

## DETECÇÃO DE INTENÇÕES DE COMPRA E SERVIÇOS COM USO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

## DETECTION OF PURCHASE AND SERVICE INTENTIONS USING ARTIFICIAL INTELLIGENCE TECHNIQUES

## DETECCIÓN DE INTENCIÓNES DE COMPRA Y DE SERVICIOS MEDIANTE TÉCNICAS DE INTELIGENCIA ARTIFICIAL



<https://doi.org/10.56238/ERR01v10n4-031>

**Bruno da Silva Sousa**

Bacharel em Ciência da Computação  
Instituição: Universidade Estadual do Piauí  
E-mail: bdasilvasousa@aluno.uespi.br  
Orcid: <https://orcid.org/0009-0001-6424-7446>

**Dario Brito Calçada**

Doutor em Computação e Matemática Computacional  
Instituição: Universidade Estadual do Piauí  
E-mail: dariobcalcada@phb.uespi.br  
Orcid: <https://orcid.org/0000-0003-3313-7874>

### RESUMO

**Introdução:** A crescente complexidade das interações em plataformas de e-commerce impõe desafios significativos à recuperação precisa de produtos e serviços, particularmente em consultas curtas e ruidosas em português brasileiro. Nesse contexto, modelos monolíngues baseados em arquiteturas *Transformer* emergem como soluções promissoras para capturar as nuances linguísticas locais, visando a redução de reformulações e do abandono de busca, elementos críticos para a experiência do usuário e taxas de conversão. **Objetivo:** O presente trabalho teve como objetivo principal desenvolver e avaliar um classificador de intenções de compra e contratação de serviços em consultas de e-commerce, empregando a técnica de transferência de aprendizado com modelos BERT pré-treinados especificamente para o português brasileiro. **Materiais e Métodos:** Para tanto, foi coletado e anotado manualmente um corpus de consultas obtidas na região de Parnaíba, categorizado em 92 intenções distintas. O pré-processamento incluiu limpeza, anonimização, normalização textual e a aplicação de técnicas de balanceamento e *data augmentation* para mitigar o desequilíbrio de classes. Realizou-se o *fine-tuning* do modelo `neuralmind/bert-large-portuguese-cased` (BERTimbau) utilizando as bibliotecas Hugging Face e PyTorch, com tokenização configurada para `max\_length = 128`. A divisão dos dados foi estratificada (`test\_size = 0,10`), complementada por validação cruzada estratificada, e o treinamento incorporou estratégias de regularização e *early stopping*. A avaliação de desempenho baseou-se em métricas como acurácia, precisão, *recall* e F1-score, com a fixação de sementes e o salvamento de *checkpoints* para garantir a reprodutibilidade. **Resultado:** No conjunto de validação, composto por 2.940 amostras, o modelo alcançou uma acurácia global de 96,43% e um F1-score médio ponderado de aproximadamente 0,96. As matrizes de confusão exibiram um padrão predominantemente diagonal, indicando alta precisão, embora se tenha observado maior confusão

entre categorias semanticamente próximas e um desempenho ligeiramente inferior em classes com baixo suporte. O treinamento demonstrou convergência eficiente em 3–5 épocas, sendo o melhor *checkpoint* selecionado com base na acurácia. Conclusão: Os achados demonstram que a transferência de aprendizado com modelos BERT em português constitui uma solução viável e altamente eficaz para a detecção de intenções em consultas de e-commerce. Este trabalho resultou na entrega de um *dataset* rotulado de alta qualidade e de um modelo funcional, que pode ser integrado para aprimorar a experiência do usuário e otimizar sistemas de recomendação. Para futuras implementações em produção, recomenda-se a integração de sinais contextuais (como histórico do usuário e atributos de produto), bem como a exploração de *oversampling* e hierarquização de categorias para refinar o desempenho em classes mais desafiadoras.

**Palavras-chave:** Busca Semântica. Processamento de Linguagem Natural (PLN). Classificação de Consultas. Sistemas de Recomendação. Transformers.

## ABSTRACT

**Introduction:** The increasing complexity of interactions on e-commerce platforms poses significant challenges to the precise retrieval of products and services, particularly for short and noisy queries in Brazilian Portuguese. In this context, monolingual models based on Transformer architectures emerge as promising solutions to capture local linguistic nuances, aiming to reduce reformulations and search abandonment, which are critical elements for user experience and conversion rates. **Objective:** The main objective of this work was to develop and evaluate a classifier for purchase and service intention detection in e-commerce queries, employing the transfer learning technique with BERT models specifically pre-trained for Brazilian Portuguese. **Materials and Methods:** To this end, a corpus of queries obtained from the Parnaíba region was manually collected and annotated, categorized into 92 distinct intentions. Pre-processing included cleaning, anonymization, textual normalization, and the application of balancing and data augmentation techniques to mitigate class imbalance. Fine-tuning of the *neuralmind/bert-large-portuguese-cased* (BERTimbau) model was performed using the Hugging Face and PyTorch libraries, with tokenization configured for `max_length = 128`. Data splitting was stratified (`test_size = 0.10`), complemented by stratified cross-validation, and training incorporated regularization and early stopping strategies. **Performance evaluation:** was based on metrics such as accuracy, precision, recall, and F1-score, with seed fixing and *checkpoint* saving to ensure reproducibility. **Result:** On the validation set, comprising 2,940 samples, the model achieved a global accuracy of 96.43% and a weighted average F1-score of approximately 0.96. Confusion matrices exhibited a predominantly diagonal pattern, indicating high precision, although greater confusion was observed between semantically similar categories and slightly lower performance in classes with low support. **Training:** demonstrated efficient convergence within 3–5 epochs, with the best *checkpoint* selected based on accuracy. **Conclusion:** The findings demonstrate that transfer learning with BERT models in Portuguese constitutes a viable and highly effective solution for intention detection in e-commerce queries. This work resulted in the delivery of a high-quality labeled dataset and a functional model, which can be integrated to enhance user experience and optimize recommendation systems. For future production implementations, the integration of contextual signals (such as user history and product attributes), as well as the exploration of oversampling and category hierarchization, is recommended to refine performance in more challenging classes.

**Keywords:** Semantic Search. Natural Language Processing (Nlp). Query Classification. Recommendation Systems. Transformers.

## RESUMEN

**Introducción:** La creciente complejidad de las interacciones en plataformas de comercio electrónico plantea desafíos significativos para la recuperación precisa de productos y servicios, particularmente en consultas cortas y ruidosas en portugués brasileño. En este contexto, los modelos monolingües basados en arquitecturas Transformer emergen como soluciones prometedoras para capturar los matices lingüísticos locales, con el objetivo de reducir las reformulaciones y el abandono de la búsqueda, elementos críticos para la experiencia del usuario y las tasas de conversión. **Objetivo:** El presente trabajo tuvo como objetivo principal desarrollar y evaluar un clasificador de intenciones de compra y contratación de servicios en consultas de comercio electrónico, empleando la técnica de aprendizaje por transferencia con modelos BERT pre-entrenados específicamente para el portugués brasileño. **Materiales y Métodos:** Para ello, se recopiló y anotó manualmente un corpus de consultas obtenidas en la región de Parnaíba, categorizado en 92 intenciones distintas. El preprocesamiento incluyó limpieza, anonimización, normalización textual y la aplicación de técnicas de balanceo y aumento de datos (*data augmentation*) para mitigar el desequilibrio de clases. Se realizó el ajuste fino (*fine-tuning*) del modelo *neuralmind/bert-large-portuguese-cased* (BERTimbau) utilizando las bibliotecas Hugging Face y PyTorch, con la tokenización configurada para `max_length = 128`. La división de los datos fue estratificada (`test_size = 0,10`), complementada por validación cruzada estratificada, y el entrenamiento incorporó estrategias de regularización y parada temprana (*early stopping*). La evaluación del rendimiento se basó en métricas como la exactitud (*accuracy*), precisión, exhaustividad (*recall*) y F1-score, con la fijación de semillas y el guardado de puntos de control (*checkpoints*) para garantizar la reproducibilidad. **Resultado:** En el conjunto de validación, compuesto por 2.940 muestras, el modelo alcanzó una exactitud global del 96,43% y un F1-score promedio ponderado de aproximadamente 0,96. Las matrices de confusión exhibieron un patrón predominantemente diagonal, indicando alta precisión, aunque se observó una mayor confusión entre categorías semánticamente cercanas y un rendimiento ligeramente inferior en clases con bajo soporte. El entrenamiento demostró una convergencia eficiente en 3–5 épocas, seleccionándose el mejor punto de control basándose en la exactitud. **Conclusión:** Los hallazgos demuestran que el aprendizaje por transferencia con modelos BERT en portugués constituye una solución viable y altamente eficaz para la detección de intenciones en consultas de comercio electrónico. Este trabajo resultó en la entrega de un conjunto de datos etiquetado de alta calidad y un modelo funcional, que puede integrarse para mejorar la experiencia del usuario y optimizar los sistemas de recomendación. Para futuras implementaciones en producción, se recomienda la integración de señales contextuales (como el historial del usuario y los atributos del producto), así como la exploración del sobremuestreo (*oversampling*) y la jerarquización de categorías para refinar el rendimiento en clases más desafiantes.

**Palabras clave:** Búsqueda Semántica. Procesamiento de Lenguaje Natural (PLN). Clasificación de Consultas. Sistemas de Recomendación. Transformers.

## 1 INTRODUÇÃO

O cenário contemporâneo do comércio eletrônico é marcado por uma expansão digital sem precedentes, que redefiniu fundamentalmente a interação entre consumidores e o vasto universo de produtos e serviços disponíveis. Neste ambiente dinâmico, a eficácia dos mecanismos de busca e recuperação de informações assume um papel central, influenciando diretamente a experiência do usuário e as taxas de conversão (Khurana et al., 2023). Contudo, os sistemas tradicionais de recuperação de informação frequentemente demonstram limitações significativas na interpretação semântica de consultas, especialmente em línguas com rica complexidade morfológica e sintática como o português brasileiro. Essa lacuna resulta em assimetria informacional, onde a real intenção de busca do usuário não é plenamente compreendida, levando a resultados subótimos e à reformulação exaustiva de consultas (Borges, 2024; Xie et al., 2024).

Para mitigar essas deficiências, o campo da Inteligência Artificial (IA), com particular destaque para o Processamento de Linguagem Natural (PLN), tem desenvolvido soluções inovadoras. O aprendizado profundo, caracterizado por arquiteturas computacionais multicamadas, tem impulsionado avanços substanciais em diversas tarefas linguísticas, elevando o estado da arte em reconhecimento de fala, detecção de objetos e outras aplicações (LeCun; Bengio; Hinton, 2015; Schmidhuber, 2015). Dentro deste contexto, as arquiteturas baseadas em Transformers emergiram como um paradigma revolucionário para a representação semântica, capazes de capturar dependências globais no texto de forma eficiente e paralela (Vaswani et al., 2023).

O modelo *Bidirectional Encoder Representations from Transformers* (BERT), pré-treinado em extensos corpora textuais, demonstrou uma capacidade superior de compreender as nuances contextuais da linguagem (Devlin et al., 2019). A estratégia de transferência de aprendizado, que consiste em capitalizar o conhecimento adquirido em uma tarefa-fonte para otimizar o desempenho em uma tarefa-alvo, revela-se particularmente eficaz para adaptar esses modelos robustos a domínios específicos, minimizando a necessidade de vastos conjuntos de dados rotulados (Pan; Yang, 2010). No contexto da língua portuguesa, iniciativas como o BERTimbau têm disponibilizado modelos monolíngues pré-treinados, permitindo uma análise semântica e classificação de intenções mais precisa e adaptada às particularidades linguísticas locais (Caseli; Freitas; Viola, 2022; Peeters; Bizer, 2022; Specia; Rino, 2002).

Apesar dos avanços, a aplicação e a avaliação rigorosa desses modelos para a detecção de intenções de compra e contratação de serviços em consultas de e-commerce em português brasileiro – um cenário frequentemente caracterizado por consultas curtas e ruidosas – ainda representam um desafio e uma área de pesquisa com significativas implicações práticas. Diante disso, o presente estudo tem como objetivo principal desenvolver e avaliar um classificador de intenções de compra e

contratação de serviços em consultas de e-commerce, empregando técnicas de transferência de aprendizado com modelos BERT pré-treinados para o português brasileiro. Para tanto, foi coletado e anotado um corpus de consultas da região de Parnaíba, sobre o qual o modelo *neuralmind/bert-large-portuguese-cased* foi ajustado (*fine-tuned*) e avaliado. Os resultados esperados incluem a validação da viabilidade e eficácia do aprendizado por transferência para a classificação de intenções em português, a disponibilização de um *dataset* rotulado de alta qualidade e um modelo funcional. Esta pesquisa oferece subsídios técnicos cruciais para aprimorar a descoberta de produtos, personalizar a experiência do usuário e otimizar as estratégias de recomendação no mercado de e-commerce lusófono, contribuindo metodologicamente para o avanço do PLN em português.

## 2 REFERENCIAL TEÓRICO

A presente seção estabelece os fundamentos teóricos que sustentam a metodologia e os objetivos deste estudo, abordando os conceitos centrais do Processamento de Linguagem Natural (PLN), as bases das Redes Neurais e do Aprendizado Profundo, e a representação semântica avançada proporcionada pelos modelos *Transformer*, com foco particular no BERT.

### 2.1 DO PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)

O Processamento de Linguagem Natural (PLN) constitui um campo interdisciplinar da Inteligência Artificial (IA) dedicado à interação entre computadores e a linguagem humana. Seu objetivo primordial é capacitar máquinas a compreender, interpretar e gerar linguagem natural de maneira significativa (Xie et al., 2024). O escopo do PLN é vasto, abrangendo desde tarefas lexicais e sintáticas, como a tokenização e a marcação de classes gramaticais, até desafios semânticos e pragmáticos mais complexos, como a resolução de referências e a geração de texto coerente (Khurana et al., 2023).

A relevância do PLN tem crescido exponencialmente em diversas aplicações práticas, incluindo tradução automática, detecção de spam, extração de informações e sumarização de documentos. A motivação subjacente a esses avanços reside na busca por uma interface mais intuitiva e eficiente entre usuários e sistemas computacionais, permitindo a comunicação em linguagem natural (Khurana et al., 2023). Fundamentalmente, uma linguagem natural pode ser concebida como um sistema de símbolos governado por um conjunto de regras, onde a combinação desses símbolos transmite informações. O PLN, nesse contexto, pode ser categorizado em duas vertentes principais: a Compreensão da Linguagem Natural (CLN), que se concentra na interpretação do significado, e a Geração da Linguagem Natural (GLN), que visa à produção de texto (Khurana et al., 2023). Para o



presente trabalho, a CLN é de particular interesse, pois a detecção de intenções de compra e serviço exige uma profunda compreensão semântica das consultas dos usuários.

## 2.2 DO CONCEITO DE REDES NEURAIS E APRENDIZADO PROFUNDO

No cerne das recentes inovações em PLN encontram-se as Redes Neurais (RNs), modelos computacionais inspirados na estrutura e funcionamento do cérebro humano. Uma RN padrão é composta por múltiplos processadores simples e interconectados, denominados neurônios, que produzem sequências de ativações de valor real. Os neurônios de entrada são ativados por estímulos do ambiente, enquanto outros neurônios são ativados por meio de conexões ponderadas com neurônios precedentes (Schmidhuber, 2015).

O treinamento de uma RN envolve o ajuste desses pesos sinápticos para que a rede exiba um comportamento desejado. Este processo, conhecido como aprendizado ou atribuição de créditos, busca encontrar a configuração de pesos que otimize a performance da rede em uma determinada tarefa. Em problemas complexos, onde longas cadeias causais de estágios computacionais transformam a ativação agregada da rede, o Aprendizado Profundo (*Deep Learning*) surge como uma abordagem essencial. O Aprendizado Profundo permite que modelos computacionais, caracterizados por múltiplas camadas de processamento, aprendam representações de dados com diversos níveis de abstração (LeCun; Bengio; Hinton, 2015; Schmidhuber, 2015). Esses métodos têm aprimorado drasticamente o estado da arte em áreas como reconhecimento de fala e visão computacional, ao descobrir estruturas complexas em grandes conjuntos de dados, utilizando algoritmos como a retropropagação para ajustar os parâmetros internos do modelo (LeCun; Bengio; Hinton, 2015).

Uma técnica particularmente relevante dentro do Aprendizado Profundo é a Transferência de Aprendizado (*Transfer Learning*). Conforme Pan e Yang (2010), a transferência de aprendizado consiste em capitalizar o conhecimento adquirido em uma tarefa ou domínio-fonte para aprimorar o desempenho em uma tarefa ou domínio-alvo. Esta abordagem é especialmente valiosa em cenários práticos onde os dados de treinamento e teste não compartilham o mesmo espaço ou distribuição, ou quando a coleta de grandes volumes de dados rotulados para a tarefa-alvo é inviável. Pan e Yang (2010) categorizam os cenários de transferência em indutiva (tarefa-alvo difere, rótulos disponíveis no alvo), transdutiva (tarefa a mesma, domínios diferem, usa dados não rotulados do alvo) e não supervisionada (sem rótulos em nenhum domínio). Além disso, eles identificam quatro formas de “o que transferir”: amostras, representações de características, parâmetros de modelos e relações estruturais. A aplicação da transferência de aprendizado é crucial para este estudo, pois permite o ajuste de modelos de linguagem pré-treinados em larga escala para a tarefa específica de detecção de

intenções em consultas de e-commerce em português, superando as limitações de dados anotados (Kim et al., 2021; Zhang et al., 2020).

### 2.3 REPRESENTAÇÃO SEMÂNTICA COM TRANSFORMERS E BERT

A representação semântica é um pilar fundamental do PLN, compreendida como a formalização do significado de sentenças, identificando conceitos e as relações entre eles (Specia; Rino, 2002). No contexto computacional, essa formalização emprega metalinguagens específicas para descrever o vocabulário e a gramática, resultando em estruturas claras, não ambíguas e interpretáveis por sistemas. Uma abordagem proeminente para a representação semântica é o uso de vetores semânticos, onde cada palavra é mapeada para um ponto em um espaço vetorial multidimensional. Esses vetores capturam diversos aspectos do significado lexical, como similaridade (“comércio” e “negócio”), polaridade e associação, sendo construídos a partir da distribuição das palavras vizinhas (Caseli; Freitas; Viola, 2022; Specia; Rino, 2002). A construção desses espaços vetoriais, fundamentada em Álgebra Linear, permite representar não apenas palavras, mas também sentenças e documentos inteiros, geralmente por meio de matrizes de coocorrência (termo-documento ou termo-contexto).

A arquitetura *Transformer* revolucionou a forma como essas representações vetoriais são geradas e processadas. Diferentemente dos modelos recorrentes anteriores, o *Transformer* abandona a recorrência e baseia-se inteiramente em um mecanismo de atenção para capturar dependências globais entre entrada e saída, permitindo uma paralelização significativamente maior no treinamento (Vaswani et al., 2023). A arquitetura é composta por pilhas de codificadores e decodificadores. Cada camada do codificador inclui um sub-nível de atenção *multi-head*, que calcula as relações entre todos os elementos da entrada, e um sub-nível de rede *feed-forward*. O decodificador, por sua vez, adiciona uma atenção que consulta a saída do codificador, garantindo o uso de informações do texto original para a geração subsequente. A ausência de mecanismos recorrentes é compensada pela codificação posicional, que informa a ordem dos tokens através de vetores baseados em funções seno e cosseno, somados aos *embeddings* de entrada. Essa combinação de atenção *multi-head* e codificação posicional confere ao *Transformer* a capacidade de aprender dependências entre palavras de forma rápida e paralela (Vaswani et al., 2023).

Dentro da família dos *Transformers*, o modelo *Bidirectional Encoder Representations from Transformers* (BERT) destaca-se por sua capacidade de gerar representações de linguagem pré-treinadas. O BERT utiliza uma arquitetura de *Transformers* bidirecionais para capturar, em todas as camadas, o contexto tanto à esquerda quanto à direita de cada *token*. Seu pré-treinamento é realizado através de duas tarefas principais: o *Masked Language Model* (MLM), que mascara aleatoriamente

palavras no texto e as prevê com base no contexto circundante, e o *Next Sentence Prediction* (NSP), que treina o modelo a reconhecer se uma sentença sucede logicamente outra no corpus (Devlin et al., 2019). Essa abordagem bidirecional e as tarefas de pré-treinamento permitem que o BERT produza vetores contextuais profundos que podem ser ajustados (*fine-tuned*) para diversas tarefas de PLN, como *question answering* e inferência textual, com a adição de apenas uma camada de saída. A capacidade do BERT de compreender o contexto bidirecional é fundamental para a detecção precisa de intenções em consultas de e-commerce, onde a ambiguidade e a concisão das frases exigem uma compreensão contextual rica para mapear corretamente a intenção do usuário (Yulianton; Candra Noor Santi, 2024).

### 3 METODOLOGIA

A abordagem metodológica deste trabalho foi delineada para garantir o rigor científico e a reproduzibilidade dos resultados, combinando uma revisão bibliográfica sistemática com um desenho experimental aplicado. O objetivo primordial foi desenvolver, treinar e avaliar um classificador de intenções de busca baseado em modelos da família *Transformer*, especificamente adaptado para consultas em português brasileiro no contexto do comércio eletrônico.

#### 3.1 PESQUISA BIBLIOGRÁFICA E FUNDAMENTAÇÃO TEÓRICA

Inicialmente, realizou-se uma pesquisa bibliográfica abrangente com o intuito de fundamentar teoricamente as escolhas de arquiteturas de modelos, métricas de avaliação e procedimentos experimentais. Esta etapa envolveu a revisão da literatura pertinente ao Processamento de Linguagem Natural (PLN) em português, com foco em modelos pré-treinados como o BERT e suas variantes monolíngues. Foram exploradas as técnicas de transferência de aprendizado e as abordagens de avaliação empregadas em tarefas de classificação de intenção, visando identificar as melhores práticas e os desafios inerentes a este domínio. A análise crítica da literatura orientou a seleção do modelo-base, a definição dos hiperparâmetros de treinamento e a escolha das métricas de desempenho mais adequadas para a avaliação do classificador proposto.

#### 3.2 DESENHO EXPERIMENTAL

O desenho experimental compreendeu as seguintes etapas detalhadas: coleta e pré-processamento de dados, divisão e validação do conjunto de dados, configuração do modelo e treinamento, e avaliação de desempenho.

### 3.2.1 Coleta e Pré-processamento do Corpus

Um corpus de consultas de e-commerce foi coletado especificamente na região de Parnaíba, Piauí, Brasil. A escolha desta região visou capturar as nuances linguísticas e os padrões de busca locais, que podem diferir de corpora mais genéricos. O corpus foi subsequentemente anotado manualmente em 92 categorias distintas de intenção de compra e contratação de serviços por especialistas, seguindo um conjunto de diretrizes de rotulagem previamente estabelecidas para garantir consistência. A anotação manual é crucial para a criação de um *ground truth* de alta qualidade, essencial para o treinamento e avaliação de modelos de aprendizado de máquina.

Após a coleta, o conjunto de dados passou por um rigoroso processo de pré-processamento para garantir sua qualidade e adequação ao treinamento do modelo:

- Verificação e Tratamento de Inconsistências: Foram inspecionados valores ausentes e inconsistências nos dados. Linhas sem rótulo ou com informações corrompidas foram identificadas e tratadas, seja por remoção ou por imputação, conforme a natureza da inconsistência.
- Anonimização: Em conformidade com as boas práticas éticas e de privacidade de dados, todas as informações sensíveis e de identificação pessoal presentes nas consultas foram anonimizadas. Este processo envolveu a substituição de nomes, endereços, telefones e outras informações potencialmente identificáveis por *placeholders* genéricos, assegurando a proteção da privacidade dos usuários.
- Normalização Textual: As consultas foram submetidas a procedimentos de normalização para padronizar o texto. Isso incluiu a remoção de espaços excedentes, a padronização da caixa (mantendo a original, pois o modelo “cased” seria utilizado), a remoção de caracteres especiais e a uniformização de abreviações comuns, visando reduzir o ruído e aumentar a consistência dos dados de entrada.
- Codificação de Classes: Para a representação das 92 categorias de intenção, aplicou-se o “*LabelEncoder*”. Este procedimento mapeia as *strings* textuais das categorias para identificadores numéricos inteiros, o que é um requisito para o treinamento de modelos de classificação. Os mapeamentos “*id2label*” (numérico para textual) e “*label2id*” (textual para numérico) foram preservados e documentados para garantir a reprodutibilidade do experimento e facilitar o posterior *deploy* do modelo.

### 3.2.2 Divisão e Validação do Conjunto de Dados

O conjunto de dados pré-processado foi dividido em conjuntos de treinamento e avaliação (validação) utilizando uma partição estratificada. A estratificação é uma técnica crucial para garantir

que a distribuição das classes de intenção seja preservada de forma proporcional tanto no conjunto de treinamento quanto no de avaliação, o que é particularmente importante em conjuntos de dados com desequilíbrio de classes. A proporção definida para o conjunto de avaliação foi de “`test_size = 0.10`”, ou seja, 10% do total de amostras, resultando em 2.940 amostras para validação.

Adicionalmente, para obter estimativas de desempenho mais robustas e reduzir a variância associada a uma única divisão, foram conduzidas avaliações complementares por meio de validação cruzada estratificada (*k-fold cross-validation*, com  $k=5$ ). Esta técnica permite que cada amostra do conjunto de dados seja usada tanto para treinamento quanto para validação, garantindo uma avaliação mais completa da generalização do modelo.

### 3.2.3 Pré-processamento Textual para o Modelo BERT

O pré-processamento textual específico para o modelo BERT seguiu as exigências do *tokenizer* selecionado, “`BertTokenizer.from_pretrained(neuralmind/bert-large-portuguese-cased)`”. Este *tokenizer* é otimizado para o português brasileiro e preserva a capitalização (*cased*), o que é benéfico para a compreensão de entidades nomeadas e nuances semânticas. As etapas incluíram:

- Tokenização: As consultas foram convertidas em sequências de *tokens* (palavras ou subpalavras) compreendidas pelo modelo BERT.
- “`max_length = 128`”: Todas as sequências foram limitadas a um comprimento máximo de 128 *tokens*. Esta escolha foi baseada em uma análise da distribuição do comprimento das consultas no corpus, que indicou que a maioria das consultas se encaixava neste limite, e em considerações de custo computacional.
- Truncation e Padding: Consultas mais longas que “`max_length`” foram truncadas, e consultas mais curtas foram preenchidas (*padded*) com *tokens* especiais (`[PAD]`) até atingir o “`max_length`”, garantindo que todas as entradas para o modelo tivessem o mesmo formato.

### 3.2.4 Estratégias para Mitigar o Desequilíbrio de Classes

Reconhecendo que conjuntos de dados de intenção de busca frequentemente apresentam desequilíbrio entre as classes (algumas categorias com muitas amostras, outras com poucas), foram consideradas e aplicadas diversas estratégias para atenuar esse problema e evitar que o modelo viesse a superestimar as classes majoritárias:

- Reamostragem (Oversampling/Undersampling): Foram exploradas técnicas de reamostragem, como o *oversampling* de classes minoritárias (duplicando ou gerando novas

1 0 0  
1 1 0  
1 0 0  
1 0 1 0 1 1 0 0  
1 1 1 1 0 0  
1 0 0 1 0 0  
1 0 1 0 0 1 0 0 1

1 0 0 1 0 0  
1 1 0 1 1 0  
1 0 0 1 0 0  
1 0 1 0 1 0 1 1  
1 1 1 0 0 1 1 0 0  
1 0 0 1 1 0 1 0 0  
1 0 1 1 0 0 1 0 1  
1 0 1 0 1

amostras sintéticas, como via SMOTE) e/ou o *undersampling* de classes majoritárias, para balancear a distribuição das classes no conjunto de treinamento.

- **Ponderação de Perda (*compute\_class\_weight*)**: Implementou-se a ponderação de perda, onde as classes minoritárias recebem um peso maior na função de perda durante o treinamento. Isso significa que erros de classificação em classes com menor suporte são penalizados mais severamente, incentivando o modelo a aprender suas características distintivas.
- **Data Augmentation Textual**: Para aumentar a diversidade e a robustez do conjunto de treinamento, especialmente para classes com baixo suporte, foram aplicadas técnicas de *data augmentation* textual. Estas incluíram:
  - *Substituição por Sinônimos*: Palavras-chave nas consultas foram substituídas por sinônimos, utilizando recursos léxicos do português.
  - *Reformulações Leves*: Pequenas alterações na estrutura da frase ou na ordem das palavras foram introduzidas, mantendo o significado original.
  - *Simulação de Erros de Digitação*: Foram inseridos erros ortográficos comuns (e.g., troca de letras, omissão de caracteres) para simular o ruído presente em consultas reais de usuários e aumentar a robustez do modelo a variações.

### 3.2.5 Implementação e Treinamento do Modelo

A implementação do modelo foi realizada utilizando as bibliotecas *Hugging Face Transformers* e *PyTorch*, que fornecem uma estrutura robusta para trabalhar com modelos de linguagem pré-treinados. O modelo escolhido foi o “*BertForSequenceClassification*”, que consiste em uma arquitetura BERT (“*neuralmind/bert-large-portuguese-cased*”) com uma camada de classificação linear adicionada no topo, otimizada para tarefas de classificação de sequência. O número de rótulos (*num\_labels*) foi configurado dinamicamente com base no número total de categorias identificadas pelo *LabelEncoder*.

O gerenciamento do treinamento foi conduzido por meio da classe “*Trainer*” da biblioteca *Hugging Face*, que abstrai muitos dos detalhes do *loop* de treinamento e avaliação. Os principais hiperparâmetros configurados foram:

- “*num\_train\_epochs = 5*”: O número total de épocas de treinamento. Este valor foi determinado empiricamente, buscando um equilíbrio entre a convergência do modelo e a prevenção de *overfitting*.
- “*per\_device\_train\_batch\_size = 32*”: O tamanho do *batch* de treinamento por dispositivo (GPU). Um tamanho de *batch* de 32 é um valor comum que oferece um bom balanço entre estabilidade de gradiente e eficiência de memória.



- “learning\_rate = 3e-5”: A taxa de aprendizado inicial, um valor típico para fine-tuning de modelos BERT.
- “warmup\_steps = 500”: O número de passos de *warmup* para a taxa de aprendizado, onde a taxa de aprendizado aumenta gradualmente antes de decair.
- “weight\_decay = 0”: O coeficiente de regularização L2, configurado como 0 para este experimento, com a regularização sendo tratada por outras estratégias como *early stopping*.

As seguintes estratégias de treinamento foram estabelecidas para otimizar o processo e garantir a seleção do melhor modelo:

- “evaluation\_strategy = epoch”: A avaliação do modelo no conjunto de validação foi realizada ao final de cada época de treinamento.
- “save\_strategy = epoch”: Os *checkpoints* do modelo foram salvos ao final de cada época.
- “load\_best\_model\_at\_end = True”: Após o treinamento, o modelo com o melhor desempenho no conjunto de validação (conforme a métrica definida) foi automaticamente carregado.
- “metric\_for\_best\_model = accuracy”: A acurácia foi definida como a métrica principal para determinar o melhor modelo durante o treinamento.
- “fp16 = True”: Utilizou-se treinamento com precisão mista (float16), o que acelera o treinamento e reduz o consumo de memória da GPU, sem perda significativa de desempenho.
- Early Stopping: Uma estratégia de *early stopping* foi aplicada com “patience = 3” épocas consecutivas. Isso significa que o treinamento seria interrompido se a métrica de desempenho no conjunto de validação (*accuracy*) não melhorasse por três épocas consecutivas, prevenindo o *overfitting* e otimizando o tempo de treinamento.

### 3.2.6 Avaliação de Desempenho e Robustez

A avaliação do desempenho do classificador foi realizada utilizando um conjunto de métricas tradicionais de classificação, que fornecem uma visão abrangente da sua eficácia:

- Acurácia: Proporção de classificações corretas sobre o total de amostras.
- F1-score Ponderado: Média harmônica ponderada da precisão e do *recall*, considerando o suporte de cada classe. É uma métrica robusta para conjuntos de dados com classes desbalanceadas.
- Precisão (Precision) por Classe: Proporção de verdadeiros positivos sobre o total de positivos preditos para cada classe, indicando a qualidade das previsões positivas.

- Recall por Classe: Proporção de verdadeiros positivos sobre o total de positivos reais para cada classe, indicando a capacidade do modelo de encontrar todas as instâncias positivas.

Adicionalmente, foram geradas matrizes de confusão para cada classe, permitindo uma análise detalhada dos erros de classificação. As matrizes de confusão são ferramentas visuais que ajudam a identificar quais classes estão sendo confundidas entre si, fornecendo *insights* valiosos para o refinamento do modelo e do esquema de rotulagem.

Para mensurar a robustez do modelo em cenários práticos, foram conduzidos experimentos com conjuntos de dados contendo ruído controlado. Este ruído foi artificialmente introduzido para simular condições de uso real, como erros ortográficos, substituição por sinônimos e reformulações de consultas, avaliando a degradação do desempenho do modelo sob essas condições.

### 3.2.7 Reprodutibilidade e Aspectos Éticos

Para garantir a reproduzibilidade dos experimentos, todas as sementes aleatórias foram fixadas (“`torch.manual_seed(42)`”). As versões das bibliotecas utilizadas, os hiperparâmetros de treinamento e todos os registros experimentais foram devidamente documentados. Os artefatos resultantes do treinamento, como os *checkpoints* do modelo (“`model.save_pretrained()`”), o *tokenizer* (“`tokenizer.save_pretrained()`”) e os mapeamentos de rótulos, foram preservados. Além disso, os logs de treinamento e avaliação foram registrados no *TensorBoard* para visualização e análise posterior.

Conforme mencionado na seção de pré-processamento, todos os dados textuais passaram por um processo rigoroso de anonimização, assegurando a exclusão de informações sensíveis e a conformidade com as boas práticas de tratamento de dados e privacidade.

### 3.2.8 Limitações Metodológicas

É importante reconhecer as limitações inerentes a este estudo. A dependência do corpus coletado na região de Parnaíba pode afetar a generalização do modelo para catálogos de produtos ou regiões com vocabulário e padrões de busca significativamente distintos. O elevado custo computacional associado ao treinamento e *fine-tuning* de modelos de grande porte como o BERT representa uma barreira para a experimentação em larga escala. Além disso, a ausência de uma avaliação em dados fora da distribuição (*out-of-distribution*) e a necessidade de validação em ambiente de produção por meio de experimentos online (*A/B testing*) são aspectos que fundamentam recomendações para trabalhos futuros. Tais limitações, no entanto, não invalidam os achados, mas sim delimitam o escopo de aplicabilidade e apontam para direções de pesquisa subsequentes.

## 4 RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados obtidos a partir do *fine-tuning* da arquitetura BERT para a detecção de intenções de compra e serviço, seguido de uma discussão aprofundada sobre suas implicações, pontos fortes, desafios e contribuições para o campo do Processamento de Linguagem Natural (PLN) e do comércio eletrônico.

### 4.1 ANÁLISE DE DESEMPENHO GLOBAL DO MODELO

Os experimentos foram conduzidos mediante o fine-tuning da arquitetura neuralmind/bert-large-portuguese-cased sobre o corpus de consultas de e-commerce coletado e pré-processado, conforme detalhado na Seção 4. O conjunto de avaliação, composto por 2.940 amostras distribuídas entre as 92 categorias de intenção, serviu como base para a mensuração do desempenho do modelo.

O modelo alcançou uma acurácia global de 96,43% (accuracy = 0,964286) no conjunto de validação, acompanhada por um F1-score médio ponderado de aproximadamente 0,96 e médias macro de precisão e recall em torno de 0,97. Estes resultados demonstram uma notável consistência e robustez na capacidade do modelo em discriminar entre um grande número de categorias de intenção, mesmo em um cenário de consultas curtas e potencialmente ruidosas em português brasileiro. A evolução do treinamento, conforme ilustrado na Tabela 1, revela um processo de convergência eficiente e estável.

Tabela 1 – Evolução por época .

Época	Validation Loss	F1 Score	Accuracy	Training Loss
1°	0.376643	0.916213	0.917007	0.563400
2°	0.204146	0.949807	0.950000	0.235200
3°	0.180423	0.955972	0.955782	0.114200
4°	0.185791	0.958632	0.958503	0.052700
5°	0.178379	0.964434	0.964286	0.022400

Fonte: Dados do autor, 2025.

É evidenciado na Tabela 1 uma queda consistente da training loss ao longo das épocas, indicando que o modelo estava aprendendo efetivamente a mapear as consultas às suas intenções. Concomitantemente, a validation loss demonstrou uma estabilização a partir da terceira-quarta época, um indicativo de que o modelo estava generalizando bem para dados não vistos e que a estratégia de early stopping (com paciência de três épocas) foi eficaz em prevenir o overfitting. A seleção do melhor modelo com base na acurácia do conjunto de validação garantiu que a versão final do classificador representasse o ponto ótimo de generalização.

#### 4.2 ANÁLISE DE DESEMPENHO POR CLASSE E MATRIZES DE CONFUSÃO

A análise das matrizes de confusão revelou um padrão predominantemente diagonal, o que corrobora a alta acurácia global e o F1-score médio. Este padrão indica que a vasta maioria das instâncias foi corretamente classificada, com poucos erros de predição. No entanto, os erros residuais observados concentraram-se em confusões entre classes semanticamente próximas. Este fenômeno é intrínseco à natureza do problema de classificação de consultas em e-commerce, onde a brevidade e a ambiguidade das consultas podem levar a sobreposições semânticas entre categorias. Por exemplo, a distinção entre “Utensílios de Cozinha” e “Produtos Descartáveis (Embalagens e Utensílios)” pode ser tênue em consultas curtas, onde o contexto é limitado. Em um nível mais granular, o desempenho por classe apresentou variações:

- **Classes de Alto Desempenho:** Muitas categorias demonstraram um desempenho excepcional, com F1-score aproximando-se de 1,00. Exemplos notáveis incluem “Assistência Técnica para Celulares”, “Celulares e Smartphones” e “Madeireira / Comércio de Madeira”. O alto desempenho nessas classes sugere que os rótulos são consistentes e que as consultas associadas possuem sinais semânticos suficientemente distintos para que o modelo aprenda padrões discriminativos robustos. A especificidade do vocabulário e a clareza da intenção nessas categorias contribuem para a facilidade de classificação.
- **Classes de Desempenho Relativamente Inferior:** Algumas classes exibiram um desempenho ligeiramente inferior, embora ainda robusto. Por exemplo, “Produtos Descartáveis (Embalagens e Utensílios)” obteve  $precision = 0,79$ ,  $recall = 0,94$  e  $F1 = 0,86$  (suporte = 33); “Utensílios e Equipamentos de Cozinha” apresentou  $precision = 0,85$ ,  $recall = 0,85$  e  $F1 = 0,85$  (suporte = 46); classes de Eletrodomésticos/Eletrônicos variaram entre  $F1 \approx 0,90-0,91$ ; “Ferramentas e Equipamentos Industriais” ficou com  $F1 \approx 0,89$  (suporte = 30); e “Suplementos Nutricionais” exibiu  $recall$  mais alto que  $precision$ , resultando em  $F1 \approx 0,92$ . Essas diferenças podem ser atribuídas a múltiplos fatores, incluindo:
  - *Sobreposição Semântica:* Como mencionado, a proximidade conceitual entre certas categorias pode gerar ambiguidade para o modelo.
  - *Baixo Suporte:* Classes com menor número de amostras no conjunto de treinamento tendem a ser mais desafiadoras para o modelo aprender representações robustas, uma vez que há menos exemplos para generalizar.
  - *Heterogeneidade Textual:* Consultas dentro de uma mesma categoria podem apresentar grande variabilidade na formulação, dificultando a identificação de padrões consistentes.

1 0 0  
1 1 0  
1 0 0  
1 0 1 0 1 1 0 0  
1 1 1 1 0 0  
1 0 0 1 0 0  
1 0 1 0 1 0 1 0 1

1 0 0 1 0 0  
1 1 0 1 1 0  
1 0 0 1 0 0  
1 0 1 0 1 0 1  
1 1 0 0 1 1 1 1  
1 0 0 1 1 0 1 0 0  
1 0 1 0 0 1 0 1  
1 0 1 0 1

#### 4.3 DISCUSSÃO DOS ACHADOS E IMPLICAÇÕES

Os resultados obtidos confirmam a viabilidade e a eficácia da estratégia de utilizar modelos Transformer monolíngues pré-treinados em português, ajustados (fine-tuned) sobre um corpus rotulado de consultas de e-commerce, para a tarefa de detecção de intenções. A alta acurácia e o F1-score médio demonstram que a abordagem de transferência de aprendizado, conforme discutido por Pan e Yang (2010), é uma solução poderosa para domínios específicos com dados limitados, permitindo que o conhecimento linguístico geral pré-adquirido pelo BERT (Devlin et al., 2019) seja adaptado com sucesso para a tarefa de classificação de intenções.

Um dos impactos positivos mais significativos deste trabalho reside na capacidade do modelo em capturar as nuances morfológicas e as expressões idiomáticas locais do português brasileiro. O uso do BERTimbau, um modelo pré-treinado especificamente para o português, é crucial para este sucesso, pois ele incorpora o conhecimento linguístico intrínseco à língua, algo que modelos treinados em outras línguas ou em corpora multilíngues genéricos poderiam negligenciar (Caseli; Freitas; Viola, 2022; Hammes; Freitas, 2021). Essa sensibilidade linguística é fundamental para aprimorar a compreensão semântica de consultas de usuários, que frequentemente são informais, abreviadas ou contêm gírias regionais.

As implicações práticas deste classificador para plataformas de e-commerce são substanciais:

- Melhoria da Experiência do Usuário: Ao mapear com alta precisão as consultas dos usuários às suas intenções reais, o sistema pode fornecer resultados de busca mais relevantes e personalizados. Isso reduz a necessidade de reformulações sucessivas de consultas, minimizando a frustração do usuário e o abandono da busca, conforme apontado na introdução.
- Otimização de Sistemas de Recomendação: A detecção de intenções pode alimentar sistemas de recomendação, permitindo a sugestão de produtos e serviços mais alinhados aos interesses explícitos e implícitos do usuário. Isso pode levar a um aumento nas taxas de conversão e na satisfação do cliente.
- Análise de Mercado e Personalização: O classificador oferece uma ferramenta poderosa para a análise de tendências de busca e demandas de mercado. Ao categorizar automaticamente as intenções, as empresas podem identificar lacunas em seus catálogos, otimizar estratégias de marketing e personalizar ofertas de forma mais eficaz.
- Contribuição Metodológica e Artefatos: Este estudo não apenas valida uma abordagem eficaz, mas também gera dois artefatos de valor inestimável: um conjunto de dados rotulado de alta qualidade de consultas de e-commerce em português brasileiro e um modelo funcional que pode ser diretamente integrado a processos de busca e recomendação. Estes artefatos servem

como base para futuras pesquisas e para o desenvolvimento de soluções comerciais no mercado lusófono.

A robustez demonstrada pelo modelo, mesmo diante de consultas ruidosas (simuladas com erros ortográficos e reformulações), reforça sua aplicabilidade em cenários de uso real, onde a entrada do usuário raramente é perfeita. A capacidade dos *Transformers* de construir representações semânticas contextuais é a base para essa resiliência (Specia; Rino, 2002; Vaswani et al., 2023).

#### 4.4 RECOMENDAÇÕES E DIREÇÕES FUTURAS

Apesar do desempenho promissor, a análise detalhada por classe e das matrizes de confusão aponta para áreas claras de aprimoramento e direções para trabalhos futuros:

- Refinamento do Esquema de Rotulagem e Análise de Erro Manual: Para mitigar as confusões entre classes semanticamente próximas, recomenda-se a realização de uma análise de erro manual aprofundada. Este processo permitiria inspecionar as instâncias mal classificadas e, quando pertinente, revisar e padronizar o esquema de rotulagem, tornando as distinções entre categorias mais explícitas para os anotadores e, consequentemente, para o modelo.
- Estratégias Avançadas para Classes Fracas: As classes com menor suporte e maior heterogeneidade textual podem se beneficiar de técnicas mais agressivas de aumento de dados (data augmentation) ou oversampling (Pan; Yang, 2010). A geração de amostras sintéticas de alta qualidade para essas categorias pode fornecer ao modelo mais exemplos para aprender, melhorando seu desempenho.
- Modelos Hierárquicos: Quando existe uma relação natural hierárquica entre categorias (e.g., “Eletrônicos” → “Smartphones” → “Acessórios para Smartphones”), a investigação de modelos hierárquicos pode ser benéfica. Essa abordagem permite que o modelo aprenda a classificar em múltiplos níveis de abstração, potencialmente reduzindo confusões e aproveitando a estrutura inerente aos dados.
- Integração de Sinais Contextuais e Metadados: Para lidar com ambiguidades e consultas de cauda longa (*long-tail queries*), a integração de sinais adicionais é crucial. Isso inclui o histórico de navegação e compra do usuário, atributos detalhados do produto (e.g., marca, preço, especificações) e dados de cliques em resultados anteriores. Esses metadados enriquecem o sinal de entrada, fornecendo contexto adicional que o modelo pode usar para refinar suas previsões.
- Ensembles e Calibração de Probabilidades: O uso de ensembles de modelos ou a calibração de probabilidades de saída podem ser explorados para ajustar os trade-offs entre precisão e

recall em classes críticas, onde a minimização de falsos positivos ou falsos negativos é de particular importância.

- **Validação em Ambiente de Produção:** Embora os resultados de validação sejam promissores, a transição para a produção requer testes online (A/B testing) para medir o impacto real do classificador em métricas de negócio, como taxas de conversão, tempo de busca e satisfação do cliente.
- **Exploração de Modelos Mais Leves e Técnicas de Destilação:** Reconhecendo o elevado custo computacional de modelos como o BERT, futuras pesquisas podem investigar o emprego de modelos mais leves ou a aplicação de técnicas de destilação de conhecimento (knowledge distillation). Estas técnicas permitem transferir o conhecimento de um modelo grande e complexo (professor) para um modelo menor e mais eficiente (aluno), mantendo um desempenho comparável com menor custo computacional.

O classificador de intenções baseado em *Transformer* demonstrou um desempenho elevado e uma robustez notável no conjunto de dados analisado. A implementação das melhorias propostas e a exploração das direções futuras têm o potencial de aumentar ainda mais sua utilidade prática e consolidar sua contribuição para a área de busca semântica em língua portuguesa.

## 5 CONCLUSÃO

Este estudo confirmou a viabilidade e a notável eficácia da aplicação de modelos BERT pré-treinados em português para a classificação de intenções de busca de usuários em plataformas de comércio eletrônico. O modelo desenvolvido, ajustado sobre um corpus específico de consultas em português brasileiro, demonstrou um desempenho excepcional no conjunto de avaliação, atingindo uma acurácia de 96,43% e um F1-score médio de aproximadamente 0,96. Tais métricas, aliadas à estabilidade observada durante o processo de treinamento e à alta capacidade discriminatória entre a maioria das 92 categorias anotadas, validam robustamente a aplicabilidade da transferência de aprendizado em um cenário realista, caracterizado por consultas curtas e frequentemente ruidosas.

A análise aprofundada por classe e a inspeção das matrizes de confusão revelaram que as confusões residuais concentram-se predominantemente em categorias semanticamente próximas ou naquelas com menor suporte amostral. Este achado não apenas delimita as áreas para intervenção futura, como o refinamento do esquema de rotulagem, a aplicação de técnicas de *oversampling* ou *data augmentation*, e o tratamento hierárquico de categorias, mas também reforça a complexidade inerente à interpretação de linguagem natural em contextos ambíguos.

Em termos práticos, o trabalho culminou na produção de dois artefatos de valor significativo: um conjunto de dados rotulado de alta qualidade, específico para consultas de e-commerce em português brasileiro, e um modelo funcional. Estes artefatos abrem caminho para a integração direta em sistemas de busca e recomendação, permitindo o mapeamento automático de consultas a categorias de produtos e serviços e, consequentemente, o acionamento de estratégias de personalização mais eficazes. A capacidade do modelo em capturar as nuances linguísticas do português brasileiro representa um avanço crucial para aprimorar a experiência do usuário, reduzir o abandono de busca e otimizar as taxas de conversão no mercado lusófono.

Do ponto de vista metodológico, a adoção de práticas como validação cruzada estratificada, a fixação de sementes aleatórias e a preservação sistemática de todos os artefatos experimentais asseguram a reproduzibilidade do estudo e facilitam sua replicação e expansão. Este estudo, portanto, contribui tanto empiricamente, ao fornecer uma solução de alto desempenho para um problema relevante, quanto metodologicamente, ao consolidar diretrizes para a pesquisa em busca semântica em língua portuguesa, demonstrando que modelos monolíngues bem ajustados são ferramentas poderosas para a compreensão da intenção do usuário em ambientes digitais.

## REFERÊNCIAS

BORGES, Hilda Biazatti. IMPLEMENTAÇÃO DE BUSCA SEMÂNTICA EM E-COMMERCES COMO APOIO A PESQUISAS INTELIGENTES. Cachoeiro de Itapemirim: INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO ESPÍRITO SANTO, 2024.

CASELI, Helena; FREITAS, Cláudia; VIOLA, Roberta. Processamento de Linguagem Natural. In: SILVA, Ticiana et al. (Orgs.). Minicursos do SBBD 2022. 1. ed. [S.l.]: SBC, 2022. p. 1–25.

DEVLIN, Jacob et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv, , 24 maio 2019. Disponível em: <<http://arxiv.org/abs/1810.04805>>. Acesso em: 14 set. 2025

HAMMES, Luiz Otávio Alves; FREITAS, Larissa Astrogildo De. Utilizando BERTimbau para a Classificação de Emoções em Português. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA. Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2021). Brasil: Sociedade Brasileira de Computação, 29 nov. 2021. Disponível em: <<https://sol.sbc.org.br/index.php/stil/article/view/17784>>. Acesso em: 14 set. 2025

KHURANA, Diksha et al. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, v. 82, n. 3, p. 3713–3744, jan. 2023.

KIM, Hiun et al. Intent-based Product Collections for E-commerce using Pretrained Language Models. arXiv, , 15 out. 2021. Disponível em: <<http://arxiv.org/abs/2110.08241>>. Acesso em: 14 set. 2025

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 28 maio 2015.

PAN, Sinno Jialin; YANG, Qiang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, out. 2010.

PEETERS, Ralph; BIZER, Christian. Supervised Contrastive Learning for Product Matching. In: WWW '22: THE ACM WEB CONFERENCE 2022. Companion Proceedings of the Web Conference 2022. Virtual Event, Lyon France: ACM, 25 abr. 2022. Disponível em: <<https://dl.acm.org/doi/10.1145/3487553.3524254>>. Acesso em: 14 set. 2025

SCHMIDHUBER, Juergen. Deep Learning in Neural Networks: An Overview. *Neural Networks*, v. 61, p. 85–117, jan. 2015.

SPECIA, Lucia; RINO, Lucia Helena Machado. O desenvolvimento de um léxico para a geração de estruturas conceituais UNL. [S.l.: S.n.].

VASWANI, Ashish et al. Attention Is All You Need. arXiv, , 2 ago. 2023. Disponível em: <<http://arxiv.org/abs/1706.03762>>. Acesso em: 14 set. 2025

XIE, Yubo et al. Natural Language Processing with Commonsense Knowledge: A Survey. arXiv, , 13 set. 2024. Disponível em: <<http://arxiv.org/abs/2108.04674>>. Acesso em: 14 set. 2025



YULIANTON, Heribertus; CANDRA NOOR SANTI, Rina. Product Matching using Sentence-BERT: A Deep Learning Approach to E-Commerce Product Deduplication. *Engineering and Technology Journal*, v. 09, n. 12, 31 dez. 2024.

ZHANG, Han et al. Towards Personalized and Semantic Retrieval: An End-to-End Solution for E-commerce Search via Embedding Learning. *arXiv*, , 5 jun. 2020. Disponível em: <<http://arxiv.org/abs/2006.02282>>. Acesso em: 14 set. 2025