

Análise de Intenção em Conjuntos de Dados Pequenos: Comparação do Desempenho do Naive Bayes e do BERT

Rúbia M. Pereira,¹ Cassius T. Scarpin,² Pablo D. Valle³
UFPR, Curitiba, PR

Resumo. A análise de intenção textual tem se tornado uma área de grande interesse, especialmente em setores especializados, como o de petróleo e gás, onde a precisão na interpretação de textos técnicos é fundamental. Este trabalho compara o desempenho dos modelos Naive Bayes e BERT na classificação de intenções em dados específicos desse setor. O modelo BERT (*Bidirectional Encoder Representations from Transformers*), por ser baseado em aprendizado profundo e pré-treinado em grandes volumes de dados, apresenta vantagens na captura de dependências contextuais complexas. Já o Naive Bayes, por sua abordagem probabilística, oferece uma solução mais leve e eficiente computacionalmente. A partir de experimentos realizados sobre um conjunto de dados do setor, analisamos métricas como acurácia, precisão e *recall* para avaliar a eficácia de cada abordagem. Os resultados demonstram que o modelo BERT supera o Naive Bayes em termos de desempenho, especialmente em cenários com maior ambiguidade linguística, embora o Naive Bayes ainda se apresente como uma alternativa viável para aplicações com restrições computacionais.

Palavras-chave. Naive Bayes, BERT, Análise de Intenção, Óleo e Gás

1 Introdução

A identificação de intenções em textos é uma etapa relevante em diversas aplicações de Processamento de Linguagem Natural (PLN), como chatbots e sistemas de recomendação, possibilitando que o sistema interprete as solicitações dos usuários e forneça respostas ou ações adequadas [8].

A classificação correta das intenções dos usuários contribui para aprimorar as respostas e apoiar a tomada de decisões [12]. No setor de petróleo e gás, os modelos de PLN enfrentam desafios adicionais devido à disponibilidade limitada de dados e à natureza específica das informações [6].

Diferentes técnicas podem ser aplicadas nessa tarefa, desde abordagens estatísticas, como o Naive Bayes (NB), até modelos baseados em aprendizado profundo, como o BERT (*Bidirectional Encoder Representations from Transformers*). Enquanto o NB considera independência entre as palavras e apresenta menor custo computacional, o BERT permite capturar relações contextuais mais complexas nos textos.

Este estudo tem como propósito comparar o desempenho desses dois modelos na categorização de intenções em um conjunto de dados pequeno do setor de petróleo e gás, analisando sua eficácia tanto em um conjunto de teste quanto em um processo de validação cruzada.

O restante do artigo está estruturado da seguinte forma: a Seção 2 apresenta uma revisão da literatura sobre análise de intenções textuais, com foco em técnicas de PLN aplicadas ao setor de petróleo e gás. A Seção 3 descreve a metodologia adotada para comparar os modelos NB e BERT, incluindo o design do sistema e os dados utilizados. Na Seção 4, são apresentados os resultados

¹ru.biapereira@hotmail.com

²cassiusts@ufpr.br

³pablo.valle@ufpr.br

obtidos na avaliação dos modelos. Por fim, a Seção 5 discute as principais descobertas e apresenta as conclusões do estudo.

2 Referencial Teórico

A previsão de preços de petróleo envolve diversos fatores, como políticas econômicas, geopolítica e oferta e demanda. Um *chatbot* inteligente pode apoiar *stakeholders* na obtenção de previsões mais precisas, utilizando a análise de intenção para interpretar as perguntas dos usuários [7]. Por exemplo, quando um usuário questiona sobre a previsão do preço do produto XX para os próximos 180 dias, o *chatbot* deve identificar a intenção da consulta para fornecer uma resposta adequada, como uma previsão de aumento de preço.

A análise de intenção é essencial para compreender as consultas dos usuários e classificar as mensagens de forma a gerar respostas apropriadas. Modelos de PLN, como o BERT, são usados para identificar as intenções nas consultas, considerando o contexto das mensagens. Entre os métodos de classificação de intenção, destacam-se técnicas de *Machine Learning*, como o NB, e modelos baseados em *transformers*, como o BERT [7].

O modelo NB é um algoritmo de classificação baseado no Teorema de Bayes, que assume a independência entre as características dos dados. Apesar dessa suposição frequentemente não ser realista, o modelo é eficiente, especialmente em tarefas de classificação de texto, como análise de sentimentos e categorização de intenções. Sua implementação é rápida e demanda poucos recursos computacionais, o que o torna adequado para cenários com dados limitados ou necessidade de adaptação rápida [11].

O NB apresenta algumas vantagens, como a capacidade de obter bons resultados com conjuntos de dados pequenos e a possibilidade de realizar classificação multiclasse, sem a necessidade de grandes volumes de dados para treinamento. Por outro lado, ele tem limitações, como a inadequação para conjuntos de dados muito pequenos e desempenho comprometido quando as características não são independentes [10]. Alguns estudos, como o de [10], indicam que o NB pode ser eficaz com amostras pequenas, mas essa vantagem depende da independência das características. Quando essa suposição não é atendida, o desempenho do modelo pode ser prejudicado, como observado por [9]. Portanto, a aplicabilidade do NB em conjuntos de dados pequenos depende das características do problema.

O artigo de [11] compara os algoritmos Multinomial NB e Regressão Logística para a tarefa de classificação de intenções em *chatbots*. A pesquisa avaliou a precisão e eficiência dos métodos na identificação de intenções em textos dos usuários, indicando que a Regressão Logística teve um desempenho superior, oferecendo maior precisão. O estudo ressaltou a importância da escolha do modelo adequado para aumentar a eficácia dos sistemas de *chatbot* em identificar e responder intenções corretamente.

[1] desenvolveram o "HSchatbot" para prever as intenções das consultas dos alunos do ensino médio, relacionadas à escolha de carreira. Usando os classificadores Multinomial NB e *Random Forest* com técnicas de extração de características, os resultados mostraram que o *Random Forest* superou o NB em várias métricas, alcançando uma precisão superior a 90%. No entanto, o desempenho do NB foi melhor com o *CountVectorizer* do que com TF-IDF, sugerindo a necessidade de mais pesquisas sobre os fatores que afetam o desempenho do modelo.

O BERT é um modelo de linguagem baseado na arquitetura *Transformer* [13]. Desenvolvido por [4], o BERT foi projetado para entender o contexto completo de uma palavra em uma frase, processando simultaneamente as informações à esquerda e à direita da palavra, o que permite uma contextualização bidirecional. A arquitetura *Transformer* utiliza múltiplas camadas de codificadores e um mecanismo de atenção multi-cabeça, que atribui pesos diferentes às palavras em uma sequência de texto, dependendo da sua relevância contextual [4].

O BERT oferece vantagens, como uma melhor compreensão contextual, devido à sua abordagem bidirecional, e um bom desempenho em diversas tarefas de PLN, impulsionado pela capacidade de generalização adquirida através do pré-treinamento em grandes volumes de texto [4]. Ele pode ser facilmente adaptado a diversas tarefas, como resposta a perguntas e inferência de linguagem, com a adição de uma camada de saída, sem a necessidade de grandes modificações na arquitetura. Além disso, o BERT é eficaz na representação contextualizada de frases, sendo utilizado em tarefas como classificação de intenções e preenchimento de slots, por meio de um ajuste fino, similar ao aplicado em outras tarefas de PLN [3]. O pré-treinamento do BERT resulta em uma representação vetorial abrangente do texto, o que contribui para uma análise mais precisa das intenções e melhora a acurácia na identificação delas [13].

O estudo de [5] investigou a classificação de intenções em um pequeno conjunto de dados, utilizando os modelos BERT e LSTM. Com 150 classes de intenção e 100 exemplos de treinamento por classe, a pesquisa concluiu que o LSTM pode ser mais eficiente em conjuntos pequenos, alcançando maior precisão e exigindo menos tempo de ajuste do modelo. Esse estudo indicou que, apesar das vantagens do BERT, o LSTM pode ser uma alternativa eficaz em tarefas com dados limitados.

No artigo de [2], foi desenvolvido um modelo de classificação de intenções utilizando BERT, com o objetivo de categorizar perguntas dos usuários sobre veículos, com base em dados extraídos de Manuais do Proprietário. O modelo foi treinado com 254.412 registros e 199 categorias, alcançando uma precisão geral de 98,21%. Os resultados mostraram que, mesmo com erros de digitação ou gramática, a abordagem foi eficaz em classificar texto real, evidenciando a utilidade de agentes conversacionais na interpretação de intenções.

O artigo de [3] descreve um modelo de classificação de intenções e preenchimento de slots baseado em BERT, que superou modelos tradicionais em termos de precisão e capacidade de generalização. Avaliado nos conjuntos de dados ATIS e Snips, o modelo demonstrou alto desempenho, evidenciando melhorias significativas em comparação com abordagens anteriores. O estudo sugere que futuras pesquisas podem explorar conjuntos de dados maiores e integrar conhecimento externo para otimizar ainda mais os resultados do BERT.

3 Metodologia

Neste estudo, o conjunto de dados utilizado para a classificação das intenções inclui declarações coletadas para o desenvolvimento de um chatbot, que vão desde perguntas relacionadas à previsão de preço até perguntas aleatórias.

No total, são 450 perguntas, distribuídas igualmente entre as 6 categorias, com 75 perguntas em cada uma. Um exemplo das perguntas contidas na base de dados e suas respectivas categorias são apresentados na Tabela 1.

Tabela 1: Amostra do Banco de Dados e suas respectivas Categorias.

Perguntas	Categorias
"What role does supply and demand play in predicting oil and gas prices?"	Univariate Analysis
"Which factors influence oil price trends over time?"	Price Trends
"What are the oil price projections for the next ten years?"	Forecast
"What are the economic risks that could affect oil price forecasts?"	Multivariate Analysis
"What's your all-time favorite movie and why?"	Others
"What are the impacts of environmental policies on oil and gas price forecasts?"	Sentiment Analysis

Para analisar a intenção das perguntas, utilizou-se os modelos NB Multi e BERT. No caso do modelo Naive Bayes, os textos foram convertidos em vetores utilizando a técnica TF-IDF, que avalia a relevância de cada termo no contexto do conjunto de documentos. O TF-IDF é uma técnica comum de pré-processamento de texto que atribui pesos a cada palavra com base em sua frequência no documento e sua raridade no corpus global, o que ajuda a destacar as palavras mais significativas para a tarefa de classificação. O modelo Naive Bayes utilizado é o MultinomialNB, um algoritmo simples, mas eficiente, que assume a independência entre as características (termos) e é adequado para tarefas de classificação de texto. Para analisar a intenção das perguntas, utilizou-se os modelos NB Multi e BERT. No caso do modelo NB, os textos foram convertidos em vetores utilizando a técnica TF-IDF, que avalia a relevância de cada termo no contexto do conjunto de documentos. O TF-IDF é uma técnica comum de pré-processamento de texto que atribui pesos a cada palavra com base em sua frequência no documento e sua raridade no *corpus global*, o que ajuda a destacar as palavras mais significativas para a tarefa de classificação. O modelo NB utilizado é o MultinomialNB, um algoritmo simples, mas eficiente, que assume a independência entre as características (termos) e é adequado para tarefas de classificação de texto.

No caso do modelo BERT, os textos foram tokenizados utilizando o *BertTokenizer*, que converte as palavras em tokens de acordo com o vocabulário do modelo. Para limitar o comprimento das entradas, foi utilizado um tamanho máximo de 64 tokens por texto. O modelo BERT utilizado é o *bert-base-uncased*, uma versão pré-treinada do modelo BERT, que não diferencia maiúsculas de minúsculas e foi ajustado especificamente para a tarefa de classificação de textos com duas classes. O BERT é capaz de capturar o contexto das palavras de maneira bidirecional, o que o torna muito eficaz para tarefas de compreensão de linguagem.

Para avaliar os modelos, foram utilizados duas métricas: a acurácia e a validação cruzada. Para a acurácia, a base de dados foi previamente embaralhada e dividida em 50 perguntas para treino, 10 para validação e 15 para teste, e foi calculada através da Equação (1).

$$\text{Acurácia} = \frac{\text{Número de Previsões Corretas}}{\text{Número Total de Previsões}}. \quad (1)$$

Já para a métrica de validação cruzada, foi utilizado um K-fold igual a 5 garantindo que cada amostra fosse usada tanto para treino quanto para validação, sendo calculada através da Equação (2).

$$\text{Média das Acurácias} = \frac{1}{K} \sum_{n=1}^K \text{Acurácia}_i. \quad (2)$$

Em que a Acurácia_i é a acurácia obtida na iteração i e K é o número de folds.

4 Resultados

A presente pesquisa teve como objetivo comparar o desempenho dos modelos Naive Bayes e BERT quanto à acurácia e ao tempo computacional, empregando dois métodos de avaliação: acurácia geral e validação cruzada *K-fold*. No teste geral, o modelo NB apresentou acurácia de 65,56%, enquanto o BERT alcançou 70,00%, conforme demonstrado na Tabela 2. Tais resultados indicam um desempenho superior do BERT na tarefa de classificação de intenções.

Em relação ao tempo computacional, observa-se uma diferença expressiva entre os modelos: o BERT demandou 8274,243 segundos, ao passo que o NB consumiu apenas 13,3 segundos. Essa discrepância reflete a maior complexidade do BERT, cuja arquitetura baseada em *Transformers* exige um volume elevado de recursos computacionais, especialmente nas etapas de pré-treinamento e ajuste fino.

Tabela 2: Resultados dos Desempenhos e Tempo Computacional.

	Acurácia		Validação Cruzada	
	Porcentagem	(s)	Porcentagem	(s)
NB	65,56%	12,523	67,33%	13,3
BERT	70,00%	1336,947	78,89%	8274,243

Na avaliação por validação cruzada *K-fold*, que busca aferir a robustez dos modelos diante de diferentes partições dos dados, o modelo BERT também apresentou resultados superiores. Seus valores variaram entre 72,22% e 85,56%, com média de 78,89%, enquanto o NB oscilou entre 57,69% e 72,22%, com média de 67,33%. Esses dados, apresentados na Tabela 3, evidenciam que o BERT manteve desempenho consistente e elevado mesmo em diferentes subconjuntos da base, o que reforça sua estabilidade e capacidade de generalização.

Tabela 3: Resultados da Validação Cruzada K-Fold.

	NB (%)	BERT (%)
Fold 1	72,22	85,56
Fold 2	61,11	73,33
Fold 3	63,33	83,33
Fold 4	57,69	80,00
Fold 5	71,11	72,22
Média	67,33	78,89

Para melhor compreensão do processo de aprendizado do BERT, foi analisada a evolução da função de perda (*loss*) ao longo das épocas de treinamento. Tal análise é relevante para verificar se o modelo está, de fato, aprendendo e ajustando seus parâmetros de maneira eficiente. Na primeira época, a perda registrada foi de 1,7884, indicando que o modelo cometia erros significativos. Na segunda época, o valor reduziu-se para 1,1107, o que reflete um avanço considerável no ajuste dos parâmetros. Já na terceira época, a perda atingiu 1,0698, demonstrando que o modelo continuava em processo de aprendizado, embora com ganhos marginais. Esse comportamento sugere que o modelo ainda não havia atingido total convergência, apontando para a possibilidade de melhorias adicionais com um número maior de épocas ou ajustes nos hiperparâmetros.

Dessa forma, observa-se que o modelo BERT obteve desempenho superior ao NB tanto na acurácia geral quanto na validação cruzada, apresentando médias de 70,00% e 78,89%, respectivamente. Contudo, tal desempenho vem acompanhado de um custo computacional significativamente mais elevado. Assim, a escolha do modelo deve considerar o equilíbrio entre a exigência de precisão e as restrições de tempo e recursos disponíveis. O BERT mostra-se mais apropriado em contextos nos quais a acurácia é um fator crítico, enquanto o NB se configura como uma alternativa viável em cenários com limitações operacionais.

Além dos resultados principais, outras considerações podem ser destacadas com base nas análises realizadas. A menor variabilidade nos resultados do BERT entre os diferentes folds da validação cruzada evidencia sua maior estabilidade e menor sensibilidade às variações da distribuição dos dados, fortalecendo sua capacidade de generalização. A relação entre a complexidade do modelo e o desempenho preditivo torna-se clara: o BERT, embora mais oneroso em termos computacionais, oferece resultados mais robustos. Esse fator é essencial na tomada de decisão quanto ao investimento em infraestrutura computacional em aplicações sensíveis.

O desempenho inferior do NB em determinados folds evidencia as limitações de modelos probabilísticos simples diante da complexidade inerente às tarefas de Processamento de Linguagem Natural (PLN), como a análise de intenção textual. Ademais, a análise da função de perda do

BERT sugere que o modelo ainda apresentava capacidade de aprendizado após três épocas, o que indica potencial para otimizações adicionais. Por fim, o contraste entre desempenho e tempo computacional reforça a necessidade de alinhar a escolha do modelo ao contexto da aplicação. Em situações que demandam maior acurácia, o BERT representa a escolha mais adequada. Em contrapartida, o NB pode ser preferido quando se busca maior agilidade e menor custo computacional.

5 Considerações Finais

Neste trabalho, foi realizada uma comparação entre os modelos Naive Bayes e BERT para a classificação de intenção textual em um contexto especializado do setor de petróleo e gás. Os resultados demonstraram que o modelo BERT apresenta um desempenho superior em termos de acurácia, precisão e *recall*, especialmente em cenários com maior ambiguidade linguística. Sua capacidade de capturar dependências contextuais mais complexas torna-o uma alternativa robusta para aplicações que exigem alta precisão na interpretação de textos técnicos.

Por outro lado, o modelo Naive Bayes se mostrou uma abordagem mais simples e eficiente do ponto de vista computacional, sendo uma opção viável para cenários em que há restrições de processamento ou necessidade de implementação rápida. Essa análise reforça a importância da escolha do modelo considerando o contexto da aplicação e os recursos disponíveis.

Como trabalhos futuros, sugere-se a exploração de técnicas de *fine-tuning* no modelo BERT para melhor adaptação ao domínio específico, além da experimentação com outros modelos baseados em redes neurais, como *transformers* especializados, para verificar possíveis ganhos de desempenho. Também seria interessante avaliar estratégias híbridas que combinem a eficiência do Naive Bayes com a capacidade contextual do BERT, visando otimizar os resultados obtidos.

Referências

- [1] S. Assayed, K. Shaalan e M. Alkhatib. “A chatbot intent classifier for supporting high school students”. Em: **EAI Endorsed Transactions on Scalable Information Systems** (2023). DOI: <https://ssrn.com/abstract=4397536>.
- [2] S. Chakraborty, K. Y. Ohm, H. Jeon, D. H. Kim e H. J. Jin. “Intent classification of users conversation using BERT for conversational dialogue system”. Em: **2023 25th International Conference on Advanced Communication Technology (ICACT)**. 2023, pp. 65–69. DOI: <https://doi.org/10.23919/ICACT56868.2023.10079261>.
- [3] Q. Chen, Z. Zhuo e W. Wang. “Bert for joint intent classification and slot filling”. Em: **CoRR** (2019), pp. 65–69. DOI: <https://doi.org/10.48550/arXiv.1902.10909>.
- [4] J. Devlin, M-W Chang, K Lee e K Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. Em: **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**. 2019, pp. 4171–4186. DOI: <https://doi.org/10.18653/v1/N19-1423>.
- [5] A. Ezen-Can. “A Comparison of LSTM and BERT for Small Corpus”. Em: **arXiv preprint** (2020), pp. 4171–4186. DOI: <https://doi.org/10.48550/arXiv.2009.05451>.
- [6] D. D. S. M. Gomes, F. C. Cordeiro, B. S. Consoli, N. L. Santos, V. P. Moreira, R. Vieira, S. Moraes e A. G. Evsukoff. “Portuguese word embeddings for the oil and gas industry: Development and evaluation”. Em: **Computers in Industry** (2021), p. 103347. DOI: <https://doi.org/10.1016/j.compind.2020.103347>.

- [7] M. Hasan, M. J. I. Basher e M. T. R Shawon. “Bengali intent classification with generative adversarial Bert”. Em: **2023 26th International Conference on Computer and Information Technology (ICCIT)**. 2023, pp. 1–6. DOI: <https://doi.org/10.1109/ICCIT60459.2023.10440989>.
- [8] X. Huang, T. Ma, L. Jia, Y. Zhang, H. Rong e N. Alnabhan. “An effective multimodal representation and fusion method for multimodal intent recognition”. Em: **Neurocomputing** (2023), p. 126373. DOI: <https://doi.org/10.1016/j.neucom.2023.126373>.
- [9] N. Milosevic, A. Dehghantanha e K. K. R. Choo. “Machine learning aided Android malware classification”. Em: **Computers Electrical Engineering** (2017), pp. 266–274. DOI: <https://doi.org/10.1016/j.compeleceng.2017.02.013>.
- [10] A. Palanivinayagam, C. Z. El-Bayeh e R. Damaševičius. “Twenty years of machine-learning-based text classification: A systematic review”. Em: **Algorithms** (2023), p. 236. DOI: <https://doi.org/10.3390/a16050236>.
- [11] M. Y. H. Setyawan, R. M. Awangga e S. R. Efendi. “Comparison of multinomial naive bayes algorithm and logistic regression for intent classification in chatbot”. Em: **In 2018 International Conference on Applied Engineering (ICAE)**. 2018, pp. 1–5. DOI: [10.1109/INCAE.2018.8579372](https://doi.org/10.1109/INCAE.2018.8579372).
- [12] A. Singh e V. Jia T.and Nalagatla. “Generative AI enabled conversational Chatbot for drilling and production analytics”. Em: **Abu Dhabi International Petroleum Exhibition and Conference**. 2023. DOI: <https://doi.org/10.2118/216267-MS>.
- [13] T. Wu, M. Wang, Y. Xi e Z. Zhao. “Intent recognition model based on sequential information and sentence features”. Em: **Neurocomputing** (2024), p. 127054. DOI: <https://doi.org/10.1016/j.neucom.2023.127054>.