

# Algoritmos de classificação de texto na automatização dos processos orçamentação

<https://doi.org/10.21814/uminho.ed.77.7>

**Luís J. Sousa<sup>1</sup>, João Poças Martins<sup>2</sup>,  
João Santos Baptista<sup>3</sup>, Luís Sanhudo<sup>4</sup>, Pedro Mêda<sup>5</sup>**

<sup>1</sup> Faculdade de Engenharia da Universidade do Porto, Porto, 0000-0002-0789-9368

<sup>2</sup> CONSTRUCT/GEQUALTEC, FEUP – DEC, Porto, 0000-0001-9878-3792

<sup>3</sup> Associated Laboratory for Energy, Transports and Aeronautics (PROA/LAETA),  
Faculty of Engineering, University of Porto, Porto, 0000-0002-8524-5503

<sup>4</sup> BUILT CoLAB – Collaborative Laboratory for the Future Built Environment,  
Porto, 0000-0002-2578-6981

<sup>5</sup> CONSTRUCT/GEQUALTEC, Instituto da Construção,  
Faculdade de Engenharia da Universidade do Porto, Porto, 0000-0003-4380-5530

## Resumo

As aplicações de gestão da Construção incluem dados relativos à duração de tarefas, orçamentação, qualidade, segurança em obra, entre outros tópicos. No caso específico da orçamentação, as empresas de Construção são obrigadas a avaliar o âmbito de cada tarefa, mapeando as expectativas do cliente (expressas no mapa de quantidade de trabalho) para uma base de dados interna de tarefas, recursos e custos. Esta avaliação é frequentemente realizada por técnicos dentro de restrições de tempo muito austeras, apesar de os resultados obtidos através desta avaliação serem fulcrais para a qualidade e competitividade das propostas emitidas, para além de serem contratualmente vinculativas.

Com o objetivo de melhorar o desempenho desta tarefa, a presente comunicação explora a possibilidade de automatizar esta avaliação manual utilizando algoritmos de classificação de texto. Assim, propõe-se um protocolo para revisão de literatura sobre este tópico utilizando o método PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*). É realizada uma análise preliminar da literatura recolhida, permitindo a definição de um *framework* para apoiar uma abordagem automatizada à orçamentação.

Embora a automatização total não seja um objetivo verosímil, nem desejado, a curto prazo, especialmente devido à falta de especificações de construção padrão em Portugal, os algoritmos de classificação de texto podem fornecer ferramentas úteis de apoio à decisão. Estes algoritmos requerem grandes volumes de dados, que podem ser obtidos através da sua utilização contínua, pelo que será necessário mais trabalho para desenvolver fluxos de operações abrangentes.

## 1. Introdução

O processo de concurso é uma fase crucial no projeto de construção para as empresas: a qualidade das propostas pode determinar a obtenção, ou não, de contratos de execução de obra, bem como a perda ou lucro com esse mesmo projeto. Custos adicionais podem originar problemas graves para todos os intervenientes, incluindo eventuais interrupções dos trabalhos por falta de recursos [1, 2]. Uma solução para mitigar este tipo de dificuldades é prever com precisão os custos reais de obra, ou seja, prever o custo real pode promover um melhor planeamento e evitar contratempos posteriores.

Este processo requer a análise cuidada de Mapas de Quantidades e Trabalhos (MQTs), requerendo a consideração simultânea de múltiplas variáveis como duração, custo, qualidade e segurança em obra, entre outros aspetos. Cada projeto é elaborado por técnicos especializados de acordo com diferentes pressupostos como restrições de tempo que tornam estes entregáveis suscetíveis a erros e omissões. Por este motivo, existe interesse no desenvolvimento de ferramentas de apoio à decisão que permitam a mitigação destes problemas, evitando que erros se perpetuem e acumulem ao longo do processo construtivo. O presente trabalho pretende contribuir para o desenvolvimento de uma ferramenta ou sistema com este mesmo objetivo, focando as empresas nacionais do setor da engenharia civil que pretendam desenvolver processos expeditos de planeamento e de orçamentação de trabalhos de construção. Os métodos propostos não dependem do estabelecimento prévio de formatos padrão para a classificação de trabalhos de construção, embora possam claramente tirar partido de situações em que estes *standards* existam (sejam estes ao nível empresarial ou nacional).

O objetivo do projeto de investigação é transformar os dados acumulados ao longo do tempo, a partir da aglomeração de MQTs, para um formato analítico, analisá-lo e automatizá-lo a partir de uma perspetiva *big data*. O sistema deverá suportar a previsão de custos de construção e prevenir erros através de técnicas de *Machine Learning* (ML), particularmente *Natural Language Processing* (NLP), e análise de dados utilizando informação de projetos passados como referência.

Segundo o relatório anual do Instituto dos Mercados Públicos, do Imobiliário e da Construção (IMPIC), relativo à contratação pública em 2019, a taxa de conclusão registada no caso das obras públicas foi de apenas 25,96%, respeitante a 2855 contratos (contratos celebrados com indicação do preço total efetivo) [3]. Devido a esta falta de informação é difícil determinar eventuais correlações entre o nível de detalhe da informação do concurso e a eficiência da execução, especialmente em relação ao cumprimento de custo e prazo.

Os dados apresentam os principais fatores de custo conceptuais que afetam a estimativa de custos com base na utilização de dados quantitativos históricos [4]. A falta de uma base de dados fidedigna, fiável e pública pode ser um grande obstáculo à

implementação de técnicas NLP em Portugal, pois estes procedimentos necessitam de uma quantidade avultada de dados para sustentar o seu adequado funcionamento.

Em Portugal, a Parque Escolar, (EPE) celebrou um contrato com o consórcio ProNIC que teve, entre outros, o objetivo de disponibilizar uma plataforma que contém uma estrutura de codificação de trabalhos da construção – plataforma ProNIC. Esta teve o intuito de melhorar a gestão e a qualidade dos processos de Construção, desde a fase de projeto até à fase de receção da obra. Através da criação de referências tendo como base as melhores práticas e especificações técnicas dos trabalhos de construção, é possível diminuir a quantidade de erros e omissões na elaboração e análise dos cadernos de encargos. A implementação da plataforma ProNIC em todas as obras públicas em Portugal poderia permitir a normalização dos formatos dos MQTs, de tal forma que simplificaria o desenvolvimento de ferramentas como a que se pretende criar neste projeto de investigação.

Por tudo isto, neste trabalho, procurou-se qual a melhor abordagem para desconstruir estes desafios na indústria da Construção Portuguesa.

Neste artigo expôs-se o protocolo para levar a cabo uma revisão sistemática utilizando o método de PRISMA [5] e propõe-se um *framework* que representa a visão introdutória para o modelo previsto.

## 2. Metodologia

Este protocolo obedece à lista de verificação PRISMA. Os seguintes subcapítulos ilustrarão a aplicação da metodologia PRISMA à investigação em curso.

### 2.1. Objetivos

A revisão bibliográfica deve suportar o desenvolvimento, treino e validação de um algoritmo de NLP para a automatização da categorização das tarefas de um orçamento a ser usado na indústria de construção e no contexto português. Para isso, é necessário conhecer quais os principais e mais recentes métodos desenvolvidos na área. Assim este protocolo define a metodologia para responder exclusivamente as seguintes perguntas:

- 1.º Quais são as principais técnicas utilizadas na implementação de ML e NLP na orçamentação de projetos de construção?
- 2.º A que tipo de tarefas do projeto de execução foram aplicadas essas técnicas?
- 3.º De que forma se obtiveram os dados de base dos algoritmos desenvolvidos?
- 4.º Quais os resultados e desempenho obtidos por estes algoritmos?
- 5.º Qual a relação entre a utilização destas técnicas e qualidade das propostas realizadas?

## 2.2. Critérios de elegibilidade

Para ser considerado elegível, um estudo tem de abordar pelo menos um dos tópicos referidos na lista de questões relevantes apresentadas nos Objetivos. Foram considerados casos de estudo, ensaios controlados, entre outros. Qualquer artigo com dados suficientes para medir a eficácia, identificar as técnicas, métodos de implementação e o significado do resultado. Todas as revisões e atas da conferência serão excluídas. Só foram incluídos artigos de 2018 a 2021 visto que o setor de ML está em grande evolução e com muita produção científica. Apenas artigos que apliquem técnicas de ML e NLP, no contexto da engenharia civil e na fase da orçamentação foram considerados. Finalmente é importante mencionar que apenas artigos escritos na língua inglesa foram incluídos.

## 2.3. Fontes de informação

As bases de dados eletrónicas *SCOPUS*, *Web of Science*, *Dimensions*, *IEEE* e *Journal Storage* foram utilizadas para responder aos objetivos definidos na Secção 2.1. Estas bases de dados são atualmente as mais relevantes no campo da engenharia e irão apoiar a identificação de informações pertinentes.

## 2.4. Estratégia de pesquisa

As palavras-chave consideradas para esta pesquisa bibliográfica, de maneira a estender os tópicos dos artigos aos objetivos definidos foram “*Machine learning*” e “*Natural language processing*” para definir as técnicas que se pretende aplicar. “*BIM*” e “*construction*” para abordar a área onde se planeia aplicar as referidas técnicas. E “*cost*” e “*tendering*” de maneira a especificar a parte do processo construtivo sobre a qual se tenciona debruçar.

Outras *keywords* foram adicionadas de modo a abranger o número mais alargado de artigos, resultando na seguinte *string* de pesquisa:

(“*Machine Learning*” OR “*Text Classification*” OR “*Natural Language Processing*” OR  
“*AI*” OR “*Artificial Intelligence*”)  
AND (“*BIM*” OR “*Building Information Modelling*” OR “*Construction*” OR “*Civil Engi-  
neering*”)  
AND (“*Budgeting*” OR “*Budget*” OR “*Tendering*” OR “*Specification\**” OR “*Cost*”)

Devido ao número limitado de operadores booleanos da base de dados *Journal Storage* foi utilizada a seguinte *string* de pesquisa para este caso específico:

(“*Machine Learning*” OR “*Text Classification*” OR “*Natural Language Processing*”)  
AND “*Construction*”  
AND (“*Budget*” OR “*Cost*” OR “*Tendering*”)

A pesquisa iniciou-se pela inserção de cada combinação de palavras-chave numa das bases de dados eletrónicas previamente descritas; esta 1ª pesquisa será livre de qualquer limitação adicional nesta fase. O número total de artigos foi registado para conclusões estatísticas da amostra. O acompanhamento de cada estudo a partir do número inicial de artigos e do número de artigos excluídos com cada limitação começa pela data, língua, área temática, e depois pela fonte.

Seguidamente, foi feita uma análise com base no título e resumo dos artigos de modo a permitir seleccionar de entre os sobrantes quais os que contribuíram para as respostas às perguntas definidas nos Objetivos.

A estratégia final de pesquisa passou por examinar as referências dos artigos recolhidos para verificar a existência de qualquer estudo relevante que possa ser incluído na revisão.

## 2.5. Registos de estudo

### Gestão dos dados:

Os registos foram guardados num ficheiro Excel, sob a forma de tabela. Todos os artigos seleccionados foram exportados e analisados para remover duplicados. Todas as referências foram geridas utilizando o software *Endnote*.

### Processo de seleção:

À imagem da estratégia de pesquisa a seleção da literatura ocorreu em duas fases de triagem, com base nos critérios de elegibilidade definidos. Na primeira fase, os títulos e resumos foram rastreados, e na segunda fase, o texto integral de toda a literatura potencialmente relevante foi revisto. Em caso de dúvida sobre a análise inicial em torno do título e o rastreio do resumo, os artigos foram guardados para a fase seguinte. Na segunda fase, quando se verificou novamente dúvidas relativamente a inclusão de artigos, qualquer desacordo foi resolvido por consenso através da discussão entre dois revisores.

### Processo de Recolha de dados:

O número total de estudos analisados, incluindo os avaliados para critérios de exclusão e inclusão, foi determinado. Além disso, a razão para excluir estudos durante o processo de seleção foi documentada. Os dados qualitativos foram extraídos dos estudos eleitos e registados na revisão utilizando uma tabela pré-estruturada para extração de dados. A intenção foi recolher os dados que responderão com precisão às questões e objetivos da investigação. A tabela foi preenchida com os resultados combinados das diferentes bases de dados.

## 2.6. Elementos dos dados

Os dados recolhidos de artigos selecionados foram resumidos em tabelas descritivas com detalhes de publicação e tópicos definidos previamente como: técnicas utilizadas na implementação de ML e NLP, tipo de tarefas do projeto de execução abordadas, método para obtenção de dados de base, resultados e desempenho obtidos e se houve um aumento da qualidade das propostas realizadas com recurso a técnicas NLP. Também elementos como título, autor, ano entre outros foram registados para fins estatísticos.

## 2.7. Resultados e priorização

O principal objetivo desta revisão é conhecer qual(ais) a(s) melhor(es) técnica(s) para a construção de um algoritmo NLP a ser utilizado na classificação de atividade a partir de MQT e outros documentos dados de um projeto de construção com vista à elaboração de um orçamento contratualmente vinculativo. O objetivo secundário é estudar de que forma estes algoritmos podem influenciar positivamente a produtividade, sustentabilidade e gestão do processo de orçamentação na fase de concurso.

## 2.8. Risco de parcialidade

A análise ao risco de parcialidade dos artigos selecionados realizou-se por dois autores independentes de acordo com uma adaptação da ferramenta da Cochrane Collaboration para avaliar o risco de parcialidade [6].

## 2.9. Síntese de dados

A lista de verificação de PRISMA orientou este processo e a síntese dos dados realizou-se extraindo informações dos artigos elegíveis recolhidos na folha de formulário. Na tabela, os resultados significativos foram destacados.

Para este tipo de estudo, as Meta-Análises não foram aplicadas. Se um conjunto de artigos permitir o desenvolvimento de uma meta-análise, a situação será corrigida mais tarde.

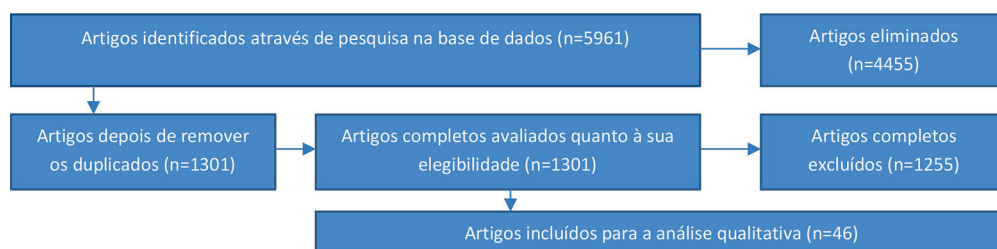
## 2.10. Confiança nas evidências acumuladas

A Revisão Sistemática é no campo da engenharia, conseqüentemente a confiança em provas cumulativas não foi aplicável.

## 3. Resultados preliminares

Os resultados preliminares da revisão realizada são apresentados na Figura 1. Como visível nesta figura, foram inicialmente identificados 5961 artigos relacionados com este tópico; contudo, após o processo de filtragem supramencionado, este número

foi reduzido a 72 artigos. Uma análise inicial destes artigos permitiu detalhar uma primeira proposta para o *framework* idealizado, bem como suportar alguns dos comentários existentes na Secção 1 do artigo. Uma análise pormenorizada dos resultados obtidos será publicada em trabalhos futuros.



**Figura 1**  
Resultados preliminares da estratégia de pesquisa.

#### 4. Abordagem proposta para o desenvolvimento da ferramenta idealizada

Um repositório de especificações dedicado a tarefas de construção civil, como por exemplo MQTs e esquemas de planeamento, é o método mais utilizado na bibliografia revista: Hong [7] selecionou aleatoriamente 27 MQTs de projetos concluídos, sendo estes fornecidos por um empreiteiro no Reino Unido. Estes mapas foram depois etiquetados manualmente utilizando sistemas britânicos de classificação da Construção, como o *Standard Method of Measurement* [8] para criar um conjunto de dados base para suporte do algoritmo de decisão proposto. Moon, [9] utilizou um processo semelhante a Hong, recolhendo sete especificações, incluindo 2 cadernos de encargo utilizados em projetos de Construção de autoestradas no Qatar e 5 especificações-padrão, ou seja, normas nacionais de países como Austrália, Reino Unido e Estados Unidos da América. Moon justifica a aplicação de normas padrão de outros países no Qatar devido a estas estarem bem organizadas, o que assegura a qualidade dos dados.

Um outro método para obtenção de dados é a utilização de modelos *Building Information Modelling* (BIM) [10-13]. Um modelo BIM capta muita informação sobre o ciclo de vida dos edifícios, sendo que é através da recuperação desta informação que os sistemas de apoio à decisão utilizam o BIM [13]. Wang [13], propõe um esquema de recuperação de informação para BIM utilizando, para isso, uma estrutura de árvore hierárquica e técnicas NLP. Através deste sistema o autor consegue aceder à informação dentro do modelo BIM através da estrutura de dados da especificação *Industry Foundation Class* (IFC), analisando esta informação com a ajuda do *International Framework for Dictionaries* (IFD) e algoritmos NLP para desenvolver um novo esquema de recuperação de informação para detetar informação multiforme associada às consultas de modelos BIM.

Estudos anteriores relativos à aplicação de NLP para auxílio na elaboração de orçamentos, apenas se aplicam a um tipo específico de tarefa [11, 14-18]: Akanbi [14] utilizou modelação semântica e técnicas de NLP no desenvolvimento de algoritmos que automatizam os processos manuais envolvidos na elaboração de um orçamento,

aplicando este modelo a um projeto de construção em madeira. O autor foca, em particular, a classificação semântica de atributos como a espessura, a aplicação e o tipo de madeira. Noutro estudo, Zang [18] aplicou um método de segmentação de palavras, baseado na correspondência máxima inversa [19], com o objetivo de analisar as especificações de construção chinesas. Ao examinar as especificações de concepção de escadas, o método proposto trata os símbolos não chineses nas especificações para fazer corresponder os contextos às especificações de design gerais chinesas.

No processo tradicional de contratação de empreitadas de obras públicas um Dono de Obra disponibiliza na plataforma de contratação os elementos que instruem o procedimento. Os MQT são elementos essenciais e é sobre estes que as empresas de construção que decidam concorrer deverão elaborar a sua proposta de preço. Função da experiência e abordagens estratégicas, as empresas trabalham os MQT de modo a identificar os custos para as diferentes tarefas previstas, bem como avaliar a existência de erros e omissões tanto no MQT como nos outros elementos. A identificação dos custos a associar às tarefas do MQT a concurso é um processo permeável a erros. A errada classificação das tarefas pode traduzir-se na perda do concurso pelo facto de a proposta não ser competitiva ou pode traduzir-se num risco acrescido para a empresa no caso de ganhar o concurso e perceber à posteriori a existência desse erro. Na literatura revista, os autores salientaram o valor acrescentado da utilização de documentos estruturados, embora não tenham sido exploradas aplicações práticas. Resta, portanto, a metodologia NLP. As diferenças destas abordagens são as seguintes:

- a primeira baseia-se no desenvolvimento de um documento legal nacional, que descreva as tarefas de forma detalhada e que é adotado pelo setor da Construção. O ProNIC numa fase inicial de desenvolvimento estruturou um leque abrangente de tarefas para a construção e reabilitação de edifícios e integrou o sistema de rubricas para a construção de estradas que tinha sido desenvolvido pela JAE – Junta Autónoma de Estradas. No contexto da Fase 3 do programa da Parque Escolar, EPE, o ProNIC ampliou a base de dados para as especificidades de construção e reabilitação de edifícios de uso escolar. Apesar de várias diligências e regulamentação no sentido de adotar o ProNIC em outras obras públicas, o processo não teve seguimento por parte do IMPIC, IP, entidade responsável pelo projeto em representação do Estado Português [20].
- a segunda abordagem reconhece a dificuldade de uniformizar todos os MQTs da indústria, utilizando uma abordagem NLP para traduzir estes documentos num formato *standard* reconhecido pelo empreiteiro (aqui identificado como MQT “*master*”). Este procedimento tem a vantagem adicional de poder ser aplicado a um projeto internacional, no qual o MQT, pode não estar estruturado de uma forma considerada comum no país de origem do empreiteiro, nem estar escrito no mesmo idioma do MQT “*master*”. Este documento contém tarefas com os preços unitários praticados pela empresa em questão. Para isso, esta abordagem foca-se na identificação de determinadas expressões e/ou palavras-chave que permitam a associação das tarefas de um MQT



qualquer com o MQT “*master*”. Para fazer esta identificação so utilizados algoritmos de aprendizagem computacional probabilísticos, como: *Naïve Bayes (NB)*, *Support vector machines (SVM)*, *K-nearest neighbors (KNN)*, *Decision tree (DT)*, *Logistic regression (LR)*, and *Feedforward neural network (FNN)* [21].

No presente artigo, focou-se a segunda solução, sendo apresentada, na Figura 2, um *framework* para o desenvolvimento de uma solução deste tipo. Como visível nesta figura, podemos generalizar a *framework* para o desenvolvimento de um modelo de classificação de texto com recurso a NLP em seis etapas [22]:

- (1) recolha dos documentos contratuais de projetos passados para formar uma base de dados inicial. Esta base de dados é vital para o sucesso desta solução, uma vez que é com ela que se fará, nas etapas seguintes, o treino e teste dos algoritmos
- (2) execução do pré-processamento dos dados, usando técnicas como *tokenization*, normalização e remoção de ruído, onde são aplicados algoritmos para obtenção de primitivas, uniformização do texto em minúscula, remoção de caracteres especiais e espaçamentos, eliminação de conectores, entre outros;
- (3) seleção das características do texto restante (*feature selection*) com base no seu poder discriminatório;
- (4) desenvolvimento dos vários tipos de algoritmos de NLP a serem treinados e testados;
- (5) treino e teste dos algoritmos desenvolvidos com recurso aos dados recolhidos na etapa 1 e seleção do melhor algoritmo com base no seu desempenho e precisão;
- (6) aplicação do algoritmo selecionado para associação das tarefas do caderno de encargos do Dono de Obra e o ficheiro “*master*” do empreiteiro, culminando na automatização da classificação das tarefas do MQT de execução que servirá de base para o orçamento do empreiteiro.

A obtenção dos dados para a primeira etapa poderá ser realizada através de referenciais como o ProNIC ou então pela cedência desta informação por parte de uma empresa de construção. É importante definir que o que se pretende fazer é uma classificação do MQT final tendo em conta o *master*; ou seja, pretende-se classificar automaticamente, não orçamentar automaticamente.

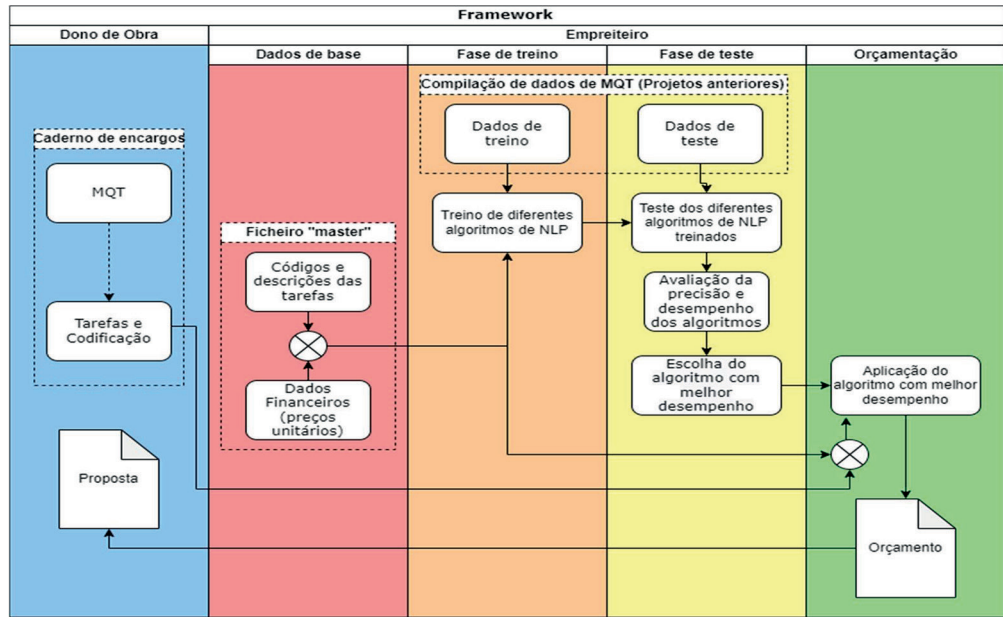


Figura 2 Framework proposto.

### 5. Conclusão e trabalhos futuros

O presente artigo teve como objetivo identificar uma solução para mitigação dos erros associados ao processo de orçamentação durante a fase de concurso, resultantes das restrições de tempo. Foi realizada uma revisão com recurso à metodologia PRISMA, da qual são apresentados os resultados preliminares.

Na literatura recolhida foram identificadas duas possíveis soluções, sendo adotada a metodologia com base em NLP. A eficiência e precisão dos modelos de NLP em tarefas como a extração de texto e análise semântica revela-se uma mais-valia para o processo de tomada de decisão na fase de orçamentação de um projeto de construção [14, 23]. As informações a serem introduzidas no modelo serão analisadas em forma de texto ou Work Breakdown Structure e podem ser obtidas a partir do modelo BIM ou de documentos relacionados com o projeto, tais como MQTs. Com base nesta solução foi proposto um *framework*, dividido em seis etapas distintas, que permite a aplicação desta abordagem ao setor da Construção em Portugal.

Como trabalhos futuros, estão planeadas duas vertentes de trabalho. Na primeira serão realizadas atualizações futuras da presente revisão, com vista à adição de artigos impactantes para a abordagem proposta. Na segunda a *framework* desenvolvida será aplicada a um contexto prático para definição clara das suas vantagens e desvantagens.

### Agradecimentos

Este trabalho foi apoiado financeiramente por: Financiamento de Base-UIDB/04708/2020 do CONSTRUCT – Instituto de I&D em Estruturas e Construções – financiado por fundos nacionais através do FCT/MCTES (PIDDAC). Este trabalho é também cofinanciado pelo Fundo Social Europeu (FSE), através do Programa

Operacional Regional do Norte (Norte 2020) [Referência de Financiamento: NORTE-06-3559-FSE-000176].”

## Referências

- [1] A. Pessoa, G. Sousa, L. M. F. Maues, F. C. Alvarenga, and D. D. Santos, “Cost Forecasting of Public Construction Projects Using Multilayer Perceptron Artificial Neural Networks: A Case Study”, *Ingenieria E Investigacion*, vol. 41, no. 3, Dec 2021, Art no. e87737, doi: 10.15446/ing.investig.v41n3.87737.
- [2] P. Jafari, M. Al Hattab, E. Mohamed, and S. Abourizk, “Automated extraction and time-cost prediction of contractual reporting requirements in construction using natural language processing and simulation”, *Applied Sciences (Switzerland)*, Article vol. 11, no. 13, 2021, Art no. 6188, doi: 10.3390/app11136188.
- [3] (2020). *Contratação Pública em Portugal 2019*.
- [4] H. H. Elmousalami, “Data on Field Canals Improvement Projects for Cost Prediction Using Artificial Intelligence”, *Data in Brief*, vol. 31, p. 105688, 2020, doi: 10.1016/j.dib.2020.105688.
- [5] M. J. Page, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”, *BMJ*, vol. 372, p. n71, 2021, doi: 10.1136/bmj.n71.
- [6] P. Montgomery, & Shepard, L. (n.d.). “Running on Empty: The Cochrane Empty Reviews Project report of findings and consensus group feedback”. [Online]. Available: <https://emptyreviews.files.wordpress.com/2011/11/madridmeetingpresentation-finalpostedits.pdf>.
- [7] Y. Hong, H. Y. Xie, G. Bhumbra, and I. Brilakis, “Comparing Natural Language Processing Methods to Cluster Construction Schedules”, *J. Constr. Eng. Manage.*, vol. 147, no. 10, Oct 2021, Art no. 04021136, doi: 10.1061/(asce)co.1943-7862.0002165.
- [8] C. C. Royal Institution of Chartered Surveyors, *SMM7: standard method of measurement of building works*. London: Royal Institution of Chartered Surveyors (in English).
- [9] S. Moon, G. Lee, and S. Chi, “Semantic text-pairing for relevant provision identification in construction specification reviews”, *Autom. Constr.*, Article vol. 128, 2021, Art no. 103780, doi: 10.1016/j.autcon.2021.103780.
- [10] M. Juszczuk, K. Zima, and W. Lelek, “Forecasting of sports fields construction costs aided by ensembles of neural networks”, *J. Civ. Eng. Manag.*, Article vol. 25, no. 7, pp. 715-729, 2019, doi: 10.3846/jcem.2019.10534.

- [11] M. Juszczuk, "Implementation of the ANNs ensembles in macro-BIM cost estimates of buildings' floor structural frames", 2018, p. 020014, doi: 10.1063/1.5030318. [Online]. Available: <https://app.dimensions.ai/details/publication/pub.1103695750>
- [12] M. Juszczuk and A. Leśniak, "Modelling Construction Site Cost Index Based on Neural Network Ensembles", *Symmetry*, vol. 11, no. 3, p. 411, 2019, doi: 10.3390/sym11030411.
- [13] J. Wang, X. A. Gao, X. P. Zhou, and Q. S. Xie, "Multi-scale information retrieval for bim using hierarchical structure modelling and natural language processing", (in english), *journal of information technology in construction*, vol. 26, pp. 409-426, 2021, doi: 10.36680/j.itcon.2021.022.
- [14] T. Akanbi and J. S. Zhang, "Design information extraction from construction specifications to support cost estimation" (in English), *Autom. Constr.*, vol. 131, NOV 2021, doi: 10.1016/j.autcon.2021.103835.
- [15] H. Elhegazy, "Artificial Intelligence for Developing Accurate Preliminary Cost Estimates for Composite Flooring Systems of Multi-Storey Buildings", *Journal of Asian Architecture and Building Engineering*, 2021, doi: 10.1080/13467581.2020.1838288.
- [16] M. Juszczuk, A. Leśniak, and K. Zima, "ANN Based Approach for Estimation of Construction Costs of Sports Fields", *Complexity*, vol. 2018, pp. 1-11, 2018, doi: 10.1155/2018/7952434.
- [17] S. Lakusic, "Use of artificial intelligence for estimating cost of integral bridges", *Journal of the Croatian Association of Civil Engineers*, vol. 73, no. 03, pp. 265-273, 2021, doi: 10.14256/jce.2831.2019.
- [18] J. Zhang, "A RMM based Word Segmentation Method for Chinese Design Specifications of Building Stairs", in *14th International Conference on Computational Intelligence and Security (CIS)*, Hangzhou, PEOPLES R CHINA, Nov 16-19 2018, 2018, pp. 277-280, doi: 10.1109/cis2018.2018.00068. [Online]. Available: <Go to ISI>://WOS:000456370300060
- [19] H. Y. Qu and W. Zhao, "A Revised BMM and RMM Algorithm of Chinese Automatic Words Segmentation", presented at the MANUFACTURING SYSTEMS AND INDUSTRY APPLICATIONS, 2011.
- [20] P. Mêda, H. Sousa, and J. Moreira, "ProNIC contributions for building refurbishment – Procedures and Technology", 2014.
- [21] F. Gürçan, "Multi-Class Classification of Turkish Texts with Machine Learning Algorithms", in *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 19-21 Oct. 2018 2018, pp. 1-5, doi: 10.1109/ISMSIT.2018.8567307.

- [22] F. Ul Hassan, T. Le, and D. H. Tran, "Multi-Class Categorization of Design-Build Contract Requirements Using Text Mining and Natural Language Processing Techniques", 2020: American Society of Civil Engineers (ASCE), pp. 1266-1274. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85096775363&partnerID=40&md5=9d03020e09d8cc3942e26264a6f8dc69>. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85096775363&partnerID=40&md5=9d03020e09d8cc3942e26264a6f8dc69>
- [23] R. Y. M. Li, H. C. Y. Li, B. Tang, and W. C. Au, "Fast AI classification for analyzing construction accidents claims", 2020: ICST, pp. 1-4, doi: 10.1145/3407703.3407705. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090386362&doi=10.1145%2f3407703.3407705&partnerID=40&md5=eb3d8cb4611d6b48d80a5e1d1ae5171c>