

# Aplicação de Aprendizado de Máquina para Classificação de Produtos no Processo de e-procurement

Angelo Miguel Marin  
UNIVALI – Universidade do  
Vale do Itajaí, SC, Brasil  
angelo.marin@outplan.com.br

Ivan Carlos dos Santos  
UNIVALI – Universidade do Vale do  
Itajaí, SC, Brasil  
ivancsantos@hotmail.com.br

Rudimar Dazzi  
UNIVALI – Universidade do  
Vale do Itajaí, SC, Brasil  
rudimar@univali.br

Anita Fernandes  
UNIVALI – Universidade do Vale  
do Itajaí, SC, Brasil  
anita.fernandes@univali.br

## ABSTRACT

Currently, the quotation and/or planning systems shopping from most companies, which are classified as e-procurement, it does not follow a standardization. Many work with free writing because it is very difficult to establish all items and possible combinations of these items in one list, to be selected by the buyer and associated with vendor description. Although there are codes like CNAE (National Classification of Economic Activities) and the NCM (Mercosur Common Number), a few companies use them to characterize the products. work is done manually most of the time, which makes the percentage of errors large, and the inconsistency between the orders is also high. Therefore, it is necessary to automate the process, to make the match between the requested products by the customer and those offered by the most assertive suppliers.

## KEYWORDS

e-procurement, Natural Language Processing, Machine Learning.

## 1 INTRODUÇÃO

Cada vez mais as empresas precisam melhorar sua performance em todas as áreas, dado o ambiente de alta competitividade em que estão inseridas [1].

Uma das áreas que gera impacto mais direto no resultado de uma empresa e a área de compras, afinal, como na maioria das vezes o valor de venda e definido pelo que o mercado está disposto a pagar, quanto menos se gastar na compra de insumos melhor pode ser o resultado [2]. Obviamente, outros fatores como processos e controle de qualidade influenciam diretamente no resultado, porém, este trabalho foca no problema do compras. Neste contexto, tem-se o *e-procurement*. que pode ser compreendido como um sistema ou ambiente digital para contato, negociação e compra de suprimentos ou serviços. Este conceito refere-se a tipos de sistemas online que

aproxime contratantes e fornecedores. Dessa forma, e mais fácil localizar empresas parceiras, que comercializam produtos necessários para sua operação [3]. Para que isto funcione, há a necessidade de padronização dos itens dos compradores e dos fornecedores.

Uma das dificuldades mais claras do departamento de compras é classificar rapidamente, e corretamente, as demandas das várias outras áreas da empresa e a partir daí selecionar fornecedores capacitados e com um bom preço para que possam atender essas necessidades. A quantidade de itens diferentes comprados por uma empresa mensalmente pode chegar a alguns milhares, dependendo do negócio, até a dezenas de milhares

Para facilitar esse trabalho é necessário agrupar esse item em “categorias de compra”, permitindo assim que um processo de compra possa tratar algumas dezenas de itens semelhantes de uma só vez e, com isso, gerar ganho operacional e até financeiro já que fornecedores tendem a oferecer melhores preços para maior volume de compra. Atualmente os sistemas de ERP (*Enterprise Resourcing Planning* - ou Software Gestão Integrada) e, mais especificamente, os sistemas de *e-procurement* [3], permitem fazer essa classificação associando, por exemplo, CANETA e PAPEL a uma categoria de material de expediente, BOTAS e LUVAS a EPI e, quem sabe, uma LUVA DE PVC a material hidráulico.

Mas tudo isso e feito de forma manual, por um ser humano que interpreta e associa cada item. Isto além de extremamente trabalhoso, possibilita muitas falhas. Não é incomum encontrar erros “grotescos” de classificação nesse cenário. Por conta disso, este trabalho apresenta a proposta de utilização de algoritmos de Processamento de Linguagem Natural [4] e Aprendizado de Máquina [5], para facilitar essa classificação.

Tendo um bom modelo de aprendizado de máquina para entender como classificar um novo item a partir de um texto livre escrito pelo solicitante, facilita todo esse trabalho do

comprador, gerando um ganho considerável de tempo e qualidade. A partir daí, bons sistemas de *e-procurement* poderão fazer o restante do trabalho ajudando o comprador a encontrar as melhores oportunidades de compra.

## 2 METODOLOGIA

Para a execução desta pesquisa, foram necessárias cinco etapas: (i) revisão bibliográfica; (ii) revisão sistemática da literatura; (iii) preparação dos dados da pesquisa; (iv) seleção, implementação, e aplicação dos algoritmos; (v) análise dos resultados.

A revisão bibliográfica focou nas técnicas de processamento de linguagem natural, e nos algoritmos de classificação que poderiam ser aplicados, como por exemplo, KNN, SVM, Redes Neurais Artificiais, dentre outros. Na revisão sistemática da literatura buscou-se trabalhos correlatos para entender as tendências da área, como por exemplo, o de Lima et al [6], que apresenta o uso dos modelos BERT [7] e BERT Multilingual [8], para classificar produtos, considerando a classificação única de produtos do Mercosul.

Os dados utilizados nesta pesquisa foram oriundos de uma base simulada de uma empresa parceira do projeto. Ela contém 3623 produtos com suas respectivas descrições. Para o escopo inicial, focou-se em classificar alguns itens:

- Canetas, como material de expediente: qualquer caneta como esferográfica, marca texto, caneta para quadro branco etc. Descartando, por exemplo, maÇANETAS, ou ferramentas do tipo caneta.
- Papéis, como material de expediente: papéis como A4, A3, papel carta etc. Descartando, por exemplo, PAPELÃO, utilizado em embalagens.
- Embalagens de papel: ao contrário de papéis que são materiais de expediente, se buscará identificar PAPELÕES utilizados para embalagens.
- Luvas, como EPI: aceitando, por exemplo, Luvas de Latex, Luvas de Proteção, Luvas isolantes e desprezando Luvas de PVC, Luvas redutoras etc., que são materiais hidráulicos.
- Luvas, como material hidráulico: o contrário das luvas de EPI, se buscará identificar aqui luvas que são materiais hidráulicos.

Realizou-se um tratamento inicial da base e retirou-se alguns registros incompletos, ficando com 3.427 registros. Analisando estes registros, obteve-se um conjunto de palavras mais frequentes, exibidas na Figura 1.

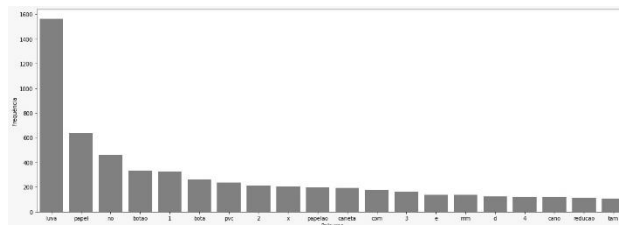


Figura 1: Palavras mais frequentes.

Para chegar a este gráfico, foi feito um pré-processamento, com remoção de *stopwords*, remoção de itens de produtos não desejados através de palavras específicas, tokenização, remoção de acentuação e padronização do texto em caixa baixa.

Em seguida classificou-se os registros nas cinco categorias: canetas, papel, embalagem, luvas como EPI e luvas como material hidráulico.

O próximo passo será implementar primeiramente os classificadores SVM e Redes Neurais Recorrentes para avaliar o desempenho. A ideia de começar por estes classificadores veio devido a estes serem bastante citados na literatura [9, 10, 11].

## 3 CONSIDERAÇÕES FINAIS

Entende-se que o resultado desse trabalho será de grande valia para facilitar o trabalho de compradores empresariais.

Uma vez consolidada como uma abordagem eficiente, ela vai preencher uma lacuna que existe hoje entre receber solicitação de compra e encontrar fornecedores. A partir daqui, utilizando boas ferramentas de *e-procurement*, que facilitam a comunicação entre compradores e fornecedores, os compradores podem alcançar uma boa base de fornecedores, também relacionados a essas categorias de compras, e receber propostas rapidamente.

## AGRADECIMENTOS

FAPESC – UNIEDU

## REFERÊNCIAS

- [1] Konyuhov, V.Y.; Gladkih, A.M.; Zott, R.S. Machine-building enterprise performance and quality improvement tools. IOP Conference Series: Materials Science and Engineering, Volume 1064, International Conference on Mechanical Engineering, Automation and Control Systems (MEACS 2020) 17th September 2020, Novosibirsk, Russian Federation. Acessado em 24/10/2022, disponível em: <https://iopscience.iop.org/article/10.1088/1757-899X/1064/1/012021/meta> [2] Brandon-Jones, A. E-procurement quality from an internal customer perspective: construct development, refinement, and replication using a mixed-methods approach, International Journal of Operations and Production Management, Vol. 37 No. 12, pp. 1741-1772, 2017 [3] Grande, M.M.; Olhê, E.A.; Rodello, I.A.; Gomes, E.J. E-Procurement: Concepts, models, and critical success factors. Rev. eletrôn. adm. (Porto Alegre) 22 (3) Sep-Dec 2016. Doi: 10.1590/1413-2311.06914.54830 [4]

Nasir, Muhammad Umar; Hussain, Muhammad Zahid; Arsalan, Ahmad; Septyanto, Arif Wicaksono. Natural Language Processing and Machine Linguistic Interpretation. 2021. 4 f. Riphah School Of Computing And Innovations, Riphah International University Lahore, Lahore, 2021. [5] Alpaydin, E. Introduction to Machine Learning. 3 rd ed.. MIT Press (MA), 2016. [6] de Lima, Roberta Rodrigues, Anita M. R. Fernandes, James Roberto Bombasar, Bruno Alves da Silva, Paul Crocker, and Valderi Reis Quietinho Leithardt. 2022. "An Empirical Comparison of Portuguese and Multilingual BERT Models for Auto-Classification of NCM Codes in International Trade" *Big Data and Cognitive Computing* 6, no. 1: 8. <https://doi.org/10.3390/bdcc6010008> [7] Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1 (Long and Short Papers). Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4171–4186. [8] Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv 2019, arXiv:1907.11692. [9] Fava, L.P.; Furtado, J.C.; Helfer, G.A.; Barbosa, J.L.V.; Beko, M.; Correia, S.D.; Leithardt, V.R.Q. A Multi-Start Algorithm for Solving the Capacitated Vehicle Routing Problem with Two-Dimensional Loading Constraints. *Symmetry* 2021, 13, 1697. [10] Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; et al. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv 2016, arXiv:1609.08144. [11] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2017; pp. 5998–6008.