: postura, posição, movimento, expressões faciais e verbais implícitas. Daft e Lengel (1986) concordam que a falta de informações em diálogos não verbais, pode comprometer a afetividade e confiança entre os participantes, mas argumentam que em algumas situações isso pode ser benéfico, como o envio de textos curtos, promovendo um entendimento claro e uma comunicação eficaz.

Walther (1994) apresentou casos onde os envolvidos declararam algumas formas de CMC como e-mail e conferência, são tão ou mais ricas que as conversas telefônicas e face-a-face. Gunawardena e Zittle (1997) realizaram um estudo que envolvia alunos em uma conferência por computador, e obteve uma avaliação positiva média, como sociável a relação entre os integrantes, e viram que aprimoraram suas experiências socioemocionais usando *emoticons* para expressar por escrito, indicadores não verbais.

No contexto da COI, a presença social refere-se a capacidade dos participantes se projetarem emocionalmente por meio da comunicação, gerando familiaridade, habilidades, motivação e compromisso organizacional por realizarem em conjunto atividades e assim passarem mais tempo usando a mídia. (GARRISON et al., 1999) . Para isso, é necessário tempo visando alcançar um nível de conforto, confiança e sentimento de pertencimento ao ambiente pelos integrantes.

Fabro e Garrison (1998) consideram a presença social crucial para uma comunidade crítica de alunos. Esse senso crítico é obtido da expressão social, em ambientes de ensino que proporcionem criar vínculos e estabelecer confiança e buscar apoio para superar as dificuldades de aprendizagem(CUTLER, 1995). O aluno tem mais facilidade de se expressar emocionalmente havendo uma comunicação aberta, baseada no desenvolver e estender as relações de troca de favores ou informações para criar uma consciência mútua de colaboração. Para isso desenvolveram um esquema de presença social dividido em três categorias (GARRISON et al., 1999; ROURKE et al., 1999), e na Tabela 1 são apontados os indicadores e definições para essas três características.

Categorias	**Indicadores**	**Definição**
Expressões de Afetividade	Expressar emoções	Expressões convencionais ou não convencionais de emoções, incluindo pontuação repetida, uso de maiúsculas, símbolos (*emoticons*).
	Uso de humor	Piadas, ironias, sarcasmo.
	Informações sobre si	Apresenta detalhes da vida fora da classe ou expressa alguma vulnerabilidade.
Comunicação Aberta	Continuar uma conversa	Usar o comando "responder" do software, em vez de começar uma conversa nova.
	Citar mensagens de outros	Usar os recursos do software para citar as mensagens inteiras dos outros, cortar e colar partes de outras mensagens.
	Fazer referência às mensagens de outros	Fazer referência ao conteúdo de outras mensagens.
	Fazer perguntas	Os alunos fazem perguntas a outros alunos ou ao professor.
	Saudar, expressar apreço, expressar concordância	Elogiar os outros ou os seus comentários; Expressar concordância com outros ou com o conteúdo de outras mensagens.
	Concordar	Concordar com outros ou com o conteúdo das suas mensagens.
Coesão de Grupo	Vocativos	Referir-se aos participantes pelo nome.
	Refere-se ao grupo usando pronomes inclusivos	Refere-se ao grupo por "nós", "nosso grupo".
	Saudações	Comunicação social: saudações, despedidas.

Tabela 1 – Dimensões, categorias e indicadores da Presença Social (adaptado de ROURKE et al. (2001))

- **Expressões de Afetividade (*Affective*):** elementos textuais que insinuam no texto características afetivas como Expressar emoções por forma de simbolos/*emoticons*, pontuações repetidas e uso de letras maiúsculas; Uso de humor como forma de ironia, sarcasmo e piadas; Informações pessoais de detalhes da vida fora da sala de aula.
- **Comunicação Aberta (*Interactive*):** O fato de um integrante da conversa responder à alguma dúvida; citar ou fazer referência a mensagem de outros para responder algo ou elogiar por ter se esforçado e tirado alguma dúvida; Fazer perguntas ao professor ou outros alunos; ou o simples fato de concordar com algo que foi postado.
- **Coesão de Grupo (*Cohesive*):** Ter uma noção que está interagindo com um grupo e a importância de todos eles, de maneira a chamar alguém pelo nome; não ser individualista e utilizar de pronomes inclusivos como "nós" ou "nosso grupo"; além de realizar uma saudação ao entrar na conversa ou se despedir ao sair dela

2.1.3 Presença de Ensino

A Presença de Ensino ou Pedagógica refere-se às facilidades que podem ser oferecidas aos processos cognitivos e sociais, a fim de obter resultados de aprendizagem significativos (ANDERSON et al., 2001b). Para Garrisson (2003), a atuação do professor na presença de ensino é fundamental na experiência educativa, principalmente no ensino a distância, onde terá que lidar com ferramentas e situações diferentes do ensino presencial. O papel do professor assim, não será apenas de lecionar a matéria que detém o conhecimento, mas também ter uma estratégia educacional à dessa outra forma de ensino, além de ser um facilitador social. Ele faz isso por gerenciar o ambiente e facilitar a aprendizagem dos alunos, lhes oferecendo um espaço mais acessível, didático, intuitivo, confortável e seguro, proporcionando as condições necessárias para o desenvolvimento de aprendizado. Em uma COI, a construção colaborativa do conhecimento por meio da presença de ensino, é dividida em 3 categorias: Desenho Instrucional e Organização; Facilitação do Discurso e Instrução Direta.

- **Desenho Institucional e Organizacional:** categoria relacionada ao planejamento referente à estrutura, processo, interação e avaliação do curso *online*. São procedimentos desempenhados dando suporte capacitando os instrutores quanto ao uso de ferramentas especializadas, transparência e clareza da estrutura. Isso pode ser feito por estabelecer uma ementa e programação do conteúdo, além das tecnologias necessárias para aplicar

os métodos desejados de ensino e acompanhamento dos alunos e atividades. Também é importante a instrução dos alunos para o uso devido das tecnologias e as formas e prazos corretos para realização de trabalhos e atividades (GARRISON; ARBAUGH, 2007a; ANDERSON et al., 2001a).

- **Facilitação do Discurso:** o curso deve oferecer recursos para que os alunos se sintam incentivados e comprometidos com a interação entre eles e na construção do conhecimento. Facilitar o discurso estabelece um vínculo de comunidade de aprendizagem que impulsiona a construção de significado, assim como a compreensão dos elementos envolvidos para a construção dela. Nessa categoria, o discente é fundamental para estimular debates, identificar áreas de acordo e desacordo, responder e incentivar as contribuições do grupo, tudo isso em um clima harmonioso e propício para a aprendizagem e construção do conhecimento (ANDERSON et al., 2001a; GARRISON; ARBAUGH, 2007a; PESSOA, 2012).
- **Aprendizagem direta:** Categoria onde os professores, com toda sua experiência, bagagem intelectual, pedagógica e acadêmica, compartilham seus conhecimentos com os alunos, por analisar comentários, apresentar informações e ferramentas relevantes, reflexão do conteúdo disponibilizado, inserir fontes de pesquisas científicas, orientar o sentido das discussões para instruir os alunos a adquirirem novos conhecimentos (ANDERSON et al., 2001a; GARRISON; ARBAUGH, 2007a).

2.2 Feedback Educacional

O objetivo principal do *Feedback* educacional é reduzir a diferença entre o desempenho do estágio atual para um nível de entendimento desejado (HATTIE; TIMPERLEY, 2007). Assim, o *feedback* se demonstra parte essencial no processo de aprendizagem, garantindo boas estratégias para alunos e professores praticarem, de acordo com as necessidades percebidas ao longo do processo (BUTLER; WINNE, 1995; SADLER, 1989; KLEIJ et al., 2015).

Para os instrutores, cabe fornecer objetivos específicos e desafiadores, em um nível apropriado às capacidades e necessidades percebidas do aluno, e assim melhorar o seu desempenho (HATTIE; TIMPERLEY, 2007; LANGER, 2011; ORRELL, 2006). Objetivos específicos são mais eficazes do que os gerais ou não específicos, pois direcionam a atenção do aluno a cumprir metas mais simples e alcançáveis (LOCKE; LATHAM, 1984), e garantem um

ambiente de aprendizado propício para desenvolverem habilidades de autorregulação e detecção de erros (HATTIE et al., 1996).

Para os alunos, receber o *feedback* pode ajudar a detectar áreas necessárias para promover a melhoria em seus conhecimentos ou habilidades, e empregar estratégias mais eficazes de aprendizagem (PARIKH et al., 2001). O aluno aumenta o esforço quando o objetivo é claro, levando a um alto grau de comprometimento por crer que a possibilidade de sucesso no final é alta (KLUGER; DENISI, 1996).

Pesquisas anteriores já apontam o uso de práticas de *feedback* para trazer benefícios na aprendizagem (NICOL; MACFARLANE-DICK, 2006; PARIKH et al., 2001), sendo parte central no desenvolvimento do aluno (HATTIE; TIMPERLEY, 2007). Mesmo assim, ainda se constata que a ocorrência de práticas subdesenvolvidas de *feedback*, limitando o progresso do aluno, pois podem criar desconfiança no processo de feedback e no professor, e resultar em problemas na motivação (CAVALCANTI et al., 2020b; MATCHA et al., 2019). O estudo de BURKE (2009), revela a insatisfação dos alunos com *feedbacks* muito curtos, negativos ou de difícil compreensão.

Nível	**Desrição**	**Exemplo**
FT	Esta categoria fala sobre as atividades de baixo nível. Por exemplo, se o aluno acertou ou errou uma atividade	"Você precisa incluir mais sobre o Tratado de Versalhes."
FP	O feedback pode ser direcionado ao processamento de informações ou processos de aprendizado que exigem a compreensão ou a conclusão da tarefa.	"Você precisa editar esta parte do texto, atendendo aos descritores que você usou, para que o leitor possa entender as nuances do seu significado,"
FR	Texto que auxilia o estudante no processo de autorregulação, como por exemplo uma autoavaliação ou melhora de confiança do aluno	"Você já conhece os principais recursos da abertura de um argumento. Verifique se você os incorporou em seu primeiro parágrafo."
FS	Feedback mais pessoal para o aluno, em geral não relacionado a atividade.	"Parabéns, você se esforçou bastante nessa atividade!"; "Você é um ótimo aluno!"; "Essa é uma resposta inteligente e muito bem elaborada!".

Tabela 2 – Níveis de Feedback
(adaptado de HATTIE; TIMPERLEY(2007))

Seguindo o modelo de Hattie e Timperley (2007), o *feedback* eficaz deve responder à três perguntas, que correspondem respectivamente às noções de *feed-up, feed-back e feed-forward*:

- **Para onde estou indo?** Deixar claro os objetivos específicos a serem alcançados.
- **Como eu estou indo?** Mensurar o progresso que está sendo feito em direção ao objetivo.
- **Onde ir a seguir?** Ter a noção das atividades que ainda precisam ser realizadas para progredir melhor, seja por meio de instrução do professor ou reflexão do aluno.

Esses três questionamentos operam em 4 níveis: Feedback sobre a Tarefa (FT), Feedback

sobre o Processamento da Tarefa (FP), Feedback sobre Autorregulação (FR) e Feedback sobre a Pessoa (FS). Na Tabela 2, foram demonstrados breves descrições e exemplos sobre cada nível, que serão melhor definidos nas seções posteriores.

2.2.1 Feedback sobre a Atividade

O *feedback* sobre a atividade ou tarefa (do termo em inglês *Feedback About the Task*), também chamado de *feedback* corretivo, permite criar estratégias para processar e compreender o conteúdo, permitindo ao aluno identificar respostas incorretas advindas de uma interpretação errada ou pela falta de uma informação significativa (HATTIE; TIMPERLEY, 2007).

Quando o FT é muito específico para a pergunta em questão, não direcionado para o nível do processo necessário para ter a confiança de responder outras perguntas, pode direcionar à uma atenção superficial, que não atingirá um desempenho de alto nível pelos alunos. (THOMPSON, 1998; KLUGER; DENISI, 1996), que levam à estratégias de tentativa e erro de menos esforço cognitivo. Assim, o FT se torna mais benéfico quando ajuda o aluno a rejeitar hipóteses erradas e fornece pistas para estratégias de pesquisas que os informem e capacitem para determinada tarefa ou situação. Isso seria a construção de hipóteses sobre a relação entre instruções, *feedback* e aprendizado pretendido.

2.2.2 Feedback sobre o Processo

O *feedback* sobre o processo da atividade ou tarefa (do termo em inglês *Feedback About the Processing of the Task*), apresenta o conteúdo complementar ao FT superficial, promovendo instruções mais específicas para processos subjacentes às outras tarefas ou informações sobre as relações do ambiente de aprendizado e as percepções do aluno (BALZER et al., 1989).

No processo de aprendizagem, uma melhor compreensão de um assunto está associada ao maior uso de estratégias diferentes por parte dos estudantes (PURDIE et al., 1996), e o FP fornece pistas que atuam como um mecanismo de indicação para busca de informações de modo mais eficaz ou rejeitar hipóteses errôneas, demonstrando ser mais eficaz no aprendizado profundo do que o FT, mas que a interação entre eles resulta em um poderoso efeito na aprendizagem (HATTIE; TIMPERLEY, 2007).

2.2.3 Feedback para Autorregulação

O *feedback* para autorregulação (do termo em inglês *Feedback About Self-Regulation*), auxilia os alunos a se tornarem ativos em seu próprio desempenho de aprendizagem, ao invés de serem passivos destinatários de instruções. Assim eles são orientados a definir metas razoáveis, persistir diante do fracasso e adotar boas estratégias para correção de erros. Isso envolve decisões sobre quais tarefas realizar baseadas no nível que conseguem solucioná-las, quando buscar ajuda e como superar obstáculos. O sentimento de capacidade os motiva a investirem esforços em tarefas desafiadoras e relevantes (PARIS; WINOGRAD, 1990).

Uma pesquisa revela fatores importantes quanto à atenção dada ao feedback em relação à expectativa de qual seja a resposta e o tipo de resultado. O feedback tem pouca atenção quando a confiança da resposta é alta e a resposta é certa, ou quando a confiança da resposta é baixa e a resposta está errada. A situação de confiança baixa e com apenas a resposta certa estimula o processo de tentativa e erro. Mas o feedback tem maior efeito quando o aluno está confiante que a resposta esteja certa, mas está errada (KULHAVY; STOCK, 1989), estimulando a adequar sua estratégia e buscar ajuda. Quando se busca uma ajuda executiva (resposta ou ajuda direta que evite tempo e trabalho) se relaciona com FT ou FP. Já quando solicita uma ajuda instrumental (sugestões ao invés de respostas), levam ao nível de FR (HATTIE; TIMPERLEY, 2007).

2.2.4 Feedback sobre a Pessoa

O *feedback* sobre a pessoa (do termo em inglês *Feedback About the Self as a Person*), ou *feedback* Pessoal, é bastante utilizado em situações de classe de aula, normalmente expressas de maneira positiva (como “Boa garota” ou “Grande esforço”), mas as vezes de maneira negativa pelas críticas. (Brophy, 1981). É um *feedback* que se demonstra imprevisível para mensurar seu impacto como positivo ou negativo, pois depende do perfil de cada aluno. Algumas pesquisas relatam reações diferentes à forma como o *feedback* é feito, e de acordo com a experiência dos alunos, que em alguns casos gostam de serem elogiados publicamente ou em particular, já outros não gostam de elogios (SHARP, 1985). Outros interpretam a forma como o *feedback* foi apresentado com sentimento que o avaliador tenha subestimado ou superestimado suas habilidades (MEINOLF et al., 1979).

Geralmente o FS apresenta-se com pouca informação sobre a tarefa e está focado no elogio pessoal. Mas em alguns casos consegue ser convertido como sinal de sucesso,

parabenizando a forma como conseguiu o êxito no seu desempenho, seja por esforço, autorregulação, engajamento ou processo. (por exemplo, ”Você é realmente ótimo porque você completou esta tarefa diligentemente aplicando esse conceito”). (HATTIE; TIMPERLEY, 2007).

2.3 Mineração de Texto

A Mineração de Texto (MT) tem o objetivo de extrair de informações úteis, por meio da identificação e exploração de padrões relevantes em bases textuais, não estruturadas ou semi-estruturadas (FELDMAN; SANGER, 2007). Segundo CHOWDHURY (2003), áreas como Aprendizagem de Máquina (AM) e Processamento de Linguagem Natural (PLN) contribuem para este processo de extração de conhecimento. A AM estuda os meios que capacitem aos computadores adquirir conhecimento e, a partir de um conjunto de dados de treinamento, estimar saídas para dados desconhecidos (NILSSON, 1996). Já PLN estuda o desenvolvimento de programas que fornecem aos computadores a capacidade realizar tarefas como: reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos, analisar sentimentos e até aprender conceitos com os textos processados (VIEIRA; LOPES, 2010).

Para Aranha e Passos (2006), a extração do conhecimento por meio da MT, se dá pela realização das seguintes etapas: coleta de dados, pré-processamento, extração do conhecimento e avaliação e interpretação dos resultados. Essas etapas serão melhor explicadas nas subseções seguintes.

2.3.1 Definição do *Corpus* e Pré-Processamento

A Definição do *Corpus* é a primeira etapa para se construir a base de dados textual que serve como fonte para manipulação e processos das etapas seguintes (ARANHA; PASSOS, 2006). Na literatura, essa base é comumente denominada de *corpus*, e é construída com uma grande quantidade de documentos previamente coletados de uma ou mais fontes. Nos AVAs, esses documentos são provenientes de campos textuais em seus bancos de dados dos processos educacionais, além de comunicações feitas por outras fontes como redes sociais, *emails*, *chats* e fóruns de ambientes *online*.

Já na fase de pré-processamento, são aplicadas técnicas, separadas ou combinadas, que preparam o *corpus* para o formato adequado que será processado pelas etapas seguintes. Dentre as principais técnicas temos: a tokenização, que identifica e separa cada unidade existente em um texto em *tokens* (pode ser uma palavra, número ou sinal) (HABERT et al., 1998); remoção de *stopwords*, que consiste em remover termos com pouca significância em um texto, como: artigos, preposições e conjunções (LO et al., 2005) ; e *stemming*,para reduzir as palavras de um texto para sua forma gramatical raiz ou inicial,(ex. "perdeu" e "perdemos" possui o radical "perd") (RAMASUBRAMANIAN; RAMYA, 2013).

2.3.2 Extração de Características

A finalidade da extração de características é criar novos dados, a partir de informações já existentes, que representem de maneira mais adequada os padrões do documento. Quanto mais relevantes forem as características do documento para o processo de classificação, melhor será a confiabilidade do classificador. (SHAH; PATEL, 2016).

Existem diversas técnicas para extração de características, como por exemplo $Bag of Words$ - onde todas as palavras do documento são características em um espaço multidimensional de vários documentos, e cada dimensão é um termo da coleção (FELDMAN; SANGER, 2007). Nessa abordagem, o vetor pode ser dimensionado em um determinado documento como menos importante para termos que aparecem com muita frequência, e mais importantes para termos que ocorrem raramente nos documentos, caracterizando a teoria da modelagem de linguagem *Term Frequency - Inverse Document Frequency* (TF-IDF) (ROBERTSON, 2004).

Também temos a técnica *Part-of-Speech Tagger* - onde é atribuída uma classe gramatical a cada palavra ou termo contido em uma sentença (VIEIRA; LIMA, 2001); e $Dependencyparsing$ - que rotula as dependências entre palavras com relações gramaticais (MARNEFFE et al., 2006), e ferramentas de categorizam as palavras e sentenças de acordo com dicionários léxico, sintático ou semântico, como *Linguistic Inquiry and Word Count* (LIWC), (PENNEBAKER et al., 2001) e *Cohesion Metrics* (Coh-Metrix) (GRAESSER et al., 2004).

2.3.3 Classificação de Texto

Na etapa da extração do conhecimento, aplicam-se algoritmos que procuram informações desconhecidas que possam ser úteis para o contexto de um problema específico. O algoritmo para classificação ou categorização textual, determina uma ou mais classes para um documento específico (SEBASTIANI, 2002).

Uma das formas para realizar essa categorização de maneira automática, é através de algoritmos de classificação baseados em aprendizado de máquina, que escolhe ou combina parâmetros de representação de modelo para classificar corretamente os exemplos conhecidos, e também novos exemplos (LEE, 2000; ROSSI, 2015). Os exemplos são representados por características significantes para determinado contexto, e também são referenciadas como *features*, atributos, variáveis (WEISS; KULIKOWSKI, 1991).

Assim, na classificação textual, um documento é interpretado como um conjunto de atributos que fornecem informações essenciais para o aprendizado do classificador descrever e categorizar um determinado documento.

2.3.3.1 Algoritmos de Classificação

No meio da computação, diversos são os algoritmos de classificação, seja para números, textos, imagens ou outros tipos de dados. Cada algoritmo se mostra importante sob um determinado contexto, combinação de variáveis e métricas. Estudos feitos diretamente com classificação de texto, como o de (RAMRAJ et al., 2018), apontam como adequados e demonstra o comparativo para a classificação de dados em formato de texto, os algoritmos: *Extreme Gradient Boosting* (XGBOOST), *Suppor Vector Machine* (SVM), *Neural Network* e *Random Forest* (RF).

Um SVM é baseado em diversos estudos teóricos sobre aprendizado estático (VAPNIK, 1995; VAPNIK, 1998a; VAPNIK, 1998b), que estabelecem principios na obtenção de classificadores com boa generalização e capacidade de prever corretamente classes de novos dados no mesmo domínio em que o aprendizado ocorreu.

LIN; WANG (2002) explica que o SVM mapeia os pontos de entrada, denominados support vectors, para definir o hiperplano ideal de separação que maximiza a margem entre duas classes nesse espaço. Adotando um conjunto de métodos de aprendizado supervisionado para analisar dados e reconhecer padrões, o SVM é usado para classificação e análise de regressão

estabelecendo uma variação no formato das margens para separação das classes (JOACHIMS, 1998).

Um Classificador de Rede Neural é um conjunto interligado neurônios que absorvem das unidades de entradas um valor real, multiplicam-no por um peso e o executam através de uma função de ativação não linear dessas unidades, e o valor que as unidades de saída determina a decisão de categorização. Ao construir várias camadas de neurônios, cada uma recebe parte das variáveis de entrada das camadas anteriores e repassa seus resultados para as próximas camadas, possibilitando a aprendizagem de funções muito complexas (KORDE; MAHENDER, 2012; HANSEN; SALAMON, 1990).

Outro método de classificação que se destaca no aprendizado de máquinas com abordagem orientada à dados, é o Aumento de Árvores Gradiente(FRIEDMAN, 2001), um método eficaz e amplamente utilizado por cientistas de dados(CHEN; GUESTRIN, 2016). A modelagem preditiva utilizando árvores de decisão, consiste em uma série de ramificações gerando vários subconjuntos separados por uma determinada regra estabelecida em vários locais de divisão (nós).

O XGBOOST, descrito por MITCHELL; FRANK (2017), impulsiona essas árvores de decisão com a construção de vários modelos sequencialmente, com cada novo modelo tentando corrigir as deficiências do modelo anterior.

O algoritmo *Random Forest*, abordado por BREIMAN (2001) e ilustrado na Figura 3 é um classificador constituído por um conjunto de preditores estruturados em árvore, onde vetores são aleatórios e distribuídos de forma idêntica e independentes, onde cada árvore emite um voto unitário para eleger classe mais popular.

Uma RF aplica o método *bagging* (*Bootstrap Aggregating*) que indica para cada árvore gerada amostras aleatórias (*bootstraps*) a partir do conjunto de treinamento inicial, no qual são escolhidos aleatoriamente exemplos para um novo subconjunto de treinamento (EFRON, 1992).

Dos benefícios para o uso de *bagging* destacam-se: a melhora do desempenho do algoritmo e a possibilidade de ter estimativas internas do erro de generalização do conjunto combinado de árvores, da força de uma árvore classificadora e de alguma combinação entre elas utilizando o método *Out-of-Bag* (OOB).

O erro OOB é o erro médio de predição onde cada amostra de treinamento X_i , utiliza apenas as árvores que não tinham X_i em sua amostra de *bootstrap*. Através desse método estima-se a relevância das variáveis, a correlação entre os classificadores e a força de predição de cada

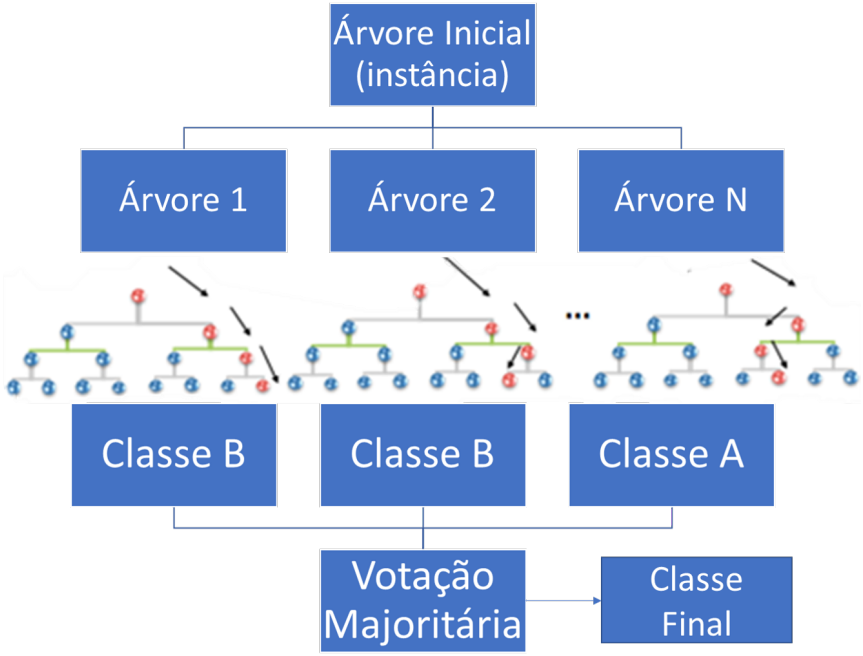

Figura 3 – Demonstração da metodologia de uma *Random Forest*
(adaptado de FU(2017))

variável. (JAMES et al., 2013; FRIEDMAN et al., 2001).

Segundo BREIMAN (2001), são propostas duas medidas de importância de atributos, uma baseada no decaimento da exatidão média, do inglês *MeanDecrease in Accuracy* (MDA), e a outra na diminuição média do índice de impureza de Gini, do inglês *Mean Decrease in Gini* (MDG). No processo de treinamento da RF, cada nó da árvore busca uma medida ótima para a divisão potencial capaz de separar os tipos de amostra pelo nó em questão. A medida é calculada pela soma de todas as diminuições na impureza de Gini a cada divisão do nó, normalizada pelo número de árvores.

Quando a RF tem uma base grande de variáveis para determinar o tipo de classe a qual ela pode pertencer, é possível estimar valores à dois parâmetros que limitam e ajustam o tamanho da árvore (KIM et al., 2006):

- ***Ntree*** **:** número de árvores usadas pela floresta aleatória.
- ***Mtry*** **:** número de número de variáveis de entrada tentadas no subconjunto aleatório em cada nó.

Explorando variações desses dois parâmetros, pode-se conseguir as combinações mais adequadas para alcançar um melhor valor para a precisão da classificação com um menor custo computacional.

2.3.3.2 Avaliação e interpretação dos resultados

Para avaliar os classificadores é preciso estabelecer métricas para identificar níveis de precisão de como os métodos envolvidos classificam corretamente. Em avaliações estatísticas, é comum verificar a confiabilidade com a Acurácia. De acordo com a ISO 572-4 da *International Organization for Standardization* (1994) [1] a acurácia é utilizada para descrever a proximidade da média de um conjunto de resultados de medição com o valor real. Já com uma perspectiva mais estatística, a acurácia trás o índice dos valores preditos condientes aos valores reais em relação a todos os resultados obtidos, a Figura 4 demonstra a formação da fórmula da acurácia pela matriz de confusão entre os valores preditos e valores reais.

		Valor Predito		Fórmula da Acurácia
		Sim	Não	
Valor Real	Sim	Verdadeiro Positivo TP	Falso Negativo FN	$Acurácia = \frac{TP + TN}{TP + TN + FP + FN}$
	Não	Falso Positivo FP	Verdadeiro Negativo TN	

Figura 4 – Matriz de Confusão e Fórmula da Acurácia.

Para tratar dados desbalanceados, usam-se medidas mais informativas que a acurácia. O coeficiente $kappa(\kappa)$ por exemplo, é um índice estatístico usado para medir a concordância entre avaliadores para itens qualitativos/categóricos (COHEN, 1960). A escala de concordância varia de acordo com os valores encontrados pelo índice, e quanto mais próximo de 1, indica um grau de concordância quase perfeita, já para valor menor ou igual a 0 indica que não teve acordo ou que foi prevista pelo acaso (LANDIS; KOCH, 1977). Quando o processo de avaliação se dá por juízes humanos geralmente visam atingir uma pontuação *kappa* acima de 0,70 (GOLDEN, 2011; WIEBE et al., 2005).

Para evitar classificações tendenciosas, são utilizadas formas de embaralhamento dos dados de treino e teste e dentre elas se destaca a técnica estatística *Cross Validation* (CV). Esse tipo de método é usado quando a distribuição está desigual, e o modelo ajustado em algumas seções da base de dados terá um resultado melhor que em outras seções. Utilizando esse método, os dados são embaralhados para evitar erros tendenciosos de como a base está originalmente organizada. Depois os dados são repartidos em $\kappa - folds$ partes iguais onde cada uma dessas partes servirá como etapa de teste do classificador, enquanto as demais partes são mantidas para

[1] https://www.iso.org/standard/11834.html

treino (KOHAVI, 1995; REZENDE, 2003). Uma maneira de complementar a técnica CV, é a estratificação (DIAMANTIDIS et al., 2000), onde a distribuição de classe em cada $fold$ é aproximadamente a mesma para os conjuntos de treinamento e teste. Esse modelo está ilustrado na Figura 5 Dessa forma se garante que todos os dados sejam utilizados, sendo uma vez para teste e $k-1$ vezes para treinamento. A média dos resultados de cada avaliação é obtida em conjunto para uma pontuação final que definirá o modelo gerado que influi menos no resultado final (PENG et al., 2004).

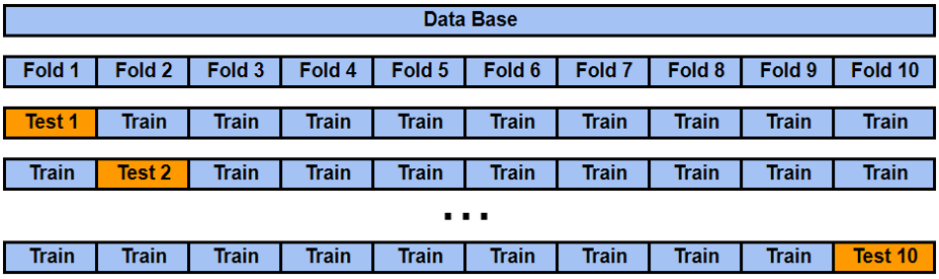

Figura 5 – Demonstração do método *Cross Validation* com $\kappa - fold = 10$ (inspirado em (REZENDE, 2003))

2.4 Tradução de Textos

Com uma maior facilidade de manipulação de textos em formato digital, algumas pesquisas envolvendo a tradução de textos se iniciaram na década de 60 (WHITE; O'CONNELL, 1993), mas foi na década de 90 que vários estudos tomaram a iniciativa para análise de qualidade das máquinas de tradução (*Machine Translations*) automática impulsionaram a exploração diferentes metodologias para a complexidade do tema, pois o julgamento é altamente subjetivo, mesmo para humanos especialistas em tradução (WHITE et al., 1994).

Segundo HUTCHINS (1998), a ferramentas desenvolvidas são resultantes de quatro tipos de demanda de tradução. O primeiro tipo corresponde ao conteúdo de divulgação, que precisa de um revisor humano para atingir o nível de qualidade requerido. O segundo tipo corresponde a assimilação parcial do conteúdo do texto, em que o usuário prefere algum entendimento, mesmo que com baixa qualidade, do que nada. O terceiro tipo envolve o intercâmbio comunicativo de acesso à conteúdos em páginas *Web*, *e-mail* e *chats*. O quarto tipo de demanda é referente ao acesso à informações de sistemas através do conteúdo dos bancos de dados.

A evolução das máquinas de tradução deram a capacidade de processamento de grandes

quantidades de textos, que seriam inviáveis serem realizadas por humanos (HUTCHINS, 1998). Com isso, essa abordagem foi adotada em diversos tipos de pesquisa, como classificação de texto (SHI et al., 2010), resumo de texto (CABRAL et al., 2014) e extração de informação (HKIRI et al., 2017).

Na Seção 4.3.1, serão detalhadas as ferramentas de tradução utilizadas na presente pesquisa, a do Google Translate e a da Microsoft Azure. Na Tabela 3 pode-se observar que ocorre traduções diferentes para o mesmo texto, o que poderá resultar em diferentes tipos de características extraídas das mensagens para o processo de classificação.

Idioma	**Texto**
Inglês	TITLE: Towards efficient web engineering approaches through flexible Process Models
Tradução com Google Translate	TÍTULO: Rumo a abordagens eficientes de engenharia da web por meio de modelos de processo flexíveis
Tradução com Microsoft Azure	TÍTULO: Para abordagens eficientes de engenharia web através de modelos de processo flexíveis

Tabela 3 – Exemplo de traduções por diferentes ferramentas.

3 Trabalhos Relacionados

Neste Capítulo são apresentados trabalhos relacionados que utilizaram presença cognitiva para desenvolver o ciclo completo da experiência de aprendizagem que é formado pelo evento desencadeador, exploração, integração e resolução. Também, veremos os trabalhos que envolvem a avaliação do feedback e o processo de classificação automatizada.

3.1 Análise de presença cognitiva

Como elucidado na Subseção 2.1, os fóruns de discussão online não proporcionam a exploração dos três tipos de presença, essenciais para um ciclo completo de aprendizagem, pois a presença de ensino envolve muita interferência do tutor, não se dando de maneira espontânea, e há muitas limitações para expressões de sentimentos, que é algo fundamental na presença social em ambientes de ensino virtuais. Assim, o presente trabalho foca no terceiro tipo de presença, a presença cognitiva, que se baseia em uma abordagem de investigação prática, desenvolvendo pensamento crítico através de experiências, envolvendo uma relação interativa e recíproca entre o mundo pessoal e compartilhado (GARRISON et al., 1999).

Na prática, por ser uma abordagem espontânea e não obrigatória, a interação através dos fóruns, várias vezes, não chega a etapa final referente ao indicador de presença cognitiva, a resolução do problema. Muitas vezes, a discussão fica paralisada na etapa de exploração, talvez pela falta de vínculo entre os alunos (CUTLER, 1995) pelas dificuldades de expressar emoções num ambiente virtual. Assim, o aluno fica receoso em participar do debate por não ter confiança da aceitação da sua resposta (GARRISON et al., 2001). Além de encontrar nas mensagens, partes que não são pertinentes à presença cognitiva, como saudações ou agradecimentos (FARROW et al., 2019).

Mesmo descrevendo as dimensões importantes da aprendizagem em comunidades onlines, o esquema quantitativo de codificação para avaliação das mensagens, quando categorizado manualmente (WATERS et al., 2015), demonstra ser uma forma muito custosa e demorada. Um exemplo foi o estudo de (GAšEVIć et al., 2015) que contou com 2 codificadores experimentais para analisar 1747 mensagens e cada um levou em torno de 130 horas para concluir a tarefa.

Esse bom índice vem a um custo de tempo muito grande e dependência de codificadores experientes, limitando pesquisas manuais a pequenos projetos, impossibilitando uma análise

em alto escala. Com tanto tempo utilizado nesta abordagem, torna-se necessário uma detecção automatizada da presença cognitiva para dar mais agilidade para que o professor ou a plataforma de ensino identifique necessidades para que uma discussão complete o ciclo chegando no indicador de Resolução.

3.1.1 Análise automática de conteúdo de mensagens de discussão

Havendo a necessidade da automação de análise de conteúdo, estudos iniciais basearam suas abordagens prioritariamente em análises estatísticas sobre frases e palavras (MCKLIN, 2004). Já CORICH et al.(2006) projetou a Ferramenta de Análise de Conteúdo Automatizada (ACAT) para classificar mensagens de discussão assíncrona de acordo com diferentes categorias de mensuração de pensamento crítico, experimentando desde sentenças sem pré-processamentos à sentenças submetidas a remoção de palavras irrelevantes e à técnica stemming.

Outro estudo elaborado por (KOVANOVIC et al., 2014a) propuseram uma classificação baseado em técnicas de mineração de texto com extração de características explorando as variações de modelos de linguagem como: N-gramas, Dependency Triplets, Named Entities e Thread Position. Adotando o classificador SVM com *Cross Validation* 10*fold*. Posteriormente no trabalho de 2016, (KOVANOVIĆ et al., 2016) propôs extrair recursos no Coh-Metrix, LIWC, Análise Semântica Latente (LSA), Entidades Nomeadas e contexto de de discussão para produzir índices das representações linguísticas e discursivas de um texto, categorizados por um classificador Random Forest. Contando com a mesma base do último estudo de KOVANOVIĆ et al.(2016), FARROW et al. (2019), explorou variações para divisão dos dados de treinamento e teste com a técnica *Cross Validation* estratificada, resultado em uma acurácia de 0.606 e 0.43 para $kappa$.

Até então foram apresentadas algumas formas de classificação de textos em Inglês. Como o foco do artigo é apresentar o diferencial de classificação em diferentes idiomas, vale ressaltar alguns estudos que realizaram a pesquisa com textos em Português. Um estudo que vale a pena ser mencionado é o de (NETO et al., 2018), que desenvolveu um método para automatizar a análise de presença cognitiva, com técnicas de mineração de texto utilizando 87 recursos extraídos de diferentes fontes (Coh-Metrix, LIWC, Incorporação de palavras, Entidades Nomeadas e Contexto de discussão), utilizando um classificador Random Forest com Cross-Validation 10-fold, chegando a constatar que o número de palavras e comprimento médio das

frases sendo de relevância para identificar altos níveis de presença cognitiva.

A pesquisa de BARBOSA et al. (2020) trabalhou com a classificação automatizada de presença cognitiva para os idiomas Português e Inglês, trabalhando com 108 indicadores psicológicos utilizando o classificador *Random Forest* e obteve os resultados para acurácia e *kappa* 0,67 e 0,32 respectivamente.

Também destaca-se a pesquisa de BARBOSA et al. (2021) que classifica os níveis de presença cognitiva e sociais de textos em Inglês e Português e trabalhando com a tradução dessas bases, onde são extraídas 185 características para o idioma Português e 187 para o idioma Inglês, e obtiveram resultados de acurácia que variam de 0,57 à 0.83, e resultado para *kappa* que variaram de 0,38 à 0,69.

Na Tabela 4 são apresentados os resultados obtidos nas pesquisas citadas. Podemos notar que a quantidade de características não é tão relevante para obter um nível de confiabilidade elevado, e sim a importância e a expressividade das características em relação às estruturas das categorias catalogadas, junto com boas práticas de mineração de texto e classificação. Assim podemos focar nessas boas práticas para realizar o comparativo entre classificadores de diferentes idiomas.

Autor	**Idiomas**	**Mensagens**	**Caract.**	**Acurácia**	**Kappa**
(MCKLIN, 2004)	EN	1997	182	0,69	0,31
(CORICH et al., 2006)	EN	74 (484 frases)	-	0,64; 0,65 e 0,71	-
(KOVANOVIC et al., 2014a)	EN	1747	13 types	0,584	0,41
(KOVANOVIĆ et al., 2016)	EN	-	205	0,703	0,63
(FARROW et al., 2019)	EN	1747	205	0,606	0,43
(NETO et al., 2018)	PT	1500	87	0,83	0,72
(BARBOSA et al., 2020)	PT/EN	1500;1747	108	0,67	0,32
(BARBOSA et al., 2021)	PT; EN	1500; 1747	185; 187	0,57; 0,73; 0,58; 0,83;	0,38; 0,52; 0,38; 0,69

Tabela 4 – Desempenho de classificação automatizada de presença cognitiva

3.2 Análise de Feedback

Há grandes implicações da revisão de feedback para avaliação em sala de aula. A avaliação pode ser definida como sendo atividades que fornecem aos professores ou alunos informações de feedback relacionadas às três perguntas de feedback, das quais se pode interpretar o quão distante é a situação atual das metas de aprendizagem (HATTIE; TIMPERLEY, 2007).

Diferente do que habitualmente nos deparamos, onde uma atividade é usada para avaliar os níveis de proficiência dos alunos, colocando mais ênfase na adequação das notas e menos na interpretação das respostas. Black e Wiliam BLACK; WILIAM (1998), concluíram que a avaliação normalmente estimula a aprendizagem superficial e mecânica, não revisam as perguntas de avaliação, e sem discutir criticamente com os colegas, há pouca reflexão sobre o que está sendo avaliado. Assim os alunos recebem pouca informação de feedback nesses casos.

A pesquisa de BANGERT-DROWNS et al. (1991) concluiram que os alunos que fizeram alguma avaliação, pontuaram melhor em exames de critério, e uma maior frequência das avaliações, sem exageros, foi associada a um melhor desempenho. As avaliações se mostram necessárias, numa quantidade moderada a ponto que permite a revisão, discussão e reflexão sobre o assunto. Por meio de avaliações, os alunos obtêm informações sobre como e o que eles entendem e entendem mal, encontrar direções e estratégias que devem seguir para melhorar e buscar assistência para suas necessidades. Já para os professores, é a oportunidade de planejar atividades e perguntas que forneçam feedback sobre a eficácia do ensino, e que saibam o que fazer a seguir.

Complementando os estudos de *feedback*, NICOL; MACFARLANE-DICK (2006) proporam 7 boas práticas do *feedback* para aumentar a capacidade do aluno se autorregular e melhorar o desempenho. Alguns trabalhos tomaram a direção da Análise de Rede Epistêmica, do inglês *Epistemic Network Analysis* (ENA), uma técnica para compreender visualmente e qualitativamente a relação entre dados, e tem sido adotadas tanto de maneira geral do campo educacional quanto sob presença cognitiva se social (SHAH; PATEL, 2016; CAVALCANTI et al., 2020a).

Autor	**Mensagens**	**Ferramenta**	**Idioma**	**Características**	**Acurácia**	**Kappa**
(CAVALCANTI et al., 2020b)	1000	RF (Random Forest)	PT	116	0.75	0.29
					0.64	0.28
					0.87	0.39
(CAVALCANTI et al., 2019)	1000	RF	PT	116	0.75	0.20
(CAVALCANTI et al., 2020a)	1000	XGBOOST	PT	116	0.91	0.82

Tabela 5 – Desempenho da classificação automatizada de textos para níveis de *feedeback*.

Por fim, estudo de CAVALCANTI et al. (2019) se aprofundou nas pesquisas envolvendo a classificação dos níveis de feedback. utilizando uma base no idioma Português com 1000 mensagens e 116 características. Como podemos ver na Tabela 5, inicialmente, o autor utilizou o algoritmo *Random Forest*, e posteriormente, aperfeiçoou a classificação mudando certos parâmetros, acrescentando boas práticas de *feedback* ou mudando o algoritmo de classificação para conseguir resultados melhores (CAVALCANTI et al., 2020a; CAVALCANTI et al., 2020b).

3.3 Limitações dos trabalhos anteriores

Como demonstrado nas seções anteriores, nenhum dos estudos chegam a trabalhar com a classificação de *feedback* juntamente com presença cognitiva.

Até 2014, havia uma limitação quanto ao número de características utilizadas na classificação de presença cognitiva(MCKLIN, 2004; CORICH et al., 2006; KOVANOVIć et al., 2014). Alguns dos trabalhos utilizaram apenas textos em inglês (KOVANOVIć et al., 2016; FARROW et al., 2019). Outros trabalharam apenas com o idioma português (NETO et al., 2018; CAVALCANTI et al., 2019; CAVALCANTI et al., 2020a; CAVALCANTI et al., 2020b).

O artigo de BARBOSA et al. (2020), mesmo que tenham abordado os dois idiomas, não utilizou a tradução de texto e analisou apenas a presença cognitiva, e não incluiu o *feedback* educacional . Já o trabalho de BARBOSA et al. (2021) utiliza a tradução, mas apenas um tradutor e não classifica os níveis de *feedback*.

O presente estudo destaca a classificação conjunta tanto para os níveis de *feedback* quanto de presença cognitiva, além de avaliar dois idiomas e a tradução deles por duas ferramentas distintas.

4 Metodologia

Este Capítulo descreve a metodologia utilizada para automatizar a análise dos níveis de *feedbak* em respostas a atividades e presença cognitiva em mensagens de fóruns de discussão assíncronas escritas em português e está organizado em cinco seções:

4.1 Etapas da Metodologia

A Figura 6 apresenta as etapas realizadas para a construção e avaliação do modelo proposto.

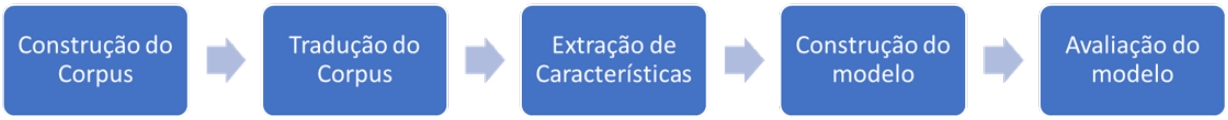

Figura 6 – Etapas da Metodologia

- **Construção do *corpus*** - processo de coleta das mensagens obtidas de fóruns online e mensagens de Ambiente Virtual de Aprendizagem (AVA) para a construção do *corpus* e sua respectivas anotações, segundo os níveis de *feedback* e níveis da presença cognitiva.
- **Tradução do *corpus*** - processo no qual são utilizadas as API do Google e Microsoft Azure para gerar 2 traduções para cada texto original;
- **Extração de Características** - utilização dos recursos para extração de características das mensagens contidas no *corpus* para construção do modelo;
- **Construção do modelo** - seleção, treinamento e otimização do classificador;
- **Avaliação do modelo** - avaliação do modelo proposto.

As tarefas realizadas em cada etapa são melhor explicadas nas próximas seções.

4.2 Descrição do *Corpus*

O presente estudo fez uso dos conjuntos de dados dos dois estudos anteriores que investigaram o processo de classificação das mensagens de discussão para os níveis de presença cognitiva.

Os conjuntos de dados em inglês e português dos estudos de (KOVANOVIć et al.,

2016) e (NETO et al., 2018) foram usados para construção dos *corpora*, plural de *corpus*, referentes à classificação automatizada dos níveis de presença cognitiva. O trabalho descrito em BARBOSA et al. (2020) também aplicou o mesmo conjunto de dados para propor um classificador multilíngue.

O conjunto de dados utilizados para construção dos *corpus* referentes à classificação automatizada dos níveis de *feedback* educacional é o mesmo utilizado por Cavalcanti et al. (2020a) (2020b). As próximas Seções apresentam mais detalhes sobre as bases utilizadas.

4.2.1 Bases de dados para classificação automática de presença cognitiva

Como mencionado na seção 3.1.1, vários estudos já foram feitos com o objetivo de analisar textos que apresentem conjunturas pertinentes à presença cognitiva, tanto de textos em inglês quanto textos em português. Utilizando os procedimentos para seleção de dados, o presente artigo segue as recomendações de (ROURKE et al., 2001) e (GARRISON et al., 2001). Utilizou-se as bases de presença cognitiva pois estão devidamente anotadas e com estudos que melhor instruem as configurações do algoritmo de classificação *Random Forest*.

O conjunto de dados em português abrange 1.500 mensagens de discussão produzidas por 215 alunos foi gerado durante quatro semanas de uma única oferta de um curso de graduação em biologia inteiramente online, oferecido em uma universidade pública brasileira. Neste caso, o fórum online centrou-se na discussão sobre uma questão / tópico levantado pelo formador, onde a participação correspondeu a 20 % da nota final do curso.

No processo de anotação, dois codificadores categorizaram cada mensagem com base nos indicadores das fases de presença cognitiva.Em termos de presença cognitiva, os codificadores obtiveram alta concordância entre avaliadores (percentual de concordância = 91,4 % e $kappa$ = 0,86). Um terceiro codificador resolveu as divergências.

O conjunto de dados em inglês contém 1.747 mensagens extraídas de seis ofertas (inverno de 2008, outono de 2008, verão de 2009, outono de 2009, inverno de 2010, inverno de 2011) de um curso intensivo de pesquisa de nível de mestrado em engenharia de software oferecido inteiramente online em uma universidade pública canadense. Assim 81 alunos gravaram e participaram de discussões online assíncronas sobre apresentações de vídeo sobre trabalhos de pesquisa relacionados a um dos tópicos do curso (ou seja, requisitos de software, design e manutenção). Esta atividade foi responsável por 15 % da nota final do curso (GAšEVIć et al.,

	Base de dados em Inglês		**Base de Dados em Português**	
Fase	**Mensagens**	**%**	**Mensagens**	**%**
Outros	140	8.01%	196	13.07%
Evento desencadeador	308	17.63%	235	15.67%
Exploração	684	39.15%	871	58.07%
Integração	508	29.08%	154	10.27%
Resolução	107	6.13%	44	2.92%
Total	1,747	100.00%	1,500	100.00%

Tabela 6 – Distribuição de mensagens para classificação de presença cognitiva

2015).

O processo de anotação para presença cognitiva seguiu uma abordagem semelhante. Usando o esquema de codificação proposto por (GARRISON et al., 2001), dois codificadores especialistas anotaram todas as mensagens de discussão online de acordo com os indicadores e fases de presença social e cognitiva, respectivamente. Em termos de presença cognitiva, a concordância entre os codificadores teve concordância percentual = 98,1 % e $kappa$ = 0,974, com apenas 32 discordâncias. As diferenças foram resolvidas por meio de discussão entre os codificadores.

Finalmente, seguindo (KOVANOVIC et al., 2014b; FERREIRA et al., 2018), as anotações sobre alguns dos indicadores (Continuando um tópico, Elogios e Vocativos) foram removidas porque quase todas as mensagens continham essas expressões. Além disso, a análise se concentrou no nível da categoria e atribuiu o rótulo final se a mensagem contiver pelo menos um indicador relacionado à categoria. A tabela 6 apresenta a distribuição final dos dados em português e inglês .

4.2.2 Bases de dados para classificação automática de *feedback* educacional

O conjunto de dados utilizado neste trabalho foi obtido de um AVA que oferece cursos *online* para uma universidade pública brasileira. O conjunto é composto por mensagens de texto que o professor respondeu atividades enviadas pelos seus alunos através da plataforma. Foram coletadas 1000 mensagens do curso de biologia (41) e literatura(959). A anotação foi

	Base de dados em Inglês			**Base de Dados em Português**		
Fase	**Classe 1**	**Classe 0**	**%**	**Classe 1**	**Classe 0**	**%**
FT	302	754	28.60%	888	112	88.80%
FP	380	676	35.98%	501	499	50.10%
FR	153	903	14.49%	8	992	0.80%
FS	221	835	20.93%	151	849	15.10%
Total	1056	-	-	1000	-	-

Tabela 7 – Distribuição de mensagens para classificação de $feedback$ educacional

realizada por especialistas seguindo as boas práticas descritas por Hattie e Timperley (2007) nos 4 níveis de feedback. Cada conteúdo de feedback foi analisado por dois codificadores especialistas separadamente. Após essa etapa, foram verificadas as diferenças entre cada dupla de especialistas. Outros dois especialistas que não participaram da primeira etapa resolveram os casos divergentes (27,8% do total). A concordância entre avaliadores teve uma porcentagem de 72,2% e a de $kappa$foi de 0,44%, valores suficientes para a experimentação de análise de conteúdo (CAVALCANTI et al., 2019; CAVALCANTI et al., 2020b).No final do processo de anotação, derivamos um conjunto de dados com classes binárias: classe 0 se uma mensagem de *feedback* não pertence a um determinado nível; classe 1 se a mensagem de feedback pertencer ao nível de *feedback*. Cada mensagem pode pertencer a mais de um nível de *feedback*.

Para a base de dados em Inglês, obtida de cursos de computação de universidades da Noruega e Escócia, 1056 mensagens receberam anotações realizadas por especialistas seguindo as boas práticas descritas por Hattie e Timperley (2007) nos 4 níveis de feedback.

A Tabela 7 mostra o conjunto de dados para esses níveis nos dois idiomas. Como o nível FR no idioma Português teve apenas oito exemplos de *feedback* de um total de 1.000, valor que não foi suficiente para caber em um modelo de aprendizado de máquina. Assim, optamos por utilizar apenas os demais níveis (FT , FP e FS) para os dois idiomas.

4.3 Tradução do *Corpus*

Nesta Seção, são explanadas as ferramentas de tradução utilizadas na presente pesquisa para extração de características de textos em diferentes idiomas. Também são apresentadas as siglas para abreviações dos bancos de dados para os textos originais e as traduções para cada

ferramenta.

4.3.1 Ferramentas Utilizadas

A linguagem de programação *Phyton*[1] (3.7) foi escolhida por apresentar bibliotecas que possibilitam a aplicação de várias abordagens necessárias para a pesquisa, como a tradução de texto, modelos de classificação e cálculos estatísticos para a avaliação e interpretação dos resultados. Essa mesma linguagem foi adotada em outras pesquisas (CAVALCANTI et al., 2020b; FERREIRA et al., 2018; BARBOSA et al., 2020).

Para a etapa de tradução, foram selecionadas duas fontes. Uma delas é a biblioteca gratuita que realiza chamadas de métodos utilizados na Interface de Programação de Aplicações, ou *Application Programming Interface* (API), do Google *Translate* [2]. Utilizando essa API, o algoritmo permite inserir um texto específico, indicar o idioma de origem e o idioma para qual o texto será traduzido, permitindo assim um rápido processamento e tradução em massa de textos.

A outra forma é o serviço de tradução fornecido por uma Transferência de Estado Representacional, ou *Representational State Transfer* (REST) da Microsoft Azure [3], que solicita cadastro na sua plataforma para escolher entre opções de assinatura gratuitas e pagas, dependendo da quantidade de palavras e sentenças a serem traduzidas em um intervalo de tempo. Com a assinatura feita, cria-se um recurso para fornecer uma chave de inscrição (*Subscription-Key*) e outros parâmetros, para configurar o algoritmo e torná-lo apto para realizar uma chamada à API REST referente ao serviço de tradução automática baseada em nuvem que responde com o texto traduzido para os idiomas indicados.

4.3.2 Abreviações para traduções

Ao todo, teremos informações sobre seis bases de dados, duas referentes aos *corporas* nos idiomas originais, sendo um para inglês e outro para português, além de duas traduções para cada *corpus*. Assim, para melhor distribuir as informações nas tabelas, abreviamos cada base com as seguintes siglas:

- **EN:** base de dados com os textos originais em Inglês, que contém 1747 mensagens para

[1] https://www.python.org/
[2] https://pypi.org/project/googletrans/
[3] https://api.cognitive.microsofttranslator.com/

presença cognitiva e 1056 mensagens para *feedback* educacional.

- **EN2PT Google:** base de dados traduzida dos textos originais em Inglês para o Português com a API do *Google Translate*.
- **EN2PT Azure:** base de dados traduzida dos textos originais em Inglês para o Português com a API do *Microsoft Azure*.
- **PT:** base de dados com os textos originais em Português, que contém 1500 mensagens para presença cognitiva e 1000 mensagens para *feedback* educacional..
- **PT2EN Google:** base de dados traduzida dos textos originais em Português para o Inglês com a API do *Google Translate*.
- **PT2EN Azure:** base de dados traduzida dos textos originais em Português para o Inglês com a API do *Microsoft Azure*.

Cabe ressaltar que cada item listado terá relação para a classificação quanto aos níveis de *feedback* educacional FT, FP e FS e para Presença Cognitiva.

4.4 Extração de Características

Para Shah e Patel (2016), a extração de características é o processo de gerar novas características a partir de informações existentes nos documentos, que permitam representar melhor os padrões do documento, gerando um classificador com resultados melhores. Esse processo pode ser realizado por métodos matemáticos, categorização por classes gramaticais e dependência entre palavras ou por meio da observação e raciocínio de especialistas de domínio (FELDMAN; SANGER, 2007; VIEIRA; LIMA, 2001; MARNEFFE et al., 2006).

A seguir, demonstraremos algumas das principais ferramentas utilizadas para extração de características para a classificação dos níveis de presença cognitiva e *feedback*, em respostas de atividades e fóruns de discussão *online* em AVAs.

4.4.1 Coh-Metrix

A ferramenta de análise textual em Inglês, Coh-Metrix, desenvolvida por pesquisadores da Universidade de Memphis - USA (GRAESSER et al., 2004), calcula índices de coesão e coerência textual sob determinadas medidas para indicar adequação de um texto ao público alvo ou identificar problemas textuais. Na sua versão livre mais recente para a lingua inglesa, Coh-

Metrix 3.0[4] , apresenta 108 índices, enquanto uma opção para a lingua portuguesa desenvolvida pela Universidade de São Paulo (USP), tendo como autor CUNHA (2015) e atualizada para a versão Coh-Metrix-PT 2.0 [5] por (SCARTON et al., 2010), contém 48 índices.

Devido aos bons resultados obtidos na classificação automatizada de fóruns de discussão, sob a perspectiva da presença cognitiva no trabalho de KOVANOVIĆ et al. (2016), o trabalho de CAMELO et al. (2020) demonstra a elaboração de uma nova versão pelo grupo de pesquisa da Aibox da UFRPE que replica 83 das principais categorias da ferramenta no idioma inglês e acrescenta 11 novas características referentes à incidência de conjunções conectivas próprias da língua portuguesa, totalizando 94 características. Assim, o presente estudo utilizou 83 características para o idioma inglês e 94 para o Português. Devido a grande quantidade de categorias, cada uma delas foram descritas na Tabela 8, e a seguir são explicados os 11 grupos das características do Coh-Metrix.

1. **Descritivo:** índices descritivos que tem 11 características envolvendo a quantidade e comprimento médio de parágrafos, frases, palavras, sílabas e letras.
2. **Coesão Referencial:** apresenta 12 características referentes à sobreposição de substantivos, argumento, radical, anáforas e palavras de conteúdo em frases adjacentes e em todas as sentenças;
3. **LSA *Latent Semantic Analysis*:** análise semântica latente que fornece 11 medidas de sobreposição semântica entre frases ou entre parágrafos.
4. **Diversidade Lexical:** apresenta 4 características referentes à variedade de palavras únicas (tipos) que ocorrem em um texto em relação ao número total de palavras (tokens).
5. **Conectivos:** apresenta 8 características de ligações coesivas entre ideias e proposições e fornecem pistas sobre a organização do texto, em diferentes tipos: causal, lógico, adversos, temporal, aditivo, positivo e negativo.
6. **Modelo de Situação:** apresenta 5 características presentes na representação mental do comprador sob um determinado.
7. **Complexidade sintática:** apresenta 6 características referentes à distância ou similaridade entre classe gramatical, palavras e lemas.
8. **Densidade de padrão sintático:** apresenta 8 características referentes à incidência de frases nominal, verbal, adverbial, de preposição, voz passiva, negação, gerúndio e infinitivo.

[4] http://cohmetrix.memphis.edu/cohmetrixhome/documentation_indices.html
[5] http://143.107.183.175:22680/

9. **Informação da Palavra:** apresenta 18 características referentes à incidência de palavras pertencentes à determinadas categorias como palavras de conteúdo (por exemplo, substantivos, verbos, adjetivos, advérbios) e palavras funcionais (por exemplo, preposições, determinantes, pronomes).
10. **Legibilidade:** apresenta 3 características referentes à diferentes fórmulas para avaliar a legibilidade de um texto.
11. **Conectivos PT:** apresenta 11 características referentes à conectivos próprios da língua portuguesa.

Descritivo	
cm.DESPC	Contagem de parágrafos, número de parágrafos
cm.DESSC	Contagem de frases, número de frases
cm.DESWC	Contagem de palavras, número de palavras
cm.DESPL	Comprimento do parágrafo, número de frases, média
cm.DESPLd	Comprimento do parágrafo, número de frases, desvio padrão
cm.DESSL	Comprimento da frase, número de palavras, média
cm.DESSLd	Comprimento da frase, número de palavras, desvio padrão
cm.DESWLsy	Comprimento da palavra, número de sílabas, média
cm.DESWLsyd	Comprimento da palavra, número de sílabas, desvio padrão
cm.DESWLlt	Comprimento da palavra, número de letras, significa
cm.DESWLltd	Comprimento da palavra, número de letras, desvio padrão
Coesão Referencial	
cm.CRFNO1	Sobreposição de substantivos, sentenças adjacentes, binários, significam
cm.CRFAO1	Sobreposição de argumento, sentenças adjacentes, binário, média
cm.CRFSO1	Sobreposição de caule, sentenças adjacentes, binário, médio
cm.CRFNOa	Sobreposição de substantivos, todas as sentenças, binários, média
cm.CRFAOa	Sobreposição de argumentos, todas as sentenças, binárias, média
cm.CRFSOa	Sobreposição de conteúdo, todas as sentenças, binárias, média
CRFCWO1	Sobreposição de palavra do conteúdo, frases adjacentes, proporcional, média
CRFCWO1d	Sobreposição de palavra do conteúdo, frases adjacentes, proporcional, desvio padrão
CRFCWOa	Sobreposição de palavra do conteúdo, todas as sentenças, proporcional, significa
CRFCWOad	Sobreposição de palavra do conteúdo, todas as frases, proporcional, desvio padrão
cm.CRFANP1	Sobreposição de anáforas, frases adjacentes
cm.CRFANPa	Sobreposição de anáforas, todas as sentenças
LSA	
cm.LSASS1	LSA sobreposição, sentenças adjacentes, média
cm.LSASS1d	Sobreposição de LSA, frases adjacentes, desvio padrão
cm.LSASSp	LSA se sobrepõem, todas as sentenças do parágrafo média
cm.LSASSpd	LSA se sobrepõem, todas as sentenças no parágrafo, desvio padrão
cm.LSAPP1	Sobreposição LSA, parágrafos adjacentes, média
cm.LSAPP1d	Sobreposição LSA, parágrafos adjacentes, desvio padrão
cm.LSAGN	LSA dado / novo, frases, média
cm.LSAGNd	LSA dado / novo, frases, desvio padrão
Diversidade Lexical	
cm.LDTTRc	Diversidade lexical, proporção de tipo-símbolo, lemas de palavras de conteúdo
cm.LDTTRa	Diversidade lexical, relação tipo-token, todas as palavras
cm.LDMTLDa	Diversidade lexical, MTLD, todas as palavras
cm.LDVOCDa	Diversidade lexical, VOCD, todas as palavras
Conectivos	
cm.CNCAll	Incidência de todos os conectivos
cm.CNCCaus	Incidência de conectivos causais
cm.CNCLogic	Incidência de conectivos lógicos
cm.CNCADC	Incidência de conectivos adversos e contrastivos
cm.CNCTemp	Incidência de conectivos temporais
cm.CNCAdd	Incidência de conectivos aditivos
cm.CNCPos	Incidência de conectivos positivos
cm.CNCNeg	Incidência de conectivos negativos
Legibilidade	
cm.RDFRE	Flesch Reading Facilidade
cm.RDFKGL	Nível de classificação Flesch-Kincaid
cm.RDL2	Legibilidade Coh-Metrix L2

Modelo de Situação	
cm.SMCAUSv	Incidência de verbo causal
cm.SMCAUSvp	Incidência de verbos causais e partículas causais
cm.SMCAUSr	Razão de partículas casuais para verbos causais
cm.SMCAUSlsa	LSA verbo sobreposição
cm.SMCAUSwn	Sobreposição de verbos do WordNet
Complexidade sintática	
cm.SYNLE	Palavras antes do verbo principal, média
cm.SYNMEDpos	Distância mínima de edição, classe gramatical
cm.SYNMEDwrd	Distância de edição mínima, todas as palavras
cm.SYNMEDlem	Distância de edição mínima, lemas
cm.SYNSTRUTa	Semelhança de sintaxe de frase, frases adjacentes, média.
cm.SYNSTRUTt	Similaridade de sintaxe de frase, todas as combinações, entre parágrafos, média
Densidade de padrão sintático	
cm.DRNP	Densidade de frase nominal, incidência
cm.DRVP	Densidade de frase verbal, incidência
cm.DRAP	Densidade de frase adverbial, incidência
cm.DRPP	Densidade de frase de preposição, incidência
cm.DRPVAL	Densidade de voz passiva sem agente, incidência
cm.DRNEG	Densidade de negação, incidência
cmDRGERUND	Densidade de gerúndios, incidência
cm.DRINF	Densidade infinitiva, incidência
Informação da Palavra	
cm.WRDNOUN	Incidência nominal
cm.WRDVERB	Incidência de verbos
cm.WRDADJ	Incidência de adjetivo
cm.WRDADV	Incidência de advérbio
cm.WRDPRO	Incidência de pronome
cm.WRDPRP1s	Incidência de pronome de primeira pessoa do singular
cm.WRDPRP1p	Incidência de pronome na primeira pessoa do plural
cm.WRDPRP2	Incidência de pronome de segunda pessoa
cmWRDPRP3s	Incidência de pronome de terceira pessoa do singular
cmWRDPRP3p	Incidência de pronome de terceira pessoa no plural
cm.WRDFRQc	Frequência de palavras CELEX para palavras de conteúdo, média
cm.WRDFRQa	Frequência de log CELEX para todas as palavras, média
cmWRDFRQmc	CELEX Log de frequência mínima para palavras de conteúdo, média
cm.WRDAOAc	Idade de aquisição para palavras de conteúdo, média
cm.WRDFAMc	Familiaridade para palavras de conteúdo, média
cm.WRDCNCc	Concretude para palavras de conteúdo, média
cm.WRDIMGc	Imagabilidade para palavras de conteúdo, média
cm.WRDMEAc	Significância, normas do Colorado, palavras de conteúdo, média
Conectivos PT	
cm.CNCAlter	Incidência de conjunções alternativas
cm.CNCConclu	Incidência de conjunções conclusivas
cm.CNCExpli	Incidência de conjunções explicativas
cm.CNCConce	Incidência de conjunções concessiva
cm.CNCCondi	Incidência de conjunções condicional
cm.CNCConfor	Incidência de conjunções conformativas
cm.CNCFinal	Incidência de conjunções finais
cm.CNCProp	Incidência de conjunções proporcionais
cm.CNCComp	Incidência de conjunções comparativas
cm.CNCConse	Incidência de conjunções consecutivas
cm.CNCInte	Incidência de conjunções integrantes

Tabela 8 – Abreviações e descrições das Características Coh-Metrix
(Adaptado da documentação oficial do Coh-Metrix (2020))

4.4.2 LIWC

A ferramenta LIWC desenvolvida por PENNEBAKER et al. (2001) com o objetivo de estudar diferentes indicadores de processos psicológicos, dentre eles o cognitivo, presente em diálogo oral ou textual. Com o tempo, foi desenvolvida uma biblioteca com milhares de palavras em inglês associadas inicialmente a 73 categorias, podendo cada palavra pertencer a mais de uma categoria. Atualmente, para o idioma inglês, são designadas 93 categorias.

Percebendo o potencial da ferramenta que conta com grande variação de características, possibilitando mais recursos para classificações com maior precisão, o Centro Interinstitucional de Linguística da Computação (NILC) e o Instituto de Matemática e Ciências da Computação da Universidadede São Paulo, juntaram esforços em 2007 para gerar um dicionário LIWC da língua portuguesa, atualmente disponíveis no site PorLex[6]. Com esse novo dicionário foram designadas 64 características (FILHO et al., 2013), incluindo as que alcançaram melhores resultados de confiabilidade e Cohen's K no trabalho de (KOVANOVIĆ et al., 2016). Devido a grande quantidade de categorias, cada uma delas foram descritas no quadro da Figura 7. A seguir são explicados os 7 grupos das características do LIWC.

- **Contagem de Palavras:** apresenta 1 característica referente à média da quantidade de palavras.
- **Variáveis de linguagem de resumo:** apresenta 7 características referentes à linguagem resumida (pensamento analítico, influência, autenticidade e tom emocional) e descritores gerais (palavras por frase, no dicionário, e com mais de 6 letras).
- **Dimensões Linguísticas:** apresenta 15 características referentes ao total de palavras de função, pronomes, artigos, sobreposições, verbos auxiliares, advérbios, conjunções e negações.
- **Outra Gramática:** apresenta 6 características referentes a outras categorias gramaticais (verbos, advérbios, comparações, interrogativos, numéricos e quantificadores).
- **Processos Psicológicos:** apresenta 33 características referentes a palavras que exploram construções psicológicas (por exemplo, afeto, cognição, processos biológicos, impulsos).
- **Orientações de tempo:** apresenta 7 características referentes à orientação geral de tempo em vez de apenas o uso do tempo verbal.
- **Preocupações pessoais:** apresenta 27 características referentes à incidência de palavras que representam atividades pessoais (por exemplo, trabalho, casa, lazer, religião).

[6] http://www.nilc.icmc.usp.br/portlex/index.php/en/liwc

Abreviações e Descrições das Features LIWC					
Média de Palavras		**Processos Psicológicos**		**Orientações de tempo**	
liwc.WC	Contagem de palavras	liwc.affect	Processos afetivos	liwc.focuspast	Foco passado
Variáveis de linguagem de resumo		liwc.posemo	Emoção positiva	liwc.focuspresent	Foco atual
liwc.Analytic	Analítico	liwc.negemo	Emoção negativa	liwc.focusfuture	Foco futuro
liwc.Clout	Influência	liwc.anx	Ansiedade	liwc.relativ	Relatividade
liwc.Authentic	Autêntico	liwc.anger	Raiva	liwc.motion	Movimento
liwc.Tone	Tom	liwc.sad	Tristeza	liwc.space	Espaço
liwc.WPS	Palavras / frase	liwc.social	Processos sociais	liwc.time	Tempo
liwc.SixLtr	Palavras> 6 letras	liwc.family	Família	**Preocupações pessoais**	
liwc.Dic	Palavras do dicionário	liwc.friend	Amigos	liwc.work	Trabalhos
Dimensões Linguísticas		liwc.female	Referências femininas	liwc.leisure	Lazer
liwc.funct	Total de palavras de função	liwc.male	Referências masculinas	liwc.home	Lar
liwc.pronoun	Total de pronomes	liwc.cogproc	Processos cognitivos	liwc.money	Dinheiro
liwc.ppron	Pronomes pessoais	liwc.insight	Entendimento	liwc.relig	Religião
liwc.i	1ª pessoa do singular	liwc.cause	Causalidade	liwc.death	Morte
liwc.we	1ª pessoa do plural	liwc.discrep	Discrepância	liwc.informal	Linguagem informal
liwc.you	2ª pessoa	liwc.tentat	Provisório	liwc.swear	Palavrões
liwc.shehe	3ª pessoa do singular	liwc.certain	Certeza	liwc.netspeak	Netspeak
liwc.they	3ª pessoa do plural	liwc.differ	Diferenciação	liwc.assent	Consentimento
liwc.ipron	Pronomes impessoais	liwc.percept	Processos perceptivos	liwc.nonflu	Não-fluências
liwc.article	Artigos	liwc.see	Ver	liwc.filler	Enchimentos
liwc.prep	Preposições	liwc.hear	Ouvir	liwc.AllPunc	Pontuação
liwc.auxverb	Verbos auxiliares	liwc.feel	Sentir	liwc.Period	Períodos
liwc.adverb	Advérbios Comuns	liwc.bio	Processos biológicos	liwc.Comma	Vírgula
liwc.conj	Conjunções	liwc.body	Corpo	liwc.Colon	Dois pontos
liwc.negate	Negações	liwc.health	Saúde	liwc.SemiC	Ponto e vírgula
Outra Gramática		liwc.sexual	Sexual	liwc.QMark	Pontos de interrogação
liwc.verb	Verbos comuns	liwc.ingest	Ingestão	liwc.Exclam	Pontos de exclamação
liwc.adj	Adjetivos comuns	liwc.drives	Drives	liwc.Dash	Traços
liwc.compare	Comparações	liwc.affiliation	Afiliação	liwc.Quote	Aspas
liwc.interrog	Interrogativos	liwc.achieve	Conquista	liwc.Apostro	Apóstrofes
liwc.number	Números	liwc.power	Poder	liwc.Parenth	Parênteses (pares)
liwc.quant	Quantificadores	liwc.reward	Recompensa	liwc.OtherP	Outra pontuação
		liwc.risk	Risco	liwc.humans	Referência à pessoas
				liwc.inhib	Inibição
				liwc.incl	Inclusão

Figura 7 – Abreviações e descrições das Características LIWC
(adaptado de PENNEBAKER et al. (2017))

4.4.3 *Social Network Analysis* (SNA)

A análise de redes sociais, SNA tem sido amplamente utilizada em pesquisas que buscam medir o nível de interação social entre os participantes de redes de interação, como as discussões online (LIU. et al., 2018; GAŠEVIĆ et al., 2018; OLIVARES et al., 2019). SNA permite identificar quais participantes são mais influentes e / ou intermediários em uma rede de interação. Portanto, as últimas características de classificação em nossa lista foram as medidas SNA mais comumente consideradas:

- ***Closeness centrality (sna.closeness)*** avalia o quão próximo um aluno da rede está de todos os outros alunos. Segundo (YUSOF et al., 2009), os alunos com maior grau de proximidade tendem a ser eficazes na divulgação de informações pela rede;
- ***Betweenness centrality (sna.betweenness)*** investiga a importância da intervenção de um aluno nas interações entre outros alunos. Portanto, um alto nível de intermediação indica a liderança do aluno entre os pares;
- ***Degree centrality (sna.degree)*** busca analisar o nível de interação dos alunos com os demais alunos presentes na rede. Assim, um maior grau de centralidade de um aluno significa que o aluno é mais influente na rede.

4.4.4 *Discussion Context Features* (DCF)

Segundo GARRISON et al. (2010) a estrutura da discussão poderia indicar aspectos da presença social e cognitiva. Por exemplo, dado que a presença cognitiva se desenvolve ao longo do tempo por meio do discurso e da reflexão, as fases de integração e resolução tendem a acontecer nas mensagens posteriores. Essas informações podem ser capturadas por esses recursos. Utilizam-se DCF como indicadores que captam o contexto e a estrutura das discussões online, conforme proposto inicialmente por KOVANOVIć et al. (2016):

- **Profundidade da mensagem (dcf.depth)**, refere-se à posição numérica da mensagem dentro a discussão;
- **Número de respostas (dcf.numReplies)**, consiste na contagem de respostas que cada mensagem da base de dados recebeu;
- **Semelhança do cosseno com a mensagem anterior / seguinte (dcf.sim.prev e dcf.sim.next)**, refere-se à semelhança do cosseno do texto de cada mensagem com a mensagem que a precede e a sucede;

- **Iniciar e terminar a discussão (dcf.first and dcf.last)**, um valor binário que indica se uma mensagem está iniciando ou terminando uma discussão.

4.5 Construção e Avaliação do Modelo

Um modelo de classificação que recebe as características extraídas e categoriza novas instâncias de acordo com as classes para qual foi treinado. Assim, a escolha do algoritmo de classificação é uma etapa importante para se obter os melhores resultados alinhados ao seu objetivo específico, variando o desempenho de acordo com determinado contexto, combinação de variáveis e métricas.

Como o presente estudo não visa comparar classificadores, então se decidiu utilizar o algoritmo *Random Forest*, alinhando-se às práticas da maioria dos trabalhos mais recentes mencionados no Capítulo 3, tanto para classificação de presença cognitiva quanto *feedback* educacional, e apresentou bom desempenho, além de ser um algoritmo caixa branca, que é possível avaliar até que ponto cada característica contribui para o classificador (FARROW et al., 2019; BARBOSA et al., 2021; CAVALCANTI et al., 2019; CAVALCANTI et al., 2020b). Das medidas utilizadas para avaliação da importância das características, a mais utilizada é o MDG, que explica a separabilidade de uma determinada característica em relação às categorias (BREIMAN, 2001).

Para criar o modelo de classificação do tipo RF, treinar e executar testes, utiliza-se a biblioteca Sckit-Learn [7] que integra disversos algorítmos de Aprendizado de Máquina (PEDREGOSA et al., 2011), incluindo outros classificadores, como SVM e Rede Neural. Já a biblioteca *SciPy*[8] fornece ao usuário comandos e classes de alto nível para manipulação e simplifica resultados de cálculos estatísticos complexos citados na 2.3.3.2.

Foram configurados os parâmetros de $ntree$ e $mtry$ para os classificadores. Para os textos de *feedback* educacional, estudos apontaram para o valor de $ntree = 500$ (CAVALCANTI et al., 2020a; CAVALCANTI et al., 2020b). Já para os textos de presença cognitiva, pesquisas apontaram para o valor de $ntree = 1000$ para certificar que o grau de convergência seja alcançada (KOVANOVIĆ et al., 2016).

Como os estudos anteriores se dedicavam a apenas um idioma, não havia variação do tipo e no número de características, então adotou-se para o $mtry$ o valor padrão *None* da

[7] https://scikit-learn.org/stable/index.html
[8] https://scipy.org/

ferramenta, utilizando assim todas as características disponíveis para cada cenário no processo de classificação.

Para a separação dos dados em conjuntos de treinamento e teste, adotou-se a técnica de CV 10- *fold* estratificada levando em consideração as recomendações propostas por FARROW et al. (2019) e BARBOSA et al. (2021).

Para avaliar os classificadores, adotou-se a acurácia e o *kappa* porque eles são amplamente usados em análise de aprendizagem e mineração de dados educacionais e nos trabalhos relacionados do Capítulo 3. Para cada caso, o classificador foi executado 30 vezes a fim de criar diferentes amostras e aplicar testes estatísticos para verificar se as amostras apresentaram resultados significativamente diferentes e com isso conseguir responder às perguntas de pesquisa.

Com esses valores, é possível realizar um teste estatístico para comparar o desempenho dos classificadores relacionados ao idioma original e os resultados obtidos com as traduções.

Assim, tornou-se necessária a realização de testes estatísticos para comprovar o real aumento ou diminuição dos resultados obtidos no processo de classificação para o idioma original e para as traduções. Para isso, primeiramente realizam-se o teste Shapiro-Wilk, para verificar se os dados dos conjuntos apresentados seguem distribuição normal (RAZALI et al., 2011). O nível de significância estabelecido foi de 95%, e assim a amostra é considerada com distribuição normal caso o valor de p do teste, seja maior que 0,05.

Posteriormente, realizam-se o teste de hipótese. A hipótese levantada é que as médias referentes às métricas encontradas como resultado das classificações são iguais. A Hipótese nula H_0 indica que a média da amostra referente à métrica analisada é igual à outra amostra. Já a Hipótese alternativa H_1 indica que as médias analisadas são estatisticamente diferentes, como podemos ver na Equação 4.1.

$$H_0 : \mu_{a1} = \mu_{a2} \text{ oposto à } H_1 : \mu_{a1} \neq \mu_{a2} \tag{4.1}$$

μ_{a1}= Média da Métrica extraída da amostra 1.

μ_{a2}= Média da Métrica extraída da amostra2.

Quando todas as amostras comparadas apresentam distribuição normal, o teste de hipótese utilizado é o *T-Student* (ZIMMERMAN, 1987), e caso uma das amostras não apresente distribuição normal, é apropriado utiliza o teste *Mann-Whitney* bicaudal, uma versão não paramétrica do *T-Student* quando se trabalha com amostras independentes (MCKNIGHT; NAJAB, 2010). Como resultado do teste *Mann-Whitney*, o valor de p maior que 0,05 indica com nível

de significância de 95% validando a hipótese nula H_0, e caso o valor de p seja menor que 0,05, rejeita-se H_0 e considera-se válida a hipótese alternativa H_1.

Como não foi encontrada normailidade em dois dos 24 processos de clasificação, evitou-se fazer testes estatísticos comparativos diferentes quando envolviam esses conjuntos, preferiu-se usar o teste estatístico não paramétrico *Mann-Whitney* em todos os cenários, assim como no artigo de BARBOSA et al. (2021).

5 Resultados

Serão apresentados os dados decorrentes das classificações para análise das métricas relevantes para determinar a eficácia da tradução automática de textos em relação ao idioma original e entre as traduções. Também, são exibidas as *features* mais relevantes para o classificador em cada domínio de cada *corpus* utilizado. Com os devidos dados apresentados, são apresentadas discussões que possibilitam responder às perguntas de pesquisa.

5.1 Resultados de Classificação e Estatísticas

Nesta seção serão apresentados os resultados obtidos com a classificação automatizada dos textos nos idiomas originais e suas traduções para os níveis de *feedback* educacional e níveis de presença cognitiva

5.1.1 Questão de Pesquisa 1

Inicialmente, avaliamos a influência da tradução de textos na classificação final dos níveis de *feedback* e presenças cognitivas. A tabela 9 apresenta a média e desvio padrão da acurácia e os resultados do coeficiente *kappa* obtidos a partir de ciclos de 30 execuções, por cada um dos classificadores propostos. Pelos resultados médios de acurácia e $Kappa$ demonstrados na Tabela 9, foram elaboradas as Tabelas 10 e 11 sintetizando os resultados das médias e testes estatísticos *Mann-Whitney* que ajudam a melhor compreender as informações explicadas nos seguintes itens. Estão destacados em negrito os valores que apresentam vantagem estatisticamente significativa da outra média, caso o valor de p seja maior que 0.05 os dois valores se encontram em negrito.

1. Quando leva-se em conta a acurácia das traduções do Inglês para o Português, os dois tradutores apontaram 2 cenários de diminuição significativa da tradução, na classificação de presença cognitiva, FP e FS. Para a classificação e FT o resultado de tradução do Google teve aumento significativo, e do Azure foi estatisticamente igual.
2. Quando leva-se em conta a acurácia das traduções do Português para o Inglês, todos os cenários de tradução apresentaram valor médio maior que no idioma original, e somente a tradução do Google para classificação de FT e a tradução do Azure para a classificação de FP apresentaram resultados estatisticamente iguais. Os 6 cenários restantes indicam que

	FT				**FP**			
	Acurácia		**Kappa**		**Acurácia**		**Kappa**	
EN	0.709	(0.004)	0.028	(0.012)	0.659	(0.005)	0.115	(0.014)
EN2PT Google	0.711	(0.004)	0.037	(0.010)	0.648	(0.006)	0.095	(0.016)
EN2PT Azure	0.708	(0.004)	0.035	(0.012)	0.655	(0.005)	0.116	(0.013)
PT	0.901	(0.004)	0.322	(0.031)	0.778	(0.007)	0.555	(0.014)
PT2EN Google	0.903	(0.004)	0.340	(0.031)	0.783	(0.008)	0.566	(0.016)
PT2EN Azure	0.903	(0.003)	0.342	(0.026)	0.779	(0.007)	0.558	(0.015)
	FS				**Presença Cognitiva**			
	Acurácia		**Kappa**		**Acurácia**		**Kappa**	
EN	0.807	(0.003)	0.183	(0.018)	0.585	(0.007)	0.385	(0.011)
EN2PT Google	0.801	(0.004)	0.165	(0.016)	0.575	(0.006)	0.381	(0.008)
EN2PT Azure	0.804	(0.003)	0.176	(0.014)	0.564	(0.006)	0.370	(0.008)
PT	0.901	(0.004)	0.546	(0.022)	0.739	(0.005)	0.520	(0.011)
PT2EN Google	0.922	(0.004)	0.687	(0.018)	0.830	(0.005)	0.697	(0.008)
PT2EN Azure	0.923	(0.004)	0.683	(0.016)	0.833	(0.005)	0.712	(0.008)

Tabela 9 – Resultados obtidos para níveis de *feedback* e presença cognitiva.

houve aumento dos resultados com a tradução para o Inglês.

3. Quando leva-se em conta o *kappa* das traduções do Inglês para o Português, os dois tradutores apontam aumento dos resultados médios para classificação de FT. Já nos cenários de classificação para FP e FS do Google, e de presença cognitiva do Azure, houve diminuição significativa na tradução. Os três cenários restantes validaram H_0 apontando igualdade entre as amostras.
4. Quando leva-se em conta o *kappa* das traduções do Português para o Inglês, todos os cenários de classificação apresentaram valor médio maior que no idioma original, a maioria com significância estatística comprovada, menos o cenário de classificação de FP que considera os resultados médios das amostras iguais.

Apesar de ter resultados de *kappa* que em termos estatísticos indicam níveis de concordância baixo, a maioria dos valores encontrados assemelham-se aos resultados descritos

	EN2PT Google			**EN2PT Azure**			**PT2EN Google**			**PT2EN Azure**		
	EN	**p**	**PT**	**EN**	**p**	**PT**	**PT**	**p**	**EN**	**PT**	**p**	**EN**
FT	0.709	0.043	**0.711**	**0.709**	**0.739**	**0.708**	**0.901**	**0.092**	**0.903**	0.901	0.045	**0.903**
FP	**0.659**	0.000	0.648	**0.659**	0.007	0.655	0.778	0.005	**0.783**	**0.778**	**0.294**	**0.779**
FS	**0.807**	0.000	0.801	**0.807**	0.010	0.804	0.901	0	**0.922**	0.901	0	**0.923**
Cognitivo	**0.585**	0.000	0.575	**0.585**	0	0.564	0.739	0	**0.830**	0.739	0	**0.833**

Tabela 10 – Médias e estatísticas para acurácia

	EN2PT Google			**EN2PT Azure**			**PT2EN Google**			**PT2EN Azure**		
	EN	**p**	**PT**	**EN**	**p**	**PT**	**PT**	**p**	**EN**	**PT**	**p**	**EN**
FT	0.028	0.005	**0.037**	0.028	0.036	**0.035**	0.322	0.033	**0.340**	0.322	0.011	**0.342**
FP	**0.115**	0.000	0.095	**0.115**	**0.959**	**0.116**	0.555	0.004	**0.566**	**0.555**	**0.311**	**0.558**
FS	**0.183**	0.000	0.165	**0.183**	**0.088**	**0.176**	0.546	0.000	**0.687**	0.546	0.000	**0.683**
Cognitivo	**0.385**	**0.290**	**0.381**	**0.385**	0.000	0.370	0.520	0.000	**0.697**	0.520	0.000	**0.712**

Tabela 11 – Médias e estatísticas para kappa

nos estudos das Subseções 3.1.1 e 3.2, indicando uma tendência dos tipos de estudo de mineração de dados para textos educacionais.

5.1.2 Questão de Pesquisa 2

O presente estudo adotou os mesmos conjuntos de características e configurações para os classificadores e metodologias usadas na literatura (FERREIRA et al., 2018; BARBOSA et al., 2021), conforme mencionado no Capítulo da 4. Assim, classificou-se as características com base na medida do índice de impureza de Gini de redução média (MDG). A seguir, reportamos o ranking com os 20 principais recursos para a classificação dos níveis de *feedback* (FT, FT e FS, e também para a presença cognitiva, ocasionando 8 cenários onde cada um deles engloba 3 resultados (textos originais e suas duas traduções), totalizando 24 classificações e 480 características ranqueadas. Também vale ressaltar que para cada cenário foram deixados em negrito as características que aparecem em mais de uma classificação.

No primeiro cenário apresentado na Tabela 12, analisamos o nível de *feedback* FT para o texto em Inglês e suas respectivas traduções. No idioma Inglês, vemos uma diversidade contando com 11 características da ferramenta Coh-Metrix e 9 características para a ferramenta LIWC. Já nos resultados para as traduções em português, todas as características são da ferramenta

Coh-Metrix.

Ao todos são 51 características da ferramenta Coh-Metrix somando 86.63 de MDG e 9 características para a ferramenta LIWC somando 12.21 de MDG. Vale ressaltar que as características: ”cm.DESSLd”, ”cm.LSAGNd”, ”cm.LSASS1d”e ”cm.RDL2”estão presentes nasclassificações de FT, tanto para o idioma original, quanto para as duas traduções.

No segundo cenário apresentado na Tabela 13, analisamos o nível de *feedback* FP para o texto em Inglês e suas respectivas traduções. Os resultados desse cenário foram similares aos do cenário anterior. No idioma Inglês, vemos uma diversidade contabilizando 11 características da ferramenta Coh-Metrix e 9 características para a ferramenta LIWC. Já nas suas traduções para o idioma Português, todas as 20 principais características foram da ferramenta Coh-Metrix.

Ao todo são 51 características da ferramenta Coh-Metrix somando 81.09 de MDG e 9 características para a ferramenta LIWC somando 11.29 de MDG. Vale ressaltar que as características ”cm.DESWLsyd”, ”cm.LSAGNd”, ”cm.LSASS1d”, ”cm.WRDIMGc”estão presentes nas três classificações de FP.

No terceiro cenário apresentado na Tabela 14, analisamos o nível de *feedback* FS para o texto em Inglês e suas respectivas traduções. No idioma Inglês, contabilizamos 11 características para a ferramenta Coh-Metrix e 9 características para a ferramenta LIWC. Na tradução para o idioma Português usando a API do Google Translate, contabilizamos 17 características para a ferramenta Coh-Metrix e 3 características para a ferramenta LIWC. Já na tradução para o idioma Português usando a API do Microsoft Azure, contabilizamos 18 características para a ferramenta Coh-Metrix e 2 características para a ferramenta LIWC.

Ao todo são 46 características da ferramenta Coh-Metrix somando 70.78 de MDG e 14 características para a ferramenta LIWC somando 33.42 de MDG. Vale ressaltar que as características ”cm.DESSLd”, ”cm.DESWLsyd”, ”cm.WRDCNCc”, ”cm.WRDFAMc”, ”liwc.friend”estão presentes nas três classificações de FS.

No quarto cenário apresentado na Tabela 15, analisamos o nível de *feedback* FT para o texto em Português e suas respectivas traduções. No idioma Português, contabiliza-se todas as 20 principais características da ferramenta Coh-Metrix. Na tradução para o idioma Inglês usando a API do Google Translate, contabilizamos 9 características para a ferramenta Coh-Metrix e 11 características para a ferramenta LIWC. Já na tradução para o idioma Português usando a API do Microsoft Azure, contabilizamos 11 características para a ferramenta Coh-Metrix e 9 características para a ferramenta LIWC.

Dataset English - FT					
EN_FT		**EN2PT Google_FT**		**EN2PT Azure_FT**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
liwc.Analytic	1.92	**cm.LSASSpd**	2.72	**cm.WRDFRQmc**	2.76
cm.RDFRE	1.84	**cm.SYNMEDpos**	2.60	**cm.LSASSpd**	2.34
liwc.work	1.69	**cm.LSASS1d**	2.42	**cm.SYNMEDpos**	2.25
cm.LSAGNd	1.68	**cm.WRDAOAc**	2.21	**cm.WRDAOAc**	2.21
cm.RDL2	1.46	**cm.RDL2**	1.85	**cm.LSAGNd**	2.06
cm.WRDIMGc	1.43	**cm.WRDFRQmc**	1.83	**cm.WRDMEAc**	1.90
liwc.prep	1.36	**cm.WRDMEAc**	1.82	**cm.WRDFAMc**	1.89
cm.SYNSTRUTt	1.31	**cm.SYNLE**	1.79	**cm.LSASS1**	1.87
cm.LSASS1	1.31	**cm.DESSLd**	1.71	**cm.DESSLd**	1.75
liwc.ipron	1.31	cm.WRDCNCc	1.68	**cm.DESWLltd**	1.64
liwc.pronoun	1.26	**cm.WRDFAMc**	1.67	**cm.LSASS1d**	1.60
cm.LSASS1d	1.26	cm.CRFCWO1	1.66	**cm.DESWLsyd**	1.58
cm.SYNMEDlem	1.25	**cm.CRFCWOad**	1.62	**cm.RDL2**	1.53
cm.RDFKGL	1.20	**cm.DESWLsyd**	1.57	**cm.LDTTRc**	1.51
cm.DESSLd	1.20	**cm.DESWLltd**	1.53	**cm.LSASSp**	1.50
liwc.Authentic	1.19	**cm.WRDIMGc**	1.53	cm.CRFCWOa	1.48
liwc.time	1.19	**cm.LSAGNd**	1.52	**cm.CRFCWOad**	1.47
liwc.space	1.18	**cm.SMCAUSlsa**	1.46	cm.CRFCWO1d	1.45
liwc.article	1.11	**cm.LSASSp**	1.42	**cm.SMCAUSlsa**	1.44
cm.LDTTRc	1.07	cm.DESWLsy	1.41	**cm.SYNLE**	1.37

Tabela 12 – Vinte *features* mais importantes para classificação de FT pelo MDG.

Ao todo são 40 características da ferramenta Coh-Metrix somando 95.65 de MDG e 20 características para a ferramenta LIWC somando 48.65 de MDG. Vale ressaltar que as características ”cm.DESSLd”, ”cm.DESWC”, ”cm.DESWLlt”, ”cm.DESWLltd”e ”cm.DESWLsyd”estão presentes nas três classificações de FT.

Dataset English - FP					
EN_FP		**EN2PT Google_FP**		**EN2PT Azure_FP**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
cm.WRDPRP1p	1.80	**cm.WRDCNCc**	2.38	**cm.SYNLE**	2.57
cm.LDTTRa	1.59	**cm.WRDFAMc**	2.11	**cm.DESWLltd**	2.54
liwc.prep	1.40	**cm.DESWLltd**	2.11	**cm.WRDFAMc**	2.17
cm.WRDIMGc	1.39	**cm.WRDMEAc**	1.73	**cm.LSASS1d**	2.17
cm.DESWLsyd	1.35	**cm.SYNLE**	1.72	**cm.DESWLsyd**	2.06
cm.LSAGNd	1.35	**cm.DESWLsyd**	1.69	**cm.WRDCNCc**	1.99
liwc.work	1.34	**cm.SYNMEDpos**	1.67	**cm.SYNMEDpos**	1.84
liwc.adverb	1.32	**cm.CRFCWOa**	1.65	**cm.LSAGNd**	1.73
liwc.insight	1.30	**cm.WRDFRQmc**	1.52	**cm.DESSLd**	1.64
liwc.drives	1.26	**cm.CRFCWO1**	1.50	**cm.SMCAUSlsa**	1.56
cm.LDTTRc	1.25	**cm.LSAGNd**	1.49	**cm.WRDFRQmc**	1.52
liwc.adj	1.23	**cm.SMCAUSlsa**	1.48	**cm.WRDMEAc**	1.48
cm.RDFRE	1.23	**cm.WRDIMGc**	1.45	**cm.LDTTRc**	1.48
liwc.quant	1.22	**cm.WRDAOAc**	1.44	cm.LSASSpd	1.44
cm.LSASS1	1.21	**cm.DESWLlt**	1.42	**cm.DESWLsy**	1.38
liwc.compare	1.17	**cm.LSASS1d**	1.41	**cm.DESWLlt**	1.38
cm.SYNSTRUTt	1.13	**cm.DESWLsy**	1.41	**cm.WRDAOAc**	1.38
cm.SYNMEDlem	1.10	**cm.LSASS1**	1.37	**cm.WRDIMGc**	1.35
cm.LSASS1d	1.06	cm.LSASSp	1.37	**cm.CRFCWO1**	1.35
liwc.Dic	1.05	**cm.DESSLd**	1.34	**cm.CRFCWOa**	1.34

Tabela 13 – Vinte *features* mais importantes para classificação de FP pelo MDG.

No quinto cenário apresentado na Tabela 16, analisamos o nível de *feedback* FP para o texto em Português e suas respectivas traduções. No idioma Português, contabiliza-se todas as 20 principais características da ferramenta Coh-Metrix. Na tradução para o idioma Inglês usando a API do Google Translate, contabilizamos 12 características para a ferramenta Coh-Metrix e

Dataset English - FS					
EN_FS		**EN2PT_Google_FS**		**EN2PT Azure_FS**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
liwc.social	4.93	**liwc.friend**	5.29	**liwc.friend**	4.92
liwc.friend	3.23	**cm.DESWLltd**	2.39	**cm.DESWLltd**	2.41
cm.LSASS1d	2.12	**cm.WRDFAMc**	1.99	liwc.cause	2.30
liwc.prep	1.80	**cm.SMCAUSlsa**	1.99	**cm.WRDMEAc**	2.22
liwc.motion	1.75	**cm.WRDMEAc**	1.95	**cm.WRDFRQmc**	2.10
cm.LSAGN	1.56	**cm.WRDFRQmc**	1.79	**cm.WRDFAMc**	1.96
liwc.verb	1.44	**cm.DESWLsyd**	1.78	**cm.WRDCNCc**	1.79
cm.DESWLsyd	1.31	**liwc.ipron**	1.76	**cm.SMCAUSlsa**	1.66
liwc.Clout	1.27	liwc.anger	1.71	**cm.DESWLlt**	1.61
cm.WRDIMGc	1.27	**cm.DESSLd**	1.60	**cm.DESWLsyd**	1.61
cm.WRDFAMc	1.24	**cm.CRFCWO1d**	1.51	cm.LSASSp	1.50
cm.WRDCNCc	1.23	**cm.RDL2**	1.51	**cm.LDMTLDa**	1.47
cm.LSASS1	1.16	cm.LDTTRc	1.45	**cm.DESSLd**	1.46
cm.DESSLd	1.16	**cm.LDMTLDa**	1.39	**cm.SYNMEDpos**	1.43
cm.LSAGNd	1.07	**cm.WRDIMGc**	1.39	**cm.RDL2**	1.41
liwc.cogproc	1.07	cm.DESWLsy	1.38	**cm.WRDAOAc**	1.39
cm.RDFKGL	1.02	**cm.WRDAOAc**	1.38	cm.SYNLE	1.39
cm.SYNMEDlem	1.00	**cm.SYNMEDpos**	1.35	**cm.CRFCWO1d**	1.32
liwc.Dic	1.00	**cm.WRDCNCc**	1.27	cm.CRFCWOa	1.28
liwc.ipron	0.95	**cm.DESWLlt**	1.24	**cm.LSAGN**	1.27

Tabela 14 – Vinte *features* mais importantes para classificação de FS pelo MDG.

8 características para a ferramenta LIWC. Já na tradução para o idioma Português usando a API do Microsoft Azure, contabilizamos 9 características para a ferramenta Coh-Metrix e 11 características para a ferramenta LIWC.

Ao todo são 41 características da ferramenta Coh-Metrix somando 114.30 de MDG

Dataset Português - FT					
PT_FT		**PT2EN Google_FT**		**PT2EN Azure_FT**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
cm.DESWC	6.72	**liwc.WC**	8.29	**liwc.WC**	8.33
cm.WRDFRQmc	5.24	liwc.money	2.55	**cm.DESSLd**	3.02
cm.RDFKGL	4.29	liwc.verb	2.52	**cm.DESWLsyd**	2.96
cm.DESSLd	3.97	**cm.DESWC**	2.26	**liwc.social**	2.31
cm.DESSL	3.89	**liwc.social**	2.25	**cm.DESWLlt**	2.24
cm.DESWLltd	3.58	**cm.DESWLlt**	2.22	**cm.DESWLltd**	2.24
cm.WRDFAMc	3.42	**cm.DESWLltd**	2.21	**cm.CNCTemp**	2.19
cm.RDFRE	3.34	**cm.DESSLd**	2.20	**liwc.WPS**	2.12
cm.WRDIMGc	2.88	**cm.CNCTemp**	2.08	**cm.DESWC**	2.11
cm.WRDMEAc	2.37	**cm.DESWLsyd**	2.06	**liwc.AllPunc**	1.91
cm.SYNLE	2.09	**cm.RDL2**	2.01	liwc.Parenth	1.87
cm.WRDFRQc	2.00	liwc.relativ	1.84	**liwc.Colon**	1.82
cm.WRDFRQa	1.99	**cm.RDFKGL**	1.79	liwc.focuspast	1.64
cm.DESWLsy	1.59	**liwc.Colon**	1.77	liwc.Exclam	1.57
cm.WRDCNCc	1.56	liwc.function	1.49	cm.WRDPRO	1.48
cm.DESWLlt	1.51	**cm.WRDIMGc**	1.43	liwc.Dic	1.29
cm.WRDAOAc	1.46	**liwc.WPS**	1.36	**cm.WRDCNCc**	1.27
cm.LSAGN	1.42	liwc.achieve	1.34	**cm.WRDFAMc**	1.27
cm.DESWLsyd	1.42	liwc.cause	1.21	**cm.SYNLE**	1.26
cm.WRDNOUN	1.35	**liwc.AllPunc**	1.17	**cm.RDL2**	1.26

Tabela 15 – Vinte *features* mais importantes para classificação de FT pelo MDG.

e 19 características para a ferramenta LIWC somando 27.03 de MDG. Vale ressaltar que as características ”cm.DESWC”, ”cm.DESWLsyd”, ”cm.RDFRE”, ”cm.WRDMEAc”estão presentes nas três classificações de FP.

No sexto cenário apresentado na Tabela 17, analisamos o nível de *feedback* FS para o texto

Dataset Português - FP					
PT_FP		**PT2EN Google_FP**		**PT2EN Azure_FP**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
cm.LSASS1	9.93	**cm.DESWC**	9.24	**cm.DESWC**	11.18
cm.DESWC	7.77	**cm.SYNSTRUTt**	5.02	**cm.WRDPRO**	4.59
cm.WRDVERB	7.60	cm.DRPVAL	4.95	**cm.WRDMEAc**	3.27
cm.DESSL	2.66	**cm.WRDMEAc**	2.34	**cm.SYNSTRUTt**	2.93
cm.RDFKGL	2.14	**liwc.Exclam**	2.31	**liwc.Exclam**	2.33
cm.DESWLsy	2.01	**liwc.Comma**	2.03	**liwc.Comma**	2.05
cm.DESWLltd	2.00	**cm.DESSL**	1.89	**liwc.cogproc**	1.81
cm.WRDFAMc	1.96	**cm.RDFRE**	1.70	**cm.RDFRE**	1.55
cm.DESWLsyd	1.61	**liwc.cogproc**	1.60	**liwc.Colon**	1.50
cm.LSAGN	1.58	**cm.WRDPRO**	1.32	**cm.RDL2**	1.38
cm.SMCAUSlsa	1.58	**cm.LDMTLDa**	1.31	liwc.Clout	1.32
cm.WRDIMGc	1.54	**cm.RDL2**	1.22	liwc.Dic	1.31
cm.DESWLlt	1.51	**cm.DESWLsyd**	1.17	**cm.LSAGNd**	1.24
cm.LSAGNd	1.50	liwc.discrep	1.15	**cm.DESWLsyd**	1.23
cm.RDFRE	1.44	cm.DRNEG	1.13	liwc.Period	1.16
cm.WRDFRQmc	1.38	**liwc.Colon**	1.11	**cm.LDMTLDa**	1.06
cm.WRDCNCc	1.37	cm.CNCADC	1.10	liwc.WC	1.04
cm.WRDAOAc	1.34	**liwc.focuspresent**	1.09	**liwc.focuspresent**	1.02
cm.WRDMEAc	1.30	liwc.drives	1.09	liwc.function	1.02
cm.LSASSp	1.26	liwc.Authentic	1.08	liwc.auxverb	1.01

Tabela 16 – Vinte *features* mais importantes para classificação de FP pelo MDG.

em Português e suas respectivas traduções. No idioma Português, contabiliza-se 16 características da ferramenta Coh-Metrix e 4 características para a ferramenta LIWC. Na tradução para o idioma Inglês usando a API do Google Translate, contabilizamos 3 características para a ferramenta Coh-Metrix e 17 características para a ferramenta LIWC. Já na tradução para o idioma Português

usando a API do Microsoft Azure, contabilizamos 4 características para a ferramenta Coh-Metrix e 16 características para a ferramenta LIWC.

Ao todo são 23 características da ferramenta Coh-Metrix somando 43.13 de MDG e 37 características para a ferramenta LIWC somando 144.10 de MDG. Vale ressaltar que as características ”cm.WRDMEAc”, ”liwc.posemo”e ”liwc.relativ”estão presentes nas três classificações de FS.

No sétimo cenário apresentado na Tabela 18, analisamos a presença cognitiva para o texto em Inglês e suas respectivas traduções. No idioma Inglês, contabiliza-se 4 características da ferramenta Coh-Metrix, 9 características para a ferramenta LIWC, 3 características para a ferramenta SNA e 4 características para a ferramenta DCF. Na tradução para o idioma Português usando a API do Google Translate, contabiliza-se 7 características da ferramenta Coh-Metrix, 8 características para a ferramenta LIWC, 2 características para a ferramenta SNA e 3 características para a ferramenta DCF. Já na tradução para o idioma Português usando a API do Microsoft Azure,contabiliza-se 11 características da ferramenta Coh-Metrix, 4 características para a ferramenta LIWC, 2 características para a ferramenta SNA e 3 características para a ferramenta DCF.

Ao todo são 22 características da ferramenta Coh-Metrix somando 36.26 de MDG, 21 características para a ferramenta LIWC somando 31.23 de MDG, 7 características da ferramenta SNA somando 11.87 de MDG e 10 características para a ferramenta DCF somando 14.71 de MDG. Vale ressaltar que as características ”cm.DESWC”, ”dcf.depth”, ”dcf.sim.next”, ”dcf.sim.prev”, ”sna.betweenness”e ”sna.closeness”estão presentes nas três classificações de FT.

No oitavo e último cenário apresentado na Tabela 19, analisamos a presença cognitiva para o texto em Português e suas respectivas traduções. No idioma Português, contabiliza-se 10 características da ferramenta Coh-Metrix, 9 características para a ferramenta LIWC, 1 característica para a ferramenta SNA e 0 características para a ferramenta DCF. Na tradução para o idioma Inglês usando a API do Google Translate, contabiliza-se 8 características da ferramenta Coh-Metrix, 11 características para a ferramenta LIWC, 1 característica para a ferramenta SNA e 0 características para a ferramenta DCF. Já na tradução para o idioma Português usando a API do Microsoft Azure,contabiliza-se 17 características da ferramenta Coh-Metrix, 12 características para a ferramenta LIWC, 1 característica para a ferramenta SNA e 0 características para a ferramenta DCF.

Ao todo, são 25 características da ferramenta Coh-Metrix somando 48.70 de MDG, 32

Dataset Português - FS					
PT_FS		**PT2EN_Google_FS**		**PT2EN Azure_FS**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
liwc.leisure	9.97	**liwc.Exclam**	38.27	**liwc.Exclam**	37.77
cm.RDFKGL	4.01	**liwc.nonflu**	4.98	**liwc.posemo**	5.35
cm.DESSL	3.53	**liwc.posemo**	4.66	**liwc.nonflu**	4.45
cm.RDFRE	3.45	**liwc.affect**	3.48	**liwc.affect**	1.83
liwc.relativ	3.39	liwc.focuspresent	1.61	**liwc.motion**	1.80
cm.DESWLltd	2.68	**liwc.pronoun**	1.49	**liwc.Tone**	1.77
cm.LSASSp	2.59	**liwc.relativ**	1.25	**liwc.relativ**	1.30
cm.WRDMEAc	2.53	**liwc.reward**	1.18	**liwc.pronoun**	1.16
cm.LSAGN	2.38	liwc.Period	1.15	**liwc.achieve**	1.06
cm.WRDADJ	2.36	**liwc.Tone**	1.08	**liwc.reward**	1.05
liwc.posemo	2.33	**liwc.achieve**	1.00	**cm.LDTTRc**	0.97
cm.WRDFRQmc	2.13	liwc.ipron	0.87	cm.SYNMEDwrd	0.93
liwc.space	1.99	**cm.WRDMEAc**	0.76	liwc.Authentic	0.86
cm.SYNLE	1.97	**cm.LDTTRc**	0.75	**cm.WRDMEAc**	0.85
cm.DESWLsy	1.89	**liwc.work**	0.74	**cm.WRDIMGc**	0.78
cm.WRDIMGc	1.73	liwc.article	0.73	**liwc.work**	0.76
cm.DESWLsyd	1.67	**liwc.motion**	0.69	liwc.we	0.70
cm.WRDAOAc	1.55	cm.RDL2	0.66	**liwc.negate**	0.70
cm.DESWLlt	1.51	liwc.health	0.65	liwc.auxverb	0.70
cm.WRDFAMc	1.45	**liwc.negate**	0.64	liwc.Comma	0.69

Tabela 17 – Vinte *features* mais importantes para classificação de FS pelo MDG.

características para a ferramenta LIWC somando 110.44 de MDG, 3 características da ferramenta SNA somando 11.60 de MDG e 0 características para a ferramenta DCF. Vale ressaltar que as características ”cm.DESSL”, ”cm.DESWC”, ”liwc.QMark”e ”sna.closeness”estão presentes nas três classificações de FT.

Dataset English - Presença cognitiva					
EN		**EN2PT Google**		**EN2PT Azure**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
cm.DESWC	5.69	**liwc.QMark**	3.04	liwc.cogmech	4.92
liwc.QMark	2.81	**sna.betweenness**	2.06	**sna.betweenness**	2.41
dcf.depth	2.02	**liwc.funct**	1.98	**dcf.depth**	2.30
cm.LDVOCD	1.75	**sna.closeness**	1.84	**cm.SMCAUSr**	2.22
cm.DESSL	1.49	**liwc.space**	1.78	**cm.WRDNOUN**	2.10
dcf.first	1.46	**cm.DESWC**	1.77	**sna.closeness**	1.96
sna.closeness	1.21	**dcf.depth**	1.68	cm.WRDAOAc	1.79
sna.betweenness	1.21	liwc.incl	1.36	**cm.DESWLsyd**	1.66
sna.degree	1.18	**dcf.sim.next**	1.33	**dcf.sim.prev**	1.61
liwc.i	1.06	**cm.SMCAUSr**	1.32	**cm.DESWC**	1.61
liwc.you	0.90	**dcf.sim.prev**	1.31	**dcf.sim.next**	1.50
liwc.AllPunc	0.83	cm.DESSLd	1.27	**liwc.space**	1.47
liwc.focuspast	0.83	**cm.SYNLE**	1.23	**cm.DESSL**	1.46
liwc.cause	0.79	liwc.conj	1.21	**cm.SYNLE**	1.43
dcf.sim.next	0.78	**cm.WRDNOUN**	1.20	**liwc.social**	1.41
cm.SMCAUSlsa	0.77	**cm.DESWLsyd**	1.14	cm.DESWLltd	1.39
liwc.posemo	0.77	liwc.preps	1.12	**cm.DESWLsy**	1.39
liwc.compare	0.75	**liwc.social**	1.08	**liwc.funct**	1.32
liwc.ppron	0.74	liwc.ipron	1.06	cm.WRDFAMc	1.28
dcf.sim.prev	0.72	**cm.DESWLsy**	1.03	cm.LDTTRa	1.27

Tabela 18 – Vinte *features* mais importantes para classificação de presença cognitiva pelo MDG.

Recapitulando o contexto envolvido, que considerou 8 cenários diferentes, nos quais são obtidas as 20 características que têm mais impacto para a classificação do texto no idioma original e suas duas traduções, para presença cognitiva e 3 níveis de *feedback*. Com isso, os cenários apontam uma soma total de 1033.09 de MDG distribuídos entre 480 características.

Dataset Português - Presença Cognitiva					
PT		**PT2EN Google**		**PT2EN Azure**	
Características	**MDG**	**Características**	**MDG**	**Características**	**MDG**
liwc.funct	5.88	**liwc.QMark**	12.27	**liwc.WPS**	20.25
liwc.QMark	4.71	**liwc.WPS**	8.32	**liwc.QMark**	17.72
sna.closeness	4.69	**cm.DESWC**	5.24	**liwc.you**	5.42
cm.WRDNOUN	3.32	**cm.WRDPRO**	4.58	**cm.DESWC**	5.07
cm.DESWC	3.05	**liwc.you**	3.68	**sna.closeness**	4.03
liwc.preps	2.97	liwc.WC	2.89	**cm.WRDPRO**	3.22
liwc.relativ	2.35	**sna.closeness**	2.88	**liwc.assent**	1.16
liwc.incl	2.29	**cm.DRPVAL**	2.78	**liwc.Period**	0.98
liwc.social	1.74	liwc.interrog	2.24	**cm.WRDPRP1p**	0.85
cm.WRDPRP2	1.64	**cm.LDTTRa**	1.78	liwc.certain	0.77
cm.DESSL	1.60	liwc.Exclam	1.56	liwc.cause	0.75
cm.RDFKGL	1.56	**cm.WRDPRP1p**	1.49	**cm.LDTTRa**	0.73
liwc.pronoun	1.51	**liwc.assent**	1.23	liwc.posemo	0.71
liwc.article	1.50	**cm.DESSL**	1.16	**cm.DRPVAL**	0.67
cm.DESWLsyd	1.47	**liwc.Period**	1.09	liwc.number	0.65
cm.DESWLltd	1.45	liwc.Comma	1.01	**liwc.pronoun**	0.62
cm.DESWLsy	1.40	liwc.AllPunc	0.99	**liwc.ipron**	0.60
cm.RDFRE	1.35	cm.WRDPRP1s	0.96	**cm.DESSL**	0.57
cm.LDMTLDa	1.30	**cm.LDMTLDa**	0.92	liwc.informal	0.55
liwc.ipron	1.29	liwc.conj	0.75	**cm.DESWLsyd**	0.54

Tabela 19 – Vinte *features* mais importantes para classificação de presença cognitiva pelo MDG.

Pela Tabela 20, vemos que tanto em quantidade de características quanto no somatório do MDG a ferramenta Coh-Metrix teve mais que o triplo da ferramenta LIWC, para os níveis de *feedback* FT e FP. Para o nível de *feedback* FS a ferramenta LIWC teve maior somatório de MDG e o Coh-Metrix teve o maior número de ocorrências. Quanto à presença cognitiva,

EN 2 PT	**FT**		**FP**		**FS**		**Cognitivo**	
	Qtd	**Soma**	**Qtd**	**Soma**	**Qtd**	**Soma**	**Qtd**	**Soma**
Coh-Metrix	51	86.63	51	81.09	46	70.78	22	36.26
Liwc	9	12.21	9	11.29	14	33.42	21	31.23
SNA	0	0.00	0	0.00	0	0.00	7	11.87
DCF	0	0.00	0	0.00	0	0.00	10	14.71
Total	**60**	**98.84**	**60**	**92.38**	**60**	**104.20**	**60**	**94.07**
PT 2 EN	**FT**		**FP**		**FS**		**Cognitivo**	
Coh-Metrix	40	95.65	41	114.30	23	43.13	25	48.70
Liwc	20	48.65	19	27.03	37	144.10	32	110.44
SNA	0	0.00	0	0.00	0	0.00	3	11.60
DCF	0	0.00	0	0.00	0	0.00	0	0.00
Total	**60**	**144.30**	**60**	**141.33**	**60**	**187.23**	**60**	**170.74**
Geral	**FT**		**FP**		**FS**		**Cognitivo**	
Coh-Metrix	91	182.28	92	195.39	69	113.91	47	84.96
Liwc	29	60.86	28	38.32	51	177.52	53	141.67
SNA	0	0	0	0	0	0	10	23.47
DCF	0	0	0	0	0	0	10	14.71

Tabela 20 – Ocorrência de características e somatório de MDG por ferramentas para diferentes traduções.

destaca-se a ferramenta LIWCpor ter maior quantidade de características e somatório de MDG. Também repara-se na superioridade em todos os cenários do somatório de MDG das bases que foram traduzidas para o português.

Foram encontradas características 10 categorias da ferramenta Coh-Metrix, todas as 7 categorias da ferramenta LIWC, os 3 tipos de características de SNA e os 4 tipos de características de DCF.

Ferramenta	**Categoria**	**Soma**	**Qtd**
Coh-Metrix	Descritivo	189.86	84
Liwc	Preocupações pessoais	180.56	41
Coh-Metrix	Informação da Palavra	159.91	83
Liwc	Processos Psicológicos	74.29	36
Coh-Metrix	LSA	66.88	36
Liwc	Dimensões Linguísticas	56.39	38
Liwc	Variáveis de linguagem de resumo	47.19	16
Coh-Metrix	Legibilidade	46.20	25
Coh-Metrix	Complexidade sintática	42.63	24
Liwc	Orientações de tempo	28.17	18
SNA	SNA	23.47	10
Coh-Metrix	Diversidade Lexical	23.05	18
Liwc	Contagem de Plavras	20.55	4
Coh-Metrix	Coesão Referencial	17.63	12
Coh-Metrix	Modelo de Situação	15.48	10
DCF	DCF	14.71	10
Liwc	Outra Gramática	11.22	8
Coh-Metrix	Densidade de padrão sintático	9.53	4
Coh-Metrix	Conectivos	5.37	3
Coh-Metrix	**Total**	**576.54**	**299**
Liwc	**Total**	**418.37**	**161**
SNA	**Total**	**23.47**	**10**
DCF	**Total**	**14.71**	**10**

Tabela 21 – Soma de MDG e quantidade de caracerísticas por ferramenta e categorias.

Em relação às categorias, foram organizadas e ordenadas pela soma de MDG e quantidade de ocorrências dos 8 cenários na Tabela 21, deparou-se com a disparidade de três categorias que somam mais que todas as outras categorias juntas. Tratam-se das:

1. A Categoria Descritiva da ferramenta Coh-Metrix que soma 189.86 (18.38%)de MDG em 84 (17.5%) ocorrências;
2. A Categoria Preocupações Pessoais da ferramenta LIWC que soma 180.56 (17.48%)de MDG em 41 (8.54%) ocorrências;
3. A Categoria Informação da Palavra da ferramenta Coh-Metrix que soma 159.91 (15.48%)de MDG em 83 (17.29%) ocorrências;

5.1.3 Questão de Pesquisa 3

Assim como realizado nos resultados da Seção 5.1.1, para comparar o texto original com suas traduções, foram realizados testes de hipótese *Mann-Whitney* para verificar se os resultados quanto à acurácia e *Kappa* são estatisticamente iguais ou diferentes. Então foi elaborada a tabela 22 sintetizando os resultados médios e testes estatísticos que ajudam a melhor comparar e compreender as informações explicadas a seguir.

	Acurácia						**Kappa**					
	EN2PT Google vs EN2PT Azure			**PT2EN Google vs PT2ENAzure**			**EN2PT Google vs EN2PT Azure**			**PT2EN Google vs PT2ENAzure**		
	Google	**p**	**Azure**	**Google**	**p**	**Azure**	**Google**	**p**	**Azure**	**Google**	**p**	**Azure**
FT	**0.711**	0.017	0.708	**0.903**	**0.912**	**0.903**	**0.037**	**0.438**	**0.035**	**0.340**	**0.751**	**0.342**
FP	0.648	0	**0.655**	**0.783**	0.044	0.779	0.095	0	**0.116**	**0.566**	0.044	0.558
FS	0.801	0.001	**0.804**	**0.922**	**0.711**	**0.923**	0.165	0.014	**0.176**	**0.687**	**0.631**	**0.683**
Cognitivo	**0.575**	0	0.564	0.83	0.032	**0.833**	**0.381**	0	0.370	0.697	0	**0.712**

Tabela 22 – Médias e estatísticas para acurácia e kappa.

1. Quando leva-se em conta a acurácia para os textos traduzidos do Inglês para o Português, todos os resultados apresentaram diferença significativa entre as amostras. Para classificação de FT e presença cognitiva a tradução do Google teve resultados médios maiores. Já para classificação de FS e FP, o Azure teve os melhores resultados.
2. Quando leva-se em conta a acurácia para os textos traduzidos do Português para o Inglês, há igualdade estatística para as classificações de FT e FS. Para a classificação de FP o Google teve maior resultado com significância estatística, e o Azure teve vantagem na classificação de presença cognitiva.
3. Quando leva-se em conta o valor de *Kappa* para os textos traduzidos do Inglês para o

Português, a classificação de FT teve resultados estatisticamente iguais. O tradutor Azure teve resultados estatisticamente maiores na classificação de FP e FS, e o Google teve maior resultado na classificação de presença cognitiva.

4. Quando leva-se em conta o valor de *Kappa* para os textos traduzidos do Português para o Inglês, há igualdade estatística para as classificações de FT e FS. Para a classificação de FP o Google teve maior resultado com significância estatística, e o Azure teve vantagem na classificação de presença cognitiva.

5.2 Discussões

Esta Seção apresenta as discussões sobre os resultados obtidos na Seção 5.1 referentes à aplicação do método proposto para automatizar a análise da presença cognitiva e níveis de *feedback* educacional em relação às perguntas de pesquisa.

5.2.1 PP1: Eficácia da tradução automática de texto

Dos 16 cenários envolvendo acurácia com diferença de significância estatística comprovadas, 6 cenários apontam diminuição média dos resultados da tradução do Inglês para o Português; 6 cenários apontam aumento médio dos resultados com a tradução do Português para o Inglês e em apenas um resultado a tradução do Inglês pra o Português do Google apontou aumento do resultado médio.

Dos 16 cenários envolvendo *kappa* com diferença de significância estatística comprovadas, 7 cenários apontam para aumento médio dos resultados com a tradução do Português para o Inglês, 3 cenários apontam diminuição dos resultados médios com a tradução do Português para o Inglês e 2 resultados envolvendo classificação de FT da tradução do Inglês para o Português, tiveram aumento médio dos resultados.

Assim, a pesquisa não chegou a uma conclusão absoluta, quanto a tradução do Inglês para o Português, mas 9 dos 16 cenários (56,25%), apontaram uma diminuição de significância estatística dos resultados médios. Já para a tradução do Português para o Inglês, 14 dos 16 cenários (87,5%), tiveram aumento de significância estatística dos resultados médios, e os cenários restantes também apresentaram aumento mas sem diferença estatística.

Com isso nota-se a importância do texto analisado em questão estar no idioma Inglês pois

as ferramentas para extração de características para esse idioma que é o mais falado e utilizado academicamente para esses tipos de pesquisa, além de contar com mais tempo de criação e aprimoramento por parte dos desenvolvedores e comunidade.

Também vale ressaltar que o cenário "PT2EN Azure"obteve os melhores resultados para acurácia (0,833) e *kappa* (0,712) para classificação de presença cognitiva. Para classificação de níveis de *feedback* obtivemos 0.923 e 0.687 como os melhores resultados para acurácia e *kappa* respectivamente. Embora não se possa comparar diretamente aos resultados dos trabalhos relacionados nas Tabelas 4 e 5, por utilizarem algumas metodologias diferentes, os resultados deste trabalho indicam relevância das abordagens utilizadas, por se aproximarem dos resultados apresentados e no caso da acurácia apresentando até valores maiores.

5.2.2 PP2: Análise do impacto das características

A questão de pesquisa 2 focou na interpretação da importância do recurso para os diferentes modelos gerados. Para atingir esse objetivo, este estudo replicou a mesma metodologia proposta por KOVANOVIĆ et al. (2016) e BARBOSA et al. (2021), que adotaram MDG para estimar os 20 principais recursos para cada classificador criado.

Separando apenas por ferramentas de extração de características, compreende-se a ausência das ferramentas de SNA e DCF na classificação de textos para níveis de *feedback* educacional, às deixando de fora de 6 entre 8 cenários propostos. As 20 ocorrências de características das duas ferramentas, representam e 16,66% dos cenários de presença cognitiva. Já em relação ao MDG as características de SNA somam 23.47, representando 2.27% do todo e 8.86% do cenários de presença cognitiva, enquanto as as características de DCF somam 14.71, representando 1.42% do todo e 5.55% dos cenários de presença cognitiva. é importante notar que as duas ferramentas representam cerca de 4% do total das características mas impactam em 14,41% do MDG para classificação de presença cognitiva.

As ferramentas de extração de características que lideram as medidas foram : LIWC com 161 (33.54%) ocorrências totais com uma soma para MDG de 418.37 (40.5%); e Coh-Metrix com 299 (62.29%) ocorrências, quase o dobro da outra ferramenta; e com soma para MDG de 576.54 (55.81%).

Sobre as categorias das ferramentas de extração de características, há uma disparidade de três categorias que somam mais que todas as outras juntas. A Categoria Descritiva da

ferramenta Coh-Metrix que soma 189.86 (18.38%)de MDG em 84 (17.5%) ocorrências; A Categoria Preocupações Pessoais da ferramenta LIWC que soma 180.56 (17.48%)de MDG em 41 (8.54%) ocorrências; A Categoria Informação da Palavra da ferramenta Coh-Metrix que soma 159.91 (15.48%) de MDG em 83 (17.29%) ocorrências. Vale ressaltar que nenhuma das 11 características da categoria para Conectivos da língua Portuguesa da ferramenta Coh-Metrix tiveram ocorrência entre as 20 características que têm mais impacto da classificação quanto ao MDG. Como o Coh-Metrix já tem a categoria dos conectivos, não se fez necessário as variações para o idioma português, e além disso, a categoria teve o menor número de ocorrências (3) de todas as classificações.

Vale ressaltar que dos 8 cenários e 480 características envolvidas, houve uma variação entre 117 características devido à repetição de características em mais de uma classificação. Dessas 177 às 20 que mais apareceram foram todas da ferramenta Coh-Metrix contabilizando 208 ocorrências. Em relação às características que obtiveram maior soma para MDG, desacam-se:”liwc.Exclam”(83.81), ”cm.DESWC”(61.71), ”liwc.QMark”(40.55), ”liwc.WPS”(32.05),”cm.DESWLsyd”(29.88), ”cm.DESWLltd”(28.17), ”cm.WRDMEAc”(24.52), ”cm.WRDFAMc”(22.41), ”cm.DESSLd”(22.32), ”liwc.WC”(20.55) e ”cm.WRDFRQmc”(20.27). Das características de SNA na classificação de presença cognitiva o ”sna.closeness”teve os maior soma e MDG (16.61) e número de ocorrências (6). Já para as características de DCF na classificação de presença cognitiva o ”dcf.depth”teve os maior soma e MDG (6.00) e número de ocorrências (3).

Quanto a classificação de *feedback* educacional, um trabalho não focou na importância das características (CAVALCANTI et al., 2020a), e os outros dois trabalhos verificaram essa importância, mas trabalhavam com número reduzido de características (105) em relação à presente pesquisa (176/188), o que resultou em um número maior de MDG para cada característica, pois a proporção unitária é bem maior. O artigo de CAVALCANTI et al. (2019) além de trabalhar com com outro algoritmo, o XGBOOST, e realiza em uma única classificação às 7 boas práticas de *feedback*. Já o artigo de CAVALCANTI et al. trabalha com o algoritmo de classificação *Random Forest* para FT, FP e FS, mas apenas para o idioma Português e 105 características , mesmo assim percebe-se a incidência de algumas características mais relevantes como no presente estudo: ”cm.DESPC”, ”cm.DESSC”, ”cm.DESWC”e ”liwc.WPS”.

Quanto à classificação de presença cognitiva, mostra-se com um maior grau de similaridade com os trabalhos anteriormente relatados na literatura (KOVANOVIć et al., 2016;

NETO et al., 2018; FARROW et al., 2019; BARBOSA et al., 2020), com destaque para as características sobre o número de palavras ”cm.DESWC”, número de pontos de interrogação ”liwc.QMark”, condizente com as categorias Descritiva e Preocupações pessoais, com maior soma de MDG do presente trabalho.

5.2.3 PP3: Divergência entre tradutores

De acordo com os resultados médios de acurácia e *Kappa* demonstrados na Seção 5.1.3, não é possível afirmar que um tradutor obteve resultado melhor que outro. Dos 16 cenários comparativos, 5 apresentam resultados estatisticamente similares. Já nos cenários com diferença de significância estatística, o tradutor do Google Translate leva vantagem em 5 cenários e o Microsoft Azure em 6 cenários.

Dos trabalhos relacionados no Capítulo 3, a maioria faz uso de apenas um idioma. A pesquisa de BARBOSA et al. (2020) trabalhou com duas bases diferentes, mas não envolveu tradução. Já a pesquisa de BARBOSA et al. (2021), trabalhou com tradução de textos, mas utilizou apenas um tradutor.

Mesmo assim, vale ressaltar o desempenho dos tradutores, que disponibilizam gratuitamente e entregam resultados parecidos, e caso um pesquisador tenha as ferramentas de extração de características para apenas um idioma que seja diferente da base que ele esteja trabalhando, a tradução vai suprir sua necessidade para deixar seus textos aptos a serem utilizados nas suas ferramentas.

6 Considerações Finais

Esta dissertação de mestrado teve como objetivo principal analisar se o uso da tradução de textos para o idioma Inglês mantém ou melhora o desempenho dos classificadores de textos educacionais, além de comparar os resultados entre os tradutores e ordenar as caracteríticas mais relevantes para o processo de classificação.

Primeiramente, foram traduzidas as bases dos idiomas originais com as APIs do Google Translate e Micosoft Azure para diferentes idiomas, realizou-se a extração de características por diferentes ferramentas (Coh-Metrix, LIWC, SNA, DCF) com os resultados de acurácia e *kappa* envolvendo os textos originais duas traduções.

Depois dividiu-se os dados em conjuntos de treinamento e teste, adotando a técnica de CV estratificada em 10 vezes, e em cada caso o classificador foi executado 30 vezes a fim de criar diferentes amostras de resultados de acurácia e *kappa* e então aplicar testes estatísticos para verificar se as amostras apresentaram resultados significativamente diferentes. Na maioria dos cenários que houve uma diferença estatística significativa, notou-se a melhora de resultados quando se traduzia para o inglês ou percebeu-se uma diminuição nos resultados com a tradução para o portugês, com exceção de um caso.

Em seguida, ordenou-se as 20 principais características pelo MDG nos diversos cenários de classificação. Notou-se uma grande disparidade em relação a quantidade e somatório de MDG para os níveis de *feedback* educacionais FT e acsfp. Para o FS notou-se um número de ocorrências maior para o Coh-Metrix e um somatório maior para LIWC. Por fim, na presença cognitiva destaca-se com os maiores valores para quantidade e somatório de MDG para a ferramenta LIWC. Também percebeu-se que os cenários onde havia a tradução do inglês para o portuguêS, tanto para níveis de *feedback* quanto para presença cognitiva, a ocorrência de características e a soma de MDG foram superiores. No geral, levando em conta todos os cenários, a ferramenta Coh-Metrix lidera com maiores valores para para quantidade e somatório de MDG .

Por fim, quando analisou-se os resultado entre os tradutores, 8 dos 12 cenários tiveram resultados estatisticamente iguais, e os 4 cenários restantes com diferença significativa, 2 apontaram para vantagem do tradutor do Google e 2 cenários apontaram para o tradutor do Microsoft Azure. Assim considerando os resultados de tradução para as duas plataformas praticamente iguais.

As pesquisas propostas anteriormente na literatura trabalhavam apenas com um idioma

ou apenas um tipo de classificação, enquanto a presente pesquisa trabalhou com dois idiomas e tradução deles, além de contribuir com dois tipos da classificação, demonstrando vantagens do idioma Inglês nas ferramentas de extração de características e desempenho para classificação automatizada de textos educacionais referentes aos níveis *feedback* e presença cognitiva, mesmo que sejam traduzidos do português por APIs da Google ou Microsoft Azure, onde todos os cenários propostos indicaram aumento dos resultados e 87,5% deles com significância estatística pelo teste de *Mann-Whitney*. Também elencou-se as características pelo impacto que causam na classificação e identificando que apenas três categorias juntas somam 51,34% de MDG total classificado, são elas as cetegorias Descritivas e Informações da Palavra do Coh-Metrix e Preocupações Pessoais do LIWC. Também ressaltou que para a classificação de presença cognitiva, mesmo com cerca de 4% das características totais, as ferramentas de extração de características SNA e DCF representam 14,41% do total do MDG, indicando o impacto por característica significativo para a classificação.

Além disso, os resultados mostram o potencial de serem incorporados a um AVA que contenham atividades e fóruns de discussão, permitindo aos professores um melhor acompanhamento dos alunos durante o curso online e garantir iterações adequadas que otimizem a aprendizagem do conteúdo debatido e respondido.

6.1 Artigos aceitos

Para divulgação dos temas de pesquisa, foram submetidos e aceitos em conferências e periódicos qualificados na área de Ciência da Computação os seguintes artigos:

- BARBOSA, A.; FERREIRA, M.; MELLO, R. F.; LINS, R. D.; GASEVIC, D. Theimpact of automatic text translation on classification of online discussions for social andcognitive presences. In: LAK21: 11th International Learning Analytics andKnowledge Conference. New York, NY, USA: Association for Computing Machinery,2021. p. 77–87. ISBN 9781450389358. Disponível em: ⟨https://doi.org/10.1145/3448139.3448147⟩
- CAVALCANTI, A. P.; BARBOSA, A.; MELLO, R. F.; MANGAROSKA, K.; NASCIMENTO,A.; FREITAS, F.; GASEVIĆ, D. How good is my feedback? a content analysis of writtenfeedback. In:Proceedings of the Tenth International Conference on LearningAnalytics & amp; Knowledge. New York, NY, USA: Association for ComputingMachinery, 2020. (LAK '20), p. 428–437. ISBN 9781450377126. Disponível

em: ⟨https://doi.org/10.1145/3375462.3375477⟩.

6.2 Limitações da pesquisa

O presente estudo tem algumas limitações. A principal preocupação é o tamanho e os idiomas dos conjuntos de dados utilizados, que embora fossem de tamanho semelhante aos usados em trabalhos anteriores (KOVANOVIĆ et al., 2016; NETO et al., 2018; CAVALCANTI et al., 2020b; BARBOSA et al., 2021), eles ainda são relativamente pequenos, o que poderia influenciar os resultados pela falta de diversidade de cursos e áreas de cursos.

Por trabalhar com apenas 2 idiomas, inglês e português, com um número semelhante de recursos linguísticos e uma relativa proximidade linguística entre eles, não é possível afirmar o impacto da tradução para um idioma significativamente diferente (por exemplo, árabe ou chinês) que podem produzir resultados diferentes.

Por limitação de tempo e capacidade computacional, não foram utilizados outros algorítmos para classificação, como SVM e XGBOOST.

Por fim, o estudo adotou 2 tradutores de empresas renomadas no mercado, mas que são gratuitos, e poderia analisar se outros tradutores, incluindo os de ferramentas pagas, poderiam apresentar características extraídas com diferença significativa nos resultados de classificação automatizada.

6.3 Trabalhos futuros

Para trabalhos futuros, pretende-se abordar os seguintes pontos:

- Aplicar o método do presente estudo em bases maiores, que envolvam uma diversidade de cursos de divérsas áreas, incluindo as que não foram abordadas na pesquisa, como área da saúde.
- Incluir a classificação automatizada de outras formas que apontem uma melhor experiência de aprendizagem como presença social e de ensino das COI ou as 7 boas práticas de *feedback*.
- Devido limitação das ferramentas de extração de características, desenvolver um estudo para verificar a tradução de outros idiomas como árabe, mandarim ou espanhol, traduzi-los para o inglês e constatar se os resultados de classificação se aproximam com os resultados

encontrados na literatura.

- Incluir outras abordagens de extração de características para verificar o impacto na classificação, podendo incluir características próprias de outros idiomas. Por exemplo, as características extraídas para o idioma português diferentes do idioma inglês, poderiam ser acrescentadas às características extraidas para o idioma inglês, garantindo uma maior diversidade e combinações para serem analisadas pelo classificador.

Referências

ADZHARUDDIN, N. A.; LING, L. H. Learning management system (lms) among university students: Does it work. **International Journal of e-Education, e-Business, e-Management and e-Learning**, IACSIT Press, v. 3, n. 3, p. 248–252, 2013. Disponível em: ⟨https://doi.org/10.7763/IJEEEE.2013.V3.233⟩.

AMINI, M.-R.; USUNIER, N.; GOUTTE, C. et al. Learning from multiple partially observed views-an application to multilingual text categorization. In: **NIPS**. [S.l.: s.n.], 2009. v. 22, p. 28–36.

ANDERSON, T.; LIAM, R.; GARRISON, D. R.; ARCHER, W. Assessing teaching presence in a computer conferencing context. **Journal of Asynchronous Learning Networks**, Journal of the Asynchronous Learning Network, v. 5, n. 2, 2001. Disponível em: ⟨https://auspace.athabascau.ca/handle/2149/725⟩.

ANDERSON, T.; ROURKE, L.; GARRISON, D. R.; ARCHER, W. Assessing Teaching Presence in a Computer Conferencing Context. **Journal of Asynchronous Learning Networks**, v. 5, p. 1–17, 2001.

ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. **Revista Eletrônica de Sistemas de Informação**, v. 5, n. 2, 2006. Disponível em: ⟨https://doi.org/10.21529/RESI.2006.0502001⟩.

ARAUJO, E. M.; NETO, J. D. de O. Avaliação do pensamento crítico e da presença cognitiva em fórum de discussão online utilizando a análise estatística textual. In: **Proceedings of International Conference on Engineering and Computer Education**. [s.n.], 2013. v. 8, p. 113–117. Disponível em: ⟨https://copec.eu/congresses/icece2013/proc/works/26.pdf⟩.

ARBAUGH, J. B.; CLEVELAND-INNES, M.; DIAZ, S. R.; GARRISON, D. R.; ICE, P.; RICHARDSON, J. C.; SWAN, K. P. Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. **The internet and higher education**, Elsevier, v. 11, n. 3-4, p. 133–136, 2008. Disponível em: ⟨https://doi.org/10.1016/j.iheduc.2008.06.003⟩.

ARETIO, L. G. Educación a distancia hoy. **Madrid: UNED**, Universidad Nacional de Educación a Distancia (España), 1994.

BALZER, W. K.; DOHERTY, M. E. et al. Effects of cognitive feedback on performance. **Psychological bulletin**, American Psychological Association, v. 106, n. 3, p. 410–433, 1989. Disponível em: ⟨https://psycnet.apa.org/doi/10.1037/0033-2909.106.3.410⟩.

BANEA, C.; MIHALCEA, R.; WIEBE, J.; HASSAN, S. Multilingual subjectivity analysis using machine translation. In: **Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing**. [s.n.], 2008. p. 127–135. Disponível em: ⟨https://www.aclweb.org/anthology/D08-1014.pdf⟩.

BANGERT-DROWNS, R. L.; KULIK, J. A.; KULIK, C.-L. C. Effects of frequent classroom testing. **The Journal of Educational Research**, Routledge, v. 85, n. 2, p. 89–99, 1991. Disponível em: ⟨https://doi.org/10.1080/00220671.1991.10702818⟩.

BARBOSA, A.; FERREIRA, M.; MELLO, R. F.; LINS, R. D.; GASEVIC, D. The impact of automatic text translation on classification of online discussions for social and cognitive presences. In: ____. **LAK21: 11th International Learning Analytics and Knowledge Conference**. New York, NY, USA: Association for Computing Machinery, 2021. p. 77–87. ISBN 9781450389358. Disponível em: ⟨https://doi.org/10.1145/3448139.3448147⟩.

BARBOSA, G.; CAMELO, R.; CAVALCANTI, A. P.; MIRANDA, P.; MELLO, R. F.; KOVANOVIć, V.; GAsEVIć, D. Towards automatic cross-language classification of cognitive presence in online discussions. In: . New York, NY, USA: Association for Computing Machinery, 2020. ISBN 9781450377126. Disponível em: ⟨https://doi.org/10.1145/3375462.3375496⟩.

BLACK, P.; WILIAM, D. Assessment and classroom learning. **Assessment in Education: Principles, Policy & Practice**, Routledge, v. 5, n. 1, p. 7–74, 1998. Disponível em: ⟨https://doi.org/10.1080/0969595980050102⟩.

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001. Disponível em: ⟨https://doi.org/10.1023/A:1010933404324⟩.

BURKE, D. Strategies for using feedback students bring to higher education. **Assessment & Evaluation in Higher Education**, Routledge, v. 34, n. 1, p. 41–50, 2009. Disponível em: ⟨https://doi.org/10.1080/02602930801895711⟩.

BUTLER, D. L.; WINNE, P. H. Feedback and self-regulated learning: A theoretical synthesis. **Review of Educational Research**, v. 65, n. 3, p. 245–281, 1995. Disponível em: ⟨https://doi.org/10.3102/00346543065003245⟩.

CABRAL, L. d. S.; LINS, R. D.; MELLO, R. F.; FREITAS, F.; ÁVILA, B.; SIMSKE, S.; RISS, M. A platform for language independent summarization. In: **Proceedings of the 2014 ACM Symposium on Document Engineering**. New York, NY, USA: Association for Computing Machinery, 2014. (DocEng '14), p. 203–206. ISBN 9781450329491. Disponível em: ⟨https://doi.org/10.1145/2644866.2644890⟩.

CAMELO, R.; JUSTINO, S.; MELLO, R. Coh-metrix pt-br: Uma api web de análise textual para a educação. In: **Anais dos Workshops do IX Congresso Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2020. p. 179–186. ISSN 0000-0000. Disponível em: ⟨https://sol.sbc.org.br/index.php/wcbie/article/view/13043⟩.

CAVALCANTI, A.; MELLO, R.; MIRANDA, P.; FREITAS, F. Análise automática de feedback em ambientes de aprendizagem online. In: **Anais do XXXI Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2020. p. 892–901. ISSN 0000-0000. Disponível em: ⟨https://sol.sbc.org.br/index.php/sbie/article/view/12845⟩.

CAVALCANTI, A. P.; DIEGO, A.; MELLO, R. F.; MANGAROSKA, K.; NASCIMENTO, A.; FREITAS, F.; GASEVIĆ, D. How good is my feedback? a content analysis of written feedback. In: **Proceedings of the Tenth International Conference on Learning Analytics & amp; Knowledge**. New York, NY, USA: Association for Computing Machinery, 2020. (LAK '20), p. 428–437. ISBN 9781450377126. Disponível em: ⟨https://doi.org/10.1145/3375462.3375477⟩.

CAVALCANTI, A. P.; MELLO, R. Ferreira Leite de; ROLIM, V.; ANDRé, M.; FREITAS, F.; GAšEVIC, D. An analysis of the use of good feedback practices in online learning courses. In: **2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)**. [s.n.], 2019. v. 2161-377X, p. 153–157. Disponível em: ⟨https://doi.org/10.1109/ICALT.2019.00061⟩.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 785–794. ISBN 9781450342322. Disponível em: ⟨https://doi.org/10.1145/2939672.2939785⟩.

Cheniti Belcadhi, L. Personalized feedback for self assessment in lifelong learning environments based on semantic web. **Computers in Human Behavior**, v. 55, p. 562 – 570, 2016. ISSN 0747-5632. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0747563215300534⟩.

CHOWDHURY, G. G. Natural language processing. **Annual Review of Information Science and Technology**, v. 37, n. 1, p. 51–89, 2003. Disponível em: ⟨https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.1440370103⟩.

COATES, H.; JAMES, R.; BALDWIN, G. A critical examination of the effects of learning management systems on university teaching and learning. **Tertiary education and management**, Springer, v. 11, p. 19–36, 2005. Disponível em: ⟨https://doi.org/10.1007/s11233-004-3567-9⟩.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960. Disponível em: ⟨https://doi.org/10.1177/001316446002000104⟩.

CORICH, S.; HUNT, K.; HUNT, L. Computerised content analysis for measuring critical thinking within discussion forums. **Journal of e-learning and knowledge society**, Italian e-Learning Association, v. 2, n. 1, 2006. Disponível em: ⟨https://www.learntechlib.org/p/43459⟩.

CUNHA, A. L. V. d. **Coh-Metrix-Dementia: análise automática de distúrbios de linguagem nas demências utilizando Processamento de Línguas Naturais**. Tese (Doutorado) — Universidade de São Paulo, 2015.

CUTLER, R. H. Distributed presence and community in cyberspace. **Interpersonal Computing and Technology: An Electronic Journal for the 21st Century**, v. 3, n. 2, p. 12, 1995.

DAFT, R. L.; LENGEL, R. H. Organizational information requirements, media richness and structural design. **Management science**, INFORMS, v. 32, n. 5, p. 554–571, 1986. Disponível em: ⟨https://doi.org/10.1287/mnsc.32.5.554⟩.

DIAMANTIDIS, N.; KARLIS, D.; GIAKOUMAKIS, E. Unsupervised stratification of cross-validation for accuracy estimation. **Artificial Intelligence**, v. 116, n. 1, p. 1–16, 2000. ISSN 0004-3702. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S0004370299000946⟩.

EFRON, B. Bootstrap methods: Another look at the jackknife. In: ____. **Breakthroughs in Statistics: Methodology and Distribution**. New York, NY: Springer New York, 1992. p. 569–593. ISBN 978-1-4612-4380-9. Disponível em: ⟨https://doi.org/10.1007/978-1-4612-4380-9_41⟩.

EVANS, J. A.; ACEVES, P. Machine translation: Mining text for social theory. **Annual Review of Sociology**, Annual Reviews, v. 42, p. 21–50, 2016.

FABRO, K.; GARRISON, D. R. Computer conferencing and higher-order learning. **Indian journal of open learning**, v. 7, n. 1, p. 41–54, 1998.

FARROW, E.; MOORE, J.; GAsEVIć, D. Analysing discussion forum data: A replication study avoiding data contamination. In: **Proceedings of the 9th International Conference on Learning Analytics & Knowledge**. New York, NY, USA: Association for Computing Machinery, 2019. (LAK19), p. 170–179. ISBN 9781450362566. Disponível em: ⟨https://doi.org/10.1145/3303772.3303779⟩.

FELDMAN, R.; SANGER, J. **The text mining handbook: advanced approaches in analyzing unstructured data**. [S.l.]: Cambridge university press, 2007.

FERREIRA-MELLO, R.; ANDRÉ, M.; PINHEIRO, A.; COSTA, E.; ROMERO, C. Text mining in education. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, p. e1332, 2019. Disponível em: ⟨https://doi.org/10.1002/widm.1332⟩.

FERREIRA, R.; KOVANOVIĆ, V.; GAŠEVIĆ, D.; ROLIM, V. Towards combined network and text analytics of student discourse in online discussions. In: SPRINGER. **International Conference on Artificial Intelligence in Education**. 2018. p. 111–126. ISBN 978-3-319-93843-1. Disponível em: ⟨https://doi.org/10.1007/978-3-319-93843-1_9⟩.

FILHO, P. B.; PARDO, T. A. S.; ALUÍSIO, S. An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**. [S.l.: s.n.], 2013.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The elements of statistical learning**. [S.l.]: Springer series in statistics New York, NY, USA, 2001. v. 1.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 29, n. 5, p. 1189–1232, 2001. ISSN 00905364. Disponível em: ⟨http://www.jstor.org/stable/2699986⟩.

Fu, S.; Zhao, J.; Cui, W.; Qu, H. Visual analysis of mooc forums with iforum. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 1, p. 201–210, 2017. Disponível em: ⟨https://doi.org/10.1109/TVCG.2016.2598444⟩.

FU, Y. Combination of random forests and neural networks in social lending. **Journal of Financial Risk Management**, Scientific Research Publishing, v. 6, n. 4, p. 418–426, 2017. Disponível em: ⟨https://doi.org/10.4236/jfrm.2017.64030⟩.

GARRISON, D.; ANDERSON, T.; ARCHER, W. Critical inquiry in a text-based environment: Computer conferencing in higher education. **The Internet and Higher Education**, v. 2, n. 2, p. 87 – 105, 1999. ISSN 1096-7516. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S1096751600000166⟩.

GARRISON, D. R.; ANDERSON, T. E-learning in the 21st century: A framework for research and practice. **New York, Routledge**, 2003.

GARRISON, D. R.; ANDERSON, T.; ARCHER, W. Critical thinking, cognitive presence, and computer conferencing in distance education. **American Journal of distance education**, Taylor & Francis, v. 15, n. 1, p. 7–23, 2001. Disponível em: ⟨https://doi.org/10.1080/08923640109527071⟩.

____. The first decade of the community of inquiry framework: A retrospective. **The Internet and Higher Education**, v. 13, n. 1, p. 5–9, 2010. ISSN 1096-7516. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S1096751609000608⟩.

GARRISON, D. R.; ARBAUGH, J. Researching the community of inquiry framework: Review, issues, and future directions. **The Internet and Higher Education**, v. 10, n. 3, p. 157 – 172, 2007. ISSN 1096-7516. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S1096751607000358⟩.

GARRISON, D. R.; ARBAUGH, J. B. Researching the community of inquiry framework: Review, issues, and future directions. **The Internet and Higher Education**, Elsevier, v. 10, n. 3, p. 157–172, 2007. Disponível em: ⟨https://doi.org/10.1016/j.iheduc.2007.04.001⟩.

GAŠEVIĆ, D.; JOKSIMOVIĆ, S.; EAGAN, B. R.; SHAFFER, D. W. Sens: Network analytics to combine social and cognitive perspectives of collaborative learning. **Computers in Human Behavior**, Elsevier, 2018. ISSN 0747-5632. Disponível em: ⟨https://doi.org/10.1016/j.chb.2018.07.003⟩.

GAšEVIć, D.; ADESOPE, O.; JOKSIMOVIć, S.; KOVANOVIć, V. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. **The Internet and Higher Education**, v. 24, p. 53 – 65, 2015. ISSN 1096-7516. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S1096751614000700⟩.

GOLDEN, P. **Write here, write now**. 2011. Disponível em: ⟨https://www.research-live.com/article/features/write-here-write-now/id/4005303⟩.

GRAESSER, A. C.; MCNAMARA, D. S.; LOUWERSE, M. M.; CAI, Z. Coh-metrix: Analysis of text on cohesion and language. **Behavior research methods, instruments, & computers**, Springer, v. 36, n. 2, p. 193–202, 2004.

GUNAWARDENA, C. N.; ZITTLE, F. J. Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. **American Journal of Distance Education**, Routledge, v. 11, n. 3, p. 8–26, 1997. Disponível em: ⟨https://doi.org/10.1080/08923649709526970⟩.

HABERT, B.; ADDA, G.; ADDA-DECKER, M.; MARËUIL, P. B. de; FERRARI, S.; FERRET, O.; ILLOUZ, G.; PAROUBEK, P. Towards tokenization evaluation. In: **Proceedings of LREC**. [S.l.: s.n.], 1998. v. 98, p. 427–431.

HANSEN, L. K.; SALAMON, P. Neural network ensembles. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 12, n. 10, p. 993–1001, 1990.

HATTIE, J.; BIGGS, J.; PURDIE, N. Effects of learning skills interventions on student learning: A meta-analysis. **Review of Educational Research**, v. 66, n. 2, p. 99–136, 1996. Disponível em: ⟨https://doi.org/10.3102/00346543066002099⟩.

HATTIE, J.; TIMPERLEY, H. The power of feedback. **Review of Educational Research**, v. 77, n. 1, p. 81–112, 2007. Disponível em: ⟨https://doi.org/10.3102/003465430298487⟩.

HKIRI, E.; MALLAT, S.; ZRIGUI, M. Arabic-english text translation leveraging hybrid ner. In: **Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation**. [S.l.: s.n.], 2017. p. 124–131.

HUTCHINS, J. The development and use of machine translation systems and computer-based translation tools in europe, asia, and north america. Citeseer, 1998.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. Tree-based methods. In: ____. **An Introduction to Statistical Learning: with Applications in R**. New York, NY: Springer New York, 2013. p. 303–335. ISBN 978-1-4614-7138-7. Disponível em: ⟨https://doi.org/10.1007/978-1-4614-7138-7_8⟩.

JOACHIMS, T. Text categorization with support vector machines: Learning with many relevant features. In: SPRINGER. **European conference on machine learning**. [S.l.], 1998. p. 137–142.

KIM, D. S.; LEE, S. M.; PARK, J. S. Building lightweight intrusion detection system based on random forest. In: SPRINGER. **International Symposium on Neural Networks**. 2006. p. 224–230. ISBN 978-3-540-34483-4. Disponível em: ⟨https://doi.org/10.1007/11760191_33⟩.

KLEIJ, F. M. V. der; FESKENS, R. C. W.; EGGEN, T. J. H. M. Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. **Review of Educational Research**, v. 85, n. 4, p. 475–511, 2015. Disponível em: ⟨https://doi.org/10.3102/0034654314564881⟩.

KLUGER, A. N.; DENISI, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. **Psychological bulletin**, American Psychological Association, v. 119, n. 2, p. 254, 1996. Disponível em: ⟨https://pdfs.semanticscholar.org/97cc/e81ca813ed757e1e76c0023865c7dbdc7308.pdf⟩.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. **Ijcai**. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.

KORDE, V.; MAHENDER, C. N. Text classification and classifiers: A survey. **International Journal of Artificial Intelligence & Applications**, Academy & Industry Research Collaboration Center (AIRCC), v. 3, n. 2, p. 85, 2012.

KOVANOVIC, V.; JOKSIMOVIC, S.; GASEVIC, D.; HATALA, M. Automated cognitive presence detection in online discussion transcripts. In: **LAK Workshops**. [S.l.: s.n.], 2014.

____. What is the source of social capital? the association between social network position and social presence in communities of inquiry. In: EDM. **Workshop at Educational Data Mining Conference**. [S.l.], 2014.

KOVANOVIĆ, V.; JOKSIMOVIĆ, S.; WATERS, Z.; GAŠEVIĆ, D.; KITTO, K.; HATALA, M.; SIEMENS, G. Towards automated content analysis of discussion transcripts: A cognitive presence case. In: **Proceedings of the sixth international conference on learning analytics & knowledge**. [S.l.: s.n.], 2016. p. 15–24.

KOVANOVIć, V.; JOKSIMOVIć, S.; WATERS, Z.; GAsEVIć, D.; KITTO, K.; HATALA, M.; SIEMENS, G. Towards automated content analysis of discussion transcripts: A cognitive presence case. In: **Proceedings of the Sixth International Conference on Learning Analytics & Knowledge**. New York, NY, USA: Association for Computing Machinery, 2016. (LAK '16), p. 15–24. ISBN 9781450341905. Disponível em: ⟨https://doi.org/10.1145/2883851.2883950⟩.

KOVANOVIć, V.; GAšEVIć, D.; HATALA, M. Learning analytics for communities of inquiry. **Journal of Learning Analytics**, v. 1, n. 3, p. 195–198, 2014. Disponível em: ⟨https://doi.org/10.18608/jla.2014.13.21⟩.

KULHAVY, R. W.; STOCK, W. A. Feedback in written instruction: The place of response certitude. **Educational psychology review**, Springer, v. 1, n. 4, p. 279–308, 1989.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, [Wiley, International Biometric Society], v. 33, n. 1, p. 159–174, 1977. ISSN 0006341X, 15410420. Disponível em: ⟨http://www.jstor.org/stable/2529310⟩.

LANGER, P. The use of feedback in education: A complex instructional strategy. **Psychological Reports**, v. 109, n. 3, p. 775–784, 2011. PMID: 22420112. Disponível em: ⟨https://doi.org/10.2466/11.PR0.109.6.775-784⟩.

LEE, H. D. **Seleção e construção de features relevantes para o aprendizado de máquina.** Tese (Doutorado) — . Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, 2000.

LIN, C.-F.; WANG, S.-D. Fuzzy support vector machines. **IEEE transactions on neural networks**, Ieee, v. 13, n. 2, p. 464–471, 2002.

LIU., Z.; KANG., L.; DOMANSKA., M.; LIU., S.; SUN., J.; FANG., C. Social network characteristics of learners in a course forum and their relationship to learning outcomes. In: INSTICC. **Proceedings of the 10th International Conference on Computer Supported Education - Volume 2: CSEDU,**. SciTePress, 2018. p. 15–21. ISBN 978-989-758-291-2. Disponível em: ⟨https://doi.org/10.5220/0006647600150021⟩.

LO, R. T.-W.; HE, B.; OUNIS, I. Automatically building a stopword list for an information retrieval system. In: **Journal on Digital Information Management**. [S.l.: s.n.], 2005. v. 5, p. 17–24.

LOCKE, E. A.; LATHAM, G. P. Goal setting: A motivational technique that works! Prentice-Hall Englewood Cliffs, NJ, 1984. Disponível em: ⟨https://www.academia.edu/download/46606974/0090-2616_2879_2990032-920160618-1138-1deg8fw.pdf⟩.

Marin, V. J.; Pereira, T.; Sridharan, S.; Rivero, C. R. Automated personalized feedback in introductory java programming moocs. In: **2017 IEEE 33rd International Conference on Data Engineering (ICDE)**. [s.n.], 2017. p. 1259–1270. ISSN 2375-026X. Disponível em: ⟨https://doi.org/10.1109/ICDE.2017.169⟩.

MARNEFFE, M.-C. D.; MACCARTNEY, B.; MANNING, C. D. et al. Generating typed dependency parses from phrase structure parses. In: **Lrec**. [S.l.: s.n.], 2006. v. 6, p. 449–454.

MATCHA, W.; GASEVIć, D.; UZIR, N. A.; JOVANOVIć, J.; PARDO, A. Analytics of learning strategies: Associations with academic performance and feedback. In: **Proceedings of the 9th International Conference on Learning Analytics & amp; Knowledge**. New York, NY, USA: Association for Computing Machinery, 2019. (LAK19), p. 461–470. ISBN 9781450362566. Disponível em: ⟨https://doi.org/10.1145/3303772.3303787⟩.

MCGILL, T. J.; KLOBAS, J. E. A task–technology fit view of learning management system impact. **Computers & Education**, v. 52, n. 2, p. 496 – 508, 2009. ISSN 0360-1315. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0360131508001541⟩.

MCKLIN, T. E. **Analyzing Cognitive Presence in Online Courses Using an Artificial Neural Network**. Tese (Doutorado) — Georgia State University, Atlanta, GA, USA, 2004. AAI3190967. Disponível em: ⟨https://scholarworks.gsu.edu/msit_diss/1/⟩.

MCKNIGHT, P. E.; NAJAB, J. Mann-whitney u test. **The Corsini encyclopedia of psychology**, Wiley Online Library, p. 1–1, 2010.

MEHRABIAN, A. Some referents and measures of nonverbal behavior. **Behavior Research Methods & Instrumentation**, Springer, v. 1, n. 6, p. 203–207, 1968. Disponível em: ⟨https://doi.org/10.3758/BF03208096⟩.

MEINOLF, B.; URSULA, B.; MARIANNE, H.; FRITZ-OTTO, P.; HELGA, S. et al. The informational value of evaluative behavior: Influences of praise and blame on perceptions of ability. **Journal of Educational Psychology**, American Psychological Association, v. 71, n. 2, p. 259–268, 1979. Disponível em: ⟨https://psycnet.apa.org/doi/10.1037/0022-0663.71.2.259⟩.

METRIX, C. **Coh-Metrix version 3.0 indices**. [S.l.], 2020. ¡Acesso em: 15 de dez. de 2020¿. Disponível em: ⟨https://cohmetrix.memphis.edu/cohmetrixhome/documentation_indices.html⟩.

MITCHELL, R.; FRANK, E. Accelerating the xgboost algorithm using gpu computing. **PeerJ Computer Science**, v. 3, p. e127, jul. 2017. ISSN 2376-5992. Disponível em: ⟨https://doi.org/10.7717/peerj-cs.127⟩.

NETO, V.; ROLIM, V.; FERREIRA, R.; KOVANOVIĆ, V.; GAŠEVIĆ, D.; LINS, R. D.; LINS, R. Automated analysis of cognitive presence in online discussions written in portuguese. In: SPRINGER. **European Conference on Technology Enhanced Learning**. Springer International Publishing, 2018. p. 245–261. ISBN 978-3-319-98572-5. Disponível em: ⟨https://doi.org/10.1007/978-3-319-98572-5_19⟩.

NICOL, D. J.; MACFARLANE-DICK, D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. **Studies in Higher Education**, Routledge, v. 31, n. 2, p. 199–218, 2006. Disponível em: ⟨https://doi.org/10.1080/03075070600572090⟩.

NILSSON, N. J. **Introduction to machine learning: An early draft of a proposed textbook**. [S.l.]: USA; Stanford University, 1996.

OLIVARES, D.; MELLO, R. F. L. de; ADESOPE, O.; ROLIM, V.; GAŠEVIC, D.; HUNDHAUSEN, C. Using social network analysis to measure the effect of learning analytics in computing education. In: IEEE. **2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)**. 2019. v. 2161, p. 145–149. Disponível em: ⟨https://doi.org/10.1109/ICALT.2019.00044⟩.

ORRELL, J. Feedback on learning achievement: rhetoric and reality. **Teaching in Higher Education**, Routledge, v. 11, n. 4, p. 441–456, 2006. Disponível em: ⟨https://doi.org/10.1080/13562510600874235⟩.

PARIKH, A.; MCREELIS, K.; HODGES, B. Student feedback in problem based learning: a survey of 103 final year students across five ontario medical schools. **Medical Education**, v. 35, n. 7, p. 632–636, 2001. Disponível em: ⟨https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2923.2001.00994.x⟩.

PARIS, S. G.; WINOGRAD, P. Promoting metacognition and motivation of exceptional children. **Remedial and Special Education**, v. 11, n. 6, p. 7–15, 1990. Disponível em: ⟨https://doi.org/10.1177/074193259001100604⟩.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. **Journal of machine learning research**, v. 12, n. Oct, p. 2825–2830, 2011.

PENG, Y.; KOU, G.; CHEN, Z.; SHI, Y. Cross-validation and ensemble analyses on multiple-criteria linear programming classification for credit cardholder behavior. In: SPRINGER. **International Conference on Computational Science**. [S.l.], 2004. p. 931–939.

PENNEBAKER, J. W.; BOYD, R. L.; JORDAN, K.; BLACKBURN, K. **The development and psychometric properties of LIWC2015**. [S.l.], 2015. Disponível em: ⟨https://doi.org/10.15781/T29G6Z⟩.

PENNEBAKER, J. W.; FRANCIS, M. E.; BOOTH, R. J. Linguistic inquiry and word count: Liwc 2001. **Mahway: Lawrence Erlbaum Associates**, v. 71, n. 2001, p. 2001, 2001.

PESSOA, C. M. G. e T. A presença pedagógica num ambiente online criado na rede social facebook. **Educação, Formação & Tecnologias - ISSN 1646-933X**, v. 5, n. 2, 2012. ISSN ISSN 1646-933X. Disponível em: ⟨http://www.eft.educom.pt/index.php/eft/article/view/310⟩.

PURDIE, N.; HATTIE, J.; DOUGLAS, G. Student conceptions of learning and their use of self-regulated learning strategies: A cross-cultural comparison. **Journal of educational psychology**, American Psychological Association, v. 88, n. 1, p. 87, 1996. Disponível em: ⟨https://psycnet.apa.org/doi/10.1037/0022-0663.88.1.87⟩.

RAMASUBRAMANIAN, C.; RAMYA, R. Effective pre-processing activities in text mining using improved porter's stemming algorithm. **International Journal of Advanced Research in Computer and Communication Engineering**, v. 2, n. 12, p. 4536–4538, 2013.

RAMRAJ, S.; SARANYA, S.; YASHWANT, K. Comparative study of bagging, boosting and convolutional neural network for text classification. **Indian Journal of Public Health Research & Development**, Prof.(Dr) RK Sharma, v. 9, n. 9, p. 1041–1047, 2018.

RAZALI, N. M.; WAH, Y. B. et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. **Journal of statistical modeling and analytics**, v. 2, n. 1, p. 21–33, 2011.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. [S.l.]: Editora Manole Ltda, 2003.

RICHARDSON, J.; SWAN, K.; LOWENTHAL, P.; ICE, P. Social presence in online learning: Past, present, and future. In: ASSOCIATION FOR THE ADVANCEMENT OF COMPUTING IN EDUCATION (AACE). **Global Learn**. [S.l.], 2016. p. 477–483.

ROBERTSON, S. Understanding inverse document frequency: on theoretical arguments for idf. **Journal of documentation**, Emerald Group Publishing Limited, 2004.

ROSSI, R. G. **Classificação automática de textos por meio de aprendizado de máquina baseado em redes**. Tese (Doutorado) — . Tese (Doutorado em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, 2015.

ROURKE, L.; ANDERSON, T.; GARRISON, D. R.; ARCHER, W. Assessing Social Presence In Asynchronous Text-based Computer Conferencing. **The Journal of Distance Education**, v. 14, n. 2, p. 50–71, 1999. Disponível em: ⟨https://www.learntechlib.org/p/92000/⟩.

____. Methodological issues in the content analysis of computer conference transcripts. **International journal of artificial intelligence in education (IJAIED)**, v. 12, p. 8–22, 2001. Disponível em: ⟨https://telearn.archives-ouvertes.fr/hal-00197319/⟩.

SADLER, D. R. Formative assessment and the design of instructional systems. **Instructional science**, Springer, v. 18, n. 2, p. 119–144, 1989. Disponível em: ⟨https://doi.org/10.1007/BF00117714⟩.

SCARTON, C.; GASPERIN, C.; ALUISIO, S. Revisiting the readability assessment of texts in portuguese. In: SPRINGER. **Ibero-American Conference on Artificial Intelligence**. [S.l.], 2010. p. 306–315.

SEBASTIANI, F. Machine learning in automated text categorization. **ACM computing surveys (CSUR)**, ACM, v. 34, n. 1, p. 1–47, 2002.

SHAH, F. P.; PATEL, V. A review on feature selection and feature extraction for text classification. In: IEEE. **Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on**. [S.l.], 2016. p. 2264–2268.

SHANAHAN, J. G.; GREFENSTETTE, G.; QU, Y.; EVANS, D. A. Mining multilingual opinions through classification and translation. In: **Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text**. [S.l.: s.n.], 2004.

SHARP, P. Behaviour modification in the secondary school: A survey of students' attitudes to rewards and praise. **Behavioral Approaches with Children**, v. 9, p. 109–112, 1985.

SHI, L.; MIHALCEA, R.; TIAN, M. Cross language text classification by model translation and semi-supervised learning. In: **Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing**. [S.l.: s.n.], 2010. p. 1057–1067.

SPITZBERG, B. H. Preliminary development of a model and measure of computer-mediated communication (cmc) competence. **Journal of Computer-Mediated Communication**, Oxford University Press Oxford, UK, v. 11, n. 2, p. 629–666, 2006. ISSN 1083-6101. Disponível em: ⟨https://doi.org/10.1111/j.1083-6101.2006.00030.x⟩.

STANDARDIZATION, I. O. for. Standard, **Accuracy (trueness and precision) of measurement methods and results-Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method**. International Organization for Standardization, 1994. Disponível em: ⟨https://www.iso.org/standard/11834.html⟩.

Syerov, Y.; Fedushko, S.; Loboda, Z. Determination of development scenarios of the educational web forum. In: **2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT)**. [s.n.], 2016. p. 73–76. Disponível em: ⟨https://doi.org/10.1109/STC-CSIT.2016.7589872⟩.

THOMPSON, W. B. Metamemory accuracy: Effects of feedback and the stability of individual differences. **The American journal of psychology**, University of Illinois Press, v. 111, n. 1, p. 33, 1998. Disponível em: ⟨https://search.proquest.com/docview/224842869⟩.

TOLMIE, A.; BOYLE, J. Factors influencing the success of computer mediated communication (cmc) environments in university teaching: a review and case study. **Computers & Education**, Elsevier, v. 34, n. 2, p. 119–140, 2000. ISSN 0360-1315. Disponível em: ⟨https://doi.org/10.1016/S0360-1315(00)00008-7⟩.

VANDERGRIFF, I. Emotive communication online: A contextual analysis of computer-mediated communication (cmc) cues. **Journal of Pragmatics**, v. 51, p. 1 – 12, 2013. ISSN 0378-2166. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S037821661300057X⟩.

VAPNIK, V. Statistical learning theory john wiley. **New York**, 1998.

____. The support vector method of function estimation. In: **Nonlinear Modeling**. [S.l.]: Springer, 1998. p. 55–85.

VAPNIK, V. N. The nature of statistical learning. **Theory**, springer, 1995.

VIEIRA, R.; LIMA, V. L. S. Lingüística computacional: princípios e aplicações. In: SN. **Anais do XXI Congresso da SBC. I Jornada de Atualização em Inteligência Artificial**. [S.l.], 2001. v. 3, p. 47–86.

VIEIRA, R.; LOPES, L. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. **EM CORPORA**, p. 183, 2010.

WALTHER, J. B.; ANDERSON, J. F.; PARK, D. W. Interpersonal effects in computer-mediated interaction: A meta-analysis of social and antisocial communication. **Communication Research**, v. 21, n. 4, p. 460–487, 1994. Disponível em: ⟨https://doi.org/10.1177/009365094021004002⟩.

WALTHER, J. B.; HEIDE, B. V. D.; RAMIREZ, A.; BURGOON, J. K.; PEÑA, J. Interpersonal and hyperpersonal dimensions of computer-mediated communication. **The handbook of the psychology of communication technology**, Wiley Online Library, v. 1, p. 22, 2015.

WATERS, Z.; KOVANOVIĆ, V.; KITTO, K.; GAŠEVIĆ, D. Structure matters: Adoption of structured classification approach in the context of cognitive presence classification. In: SPRINGER. **AIRS**. 2015. p. 227–238. Disponível em: ⟨https://doi.org/10.1007/978-3-319-28940-3_18⟩.

WEISS, S. M.; KULIKOWSKI, C. A. **Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems**. [S.l.]: Morgan Kaufmann Publishers Inc., 1991.

WHITE, J. S.; O'CONNELL, T. A. Evaluation of machine translation. In: **HUMAN LANGUAGE TECHNOLOGY: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993**. [S.l.: s.n.], 1993.

WHITE, J. S.; O'CONNELL, T. A.; O'MARA, F. E. The arpa mt evaluation methodologies: evolution, lessons, and future approaches. In: **Proceedings of the First Conference of the Association for Machine Translation in the Americas**. [S.l.: s.n.], 1994.

WIEBE, J.; WILSON, T.; CARDIE, C. Annotating expressions of opinions and emotions in language. **Language Resources and Evaluation**, v. 39, n. 2, p. 165–210, 2005.

XU, R.; YANG, Y.; LIU, H.; HSI, A. Cross-lingual text classification via model translation with limited dictionaries. In: **Proceedings of the 25th ACM International on Conference on Information and Knowledge Management**. [S.l.: s.n.], 2016. p. 95–104.

YUSOF, N.; RAHMAN, A. A. et al. Students' interactions in online asynchronous discussion forum: A social network analysis. In: IEEE. **2009 International Conference on Education Technology and Computer**. 2009. p. 25–29. ISSN 2155-1812. Disponível em: ⟨https://doi.org/10.1109/ICETC.2009.48⟩.

ZIMMERMAN, D. W. Comparative power of student t test and mann-whitney u test for unequal sample sizes and variances. **The Journal of Experimental Education**, Taylor & Francis, v. 55, n. 3, p. 171–174, 1987.