# Métodos de Seleção de Características na Aplicação de Modelos de Aprendizagem de Máquina para o Prognóstico de Câncer: Um Mapeamento Sistemático da Literatura

# Feature Selection Methods in the Application of Machine Learning Models for Cancer Prognosis: A Systematic Mapping of Literature

# Métodos de selección de características en la aplicación de modelos de aprendizaje automático para el pronóstico del cáncer: un mapeo sistemático de la literatura

Maxwell Acioli1, Rodrigo1, Rafael2

1 Instituto de Computação, Universidade Federal de Alagoas – UFAL- AL, Maceió (AL), Brasil.

2 Professor adjunto do Instituto de Computação, Universidade Federal de Alagoas – UFAL- AL, Maceió (AL), Brasil.

Autor correspondente: Maxwell Acioli

E-mail: meas@ic.ufal.br

## Resumo

Neste trabalho, conduzimos um mapeamento sistemático para investigar quais métodos de seleção de características têm sido aplicados no domínio dos modelos de aprendizagem de máquina. Entendemos que a seleção de características é uma etapa importante no processo da construção de um modelo preditivo. Neste contexto, avaliamos quais são os métodos e técnicas mais utilizadas. Como resultado, identificamos quais sãos as técnicas e métodos mais aplicados em modelos de aprendizagem de máquina no prognóstico de câncer.

**Descritores:** Seleção de características; Aprendizagem de Máquina; Câncer

## Abstract

In this work, we conduct a systematic mapping to investigate which characteristic selection methods have been applied in the domain of machine learning models. We understand that the selection of characteristics is an important step in the process of building a predictive model. In this context, we evaluate which are the most used methods and techniques. As a result, we identified which techniques and methods are most applied in machine learning models in cancer prognosis.

**Keywords:** Features selection; Machine Learning; Cancer

## Resumen

En este trabajo, realizamos un mapeo sistemático para investigar qué métodos de selección de características se han aplicado en el dominio de los modelos de aprendizaje automático. Entendemos que la selección de características es un paso importante en el proceso de construcción de un modelo predictivo. En este contexto, evaluamos cuáles son los métodos y técnicas más utilizados. Como resultado, identificamos qué técnicas y métodos se aplican más en los modelos de aprendizaje automático en el pronóstico del cáncer.

**Descriptores:** Selección de características; Aprendizaje automático; Cáncer

## Introdução

No contexto de aprendizagem de máquina, características podem ser consideradas propriedades de um fenômeno observado. Estas são usualmente utilizadas em um processo de reconhecimento de padrões sistemático através da aplicação de modelos de aprendizagem de máquina. Geralmente características são numéricas, mas também podem ser estruturadas no formato de strings. Já a seleção de características pode ser denominada como o processo que visa a redução da quantidade de características que são utilizadas em um modelo, esta redução pode ser realizada através de processamento e análise destas, com o objetivo de encontrar quais são as características mais significativas dentre todas as propriedades existentes.

Estas instruções têm como objetivo ajudar os autores a preparar a submissão de artigo completo. Recomenda-se utilizar este *template* ou modelo para a preparação da submissão do artigo. Esta é uma representação exata do formato esperado pelo editor.

Esta modalidade trata de artigo completo e original escrito a partir de estudo científico. Para o CBIS e o respectivo Suplemento Anais de Congresso no ***Journal of Health Informatics*** serão aceitas duas modalidades de artigo:

* **Artigos Originais**: trabalhos de pesquisa com resultados inéditos e que tragam valor à área de Informática em Saúde, com, no máximo, 14 páginas.
* **Artigos de Revisão**: destinados a incorporar e sintetizar o conhecimento existente sobre o tema que aborda, com base na literatura pertinente, abrangente, crítica e sistemática, acrescido de análise e conclusão, com no máximo 14 páginas.

Os artigos fora do padrão apresentado neste modelo serão automaticamente recusados pelo corpo de revisores do CBIS. Os artigos aceitos serão publicados em um suplemento especial do periódico ***Journal of Health Informatics***.

O número de autores por trabalho está restrito a 10 (dez). Com o objetivo de garantir maior oportunidade de participação a todos, um autor pode participar de múltiplas submissões, mas deve ter somente uma submissão por modalidade como primeiro autor.

Para ter o texto publicado é obrigatório a inscrição de pelo menos um autor no CBIS 2020. A inscrição do autor é válida para no máximo 2 trabalhos aceitos em qualquer modalidade.

## Idioma, Seções e Conteúdo do Artigo

O idioma oficial do CBIS é o português, mas os artigos podem ser submetidos, também, em inglês ou espanhol. O conteúdo exigido para o artigo completo na primeira página é:

1. **Título do Artigo:** o título no idioma original do manuscrito (português, espanhol ou inglês) deve ser conciso e ilustrativo da temática abordada; O título deve estar nos três idiomas (português, inglês e espanhol), contendo no máximo doze palavras. O título do artigo deve ser escrito por extenso, sem abreviações;
2. **Autores e sua titulação:** Para descrever o nome dos autores e a titulação o limite é de 10 (dez) palavras. Lembramos que a titulação deve refletir o título mais alto e permanente do autor (veja o exemplo no início deste *template*).
3. **Resumo:** deverá estar em português (Resumo), em inglês (Abstract) e espanhol (Resumen), contendo no máximo 150 palavras. O resumo deve estar estruturado de forma a conter os seguintes itens: objetivo, métodos, resultados e conclusão.
4. **Descritores:** os autores devem indicar até três descritores, que representem a temática abordada no texto, conforme documento de chamada para a submissão de trabalhos. Os descritores devem ser apresentados nos três idiomas (português, inglês e espanhol) e devem, preferivelmente, ser selecionadas dos termos do DECS ([http://decs.bvs.br](http://decs.bvs.br/)).

Após título, resumos e descritores, recomenda-se que o artigo contenha as seguintes seções:

* 1. **Introdução** - apresenta um panorama geral do assunto que será desenvolvido no artigo. Deve fornecer um contexto geral do que se deve conhecer para entender o restante do artigo, a motivação para a pesquisa desenvolvida, o estado da arte no objeto de estudo da pesquisa, assim como o seu objetivo. (1,2)
  2. **Métodos** - descreve o método e as técnicas para a elaboração da pesquisa realizada, assim como a forma de registro e controle de informações, e as ferramentas que foram utilizadas. Além disso, em caso de experimentos descreve o cenário do experimento, da amostra, dos procedimentos e dos aspectos éticos. (1,2)
  3. **Resultados e Discussão** – descreve o que foi encontrado ao final do experimento ou da pesquisa realizada. Esta seção é composta pelos dados relevantes obtidos e sintetizados pelo autor, acompanhados da respectiva análise estatística, se aplicável. A apresentação, pura e simples, dos resultados não garante o seu pleno entendimento. O objetivo da discussão é explorar o significado dos resultados, estabelecer a comparação com achados de pesquisas de outros autores e fornecer subsídios para que se possa julgar a adequação dos argumentos, da conclusão e de todo o texto. (1,2)
  4. **Conclusão** – é o principal item do artigo, deve ser conciso e apresentar a conclusão da pesquisa. Os argumentos nesta seção devem responder à questão de pesquisa que foi traduzida no objetivo apresentado na introdução. A seção conclusão, deste modelo/*template*, apresenta considerações importantes que podem ser uteis ao se escrever a conclusão. (3)
  5. **Agradecimentos (opcional)** - aqueles que contribuíram para a elaboração do texto, mas não se enquadram como autores ou co-autores, de acordo com as definições anteriores, deverão ser relacionados na seção de Agradecimentos. Os autores devem informar se receberam assistência (financeira ou não) e identificar a entidade que a providenciou. Apoio financeiro, como bolsas de estudo e pesquisa, deve também ser mencionado na seção de Agradecimentos.
  6. **Referências** – As referências são as fontes bibliográfica utilizadas para o desenvolvimento do artigo e que foram citadas ao longo do texto. Toda menção à bibliografia deve ter sua referência no texto, caso contrário não se enquadra neste item. O limite de referências para o artigo completo é de 25. O estilo adotado para a apresentação de referências bibliográfica  é o Vancouver (<http://www.nlm.nih.gov/bsd/uniform_requirements.html>).

Para artigos de revisão sistemática recomenda-se seguir o itens do *checklist* que se encontra no artigo “Principais itens para relatar Revisões Sistemáticas e Meta-análises: A recomendação PRISMA\*” (<https://www.scielo.br/pdf/ress/v24n2/2237-9622-ress-24-02-00335.pdf>).

## Formatação Geral do Artigo

Os textos devem ser editados em processador eletrônico de texto (preferencialmente Microsoft Word ou OpenOffice), utilizando-se um arquivo no formato.DOC (Documento de Texto do Word), .RTF (Rich Text Format) ou .ODT (ODF Document Format, do OpenOffice ou BrOffice). É preciso respeitar o envio nestes formatos para que a versão final possa ser diagramada de acordo com os padrões do JHI.

O formato da página é A4 em uma única coluna, com espaçamento de 1,5 cm entre linhas, e margens de 2,5 cm em todos os lados. Para o texto use fonte Arial tamanho 12 em modo justificado com espaçamento 1,5 linha e o recuo da primeira linha de 1,27 cm. Os termos estrangeiros ao longo do texto devem ser escritos em itálico, com exceção de nomes próprios. As unidades de medida devem ser abreviadas com exatidão, de acordo com o padrão Inmetro.

As abreviaturas podem ser usadas para evitar a repetição, mas somente usando-se a forma consistente dentro de um domínio. Abreviaturas devem ser introduzidas entre parênteses após a frase completa ter sido apresentada pela primeira vez no manuscrito (em seu corpo propriamente, não nos metadados). As abreviaturas devem ser evitadas em títulos, subtítulos e no resumo. A colocação de pontos ou espaços nas abreviaturas deve ser evitada.

As notas de rodapé devem usar Times 9 pt e serem colocadas na parte inferior da coluna em que aparecem[[1]](#footnote-2). As notas de rodapé devem ser pouco usadas. Os números de nota de rodapé no texto devem estar em superscrito.

## Formatação da Seção do Artigo

Para os títulos das seções utilize fonte Arial tamanho 12 em negrito. O espaçamento antes deve ser de 18 pt e depois de 6 pt. O espaçamento é simples, à esquerda sem recuo.

## Formato da Subseção

Para os títulos das subseções utilize fonte Arial tamanho 10 em negrito. O espaçamento antes deve ser de 18 pt e depois de 6 pt. O espaçamento é simples, à esquerda sem recuo.

## Citação de Referências

Ao usar uma fonte seja ela um livro, um artigo científico, um *white paper*, um CD, um panfleto ou qualquer outra forma de publicação, deve-se ter o cuidado para não “roubar” o material de outra pessoa. A forma legalmente reconhecida de fazer referência ao que outra pessoa publicou em um artigo científico são as citações. A citação pode ser: (4)

* + Formal, direta ou transcrição – é a citação onde se transcreve tudo do autor consultado, palavra por palavra. Se for de até três linhas, usar aspas. Se for de mais que três linhas, utiliza-se uma fonte menor (tamanho 10), com recuo de 4 cm e espaçamento simples, justificado.
  + Indireta ou paráfrase – quando se escreve a ideia do(s) autor(es) consultado(s) com as suas próprias palavras.
  + Citação de citação – quando os autores não possuem acesso à fonte original. Um determinado autor é mencionado pelo fato de ter sido encontrado numa referência. Utiliza-se a expressão *apud* para indicar “citado por”.

O estilo adotado para a apresentação de referências bibliográfica  é o Vancouver (<http://www.nlm.nih.gov/bsd/uniform_requirements.html>). Vale ressaltar que o estilo Vancouver deve ser aplicado no mesmo idioma que o do conteúdo citado na referência em questão. Abreviaturas para revistas são aquelas usadas no MeSH (<http://www.nlm.nih.gov/mesh>), publicadas pela U.S National Library of Medicine. Se possível e pertinente, recomendam-se citações a artigos do **J. Health Inform.**

## Dados Suplementares

Material suplementar que não couberem no artigo podem ser referenciados ao longo do texto para um endereço com acesso pela internet. Os arquivos suplementares oferecem ao autor possibilidades adicionais de publicar aplicações de suporte (softwares), filmes, sequência de animações, arquivos de som, formulários e questionários. Existem sites especializados em arquivamento de suplementos indicamos usá-los, um exemplo é o site: <https://zenodo.org/>

## Figuras, Tabelas, Quadro e Gráficos

As tabelas e quadros, incluindo gráficos e figuras, devem ser incluídos no texto em seu local apropriado e devem ser limitados a um total de 10 (dez).  Fotografias, esquemas e cópias de telas de computador são considerados figuras. Coloque a referência da fonte de suas figuras, tabelas e gráficos. Se forem de sua autoria não é necessário referenciar. As tabelas e quadros devem ser acompanhados de cabeçalho e numerados consecutivamente em algarismos arábicos. O título da tabela ou quadro se posiciona acima do mesmo, como aparece na tabela 1 e quadro 2.

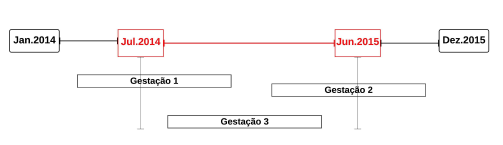
**Tabela 1 –** Exemplo de uma tabela com destaque de cores (5)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Médico** | **Especialidade** | **2013 Q2** | **2013 Q3** | **2013 Q4** | **2014 Q1** | **2014 Q2** |
| ***Msp164*** | - | 3 | 4 | 7 | 4 | 5 |
| ***Msp142*** | - | 4 | 5 | 3 | 3 | 3 |
| ***Msp242*** | Nefrologia | 5 | 10 | 10 | 14 | 12 |
| ***Msp243*** | Ortopedia e Traumatologia | 6 | 8 | 8 | 6 | 8 |
| ***Msp140*** | - | 7 | 2 | 4 | 5 | 4 |
| ***Msp148*** | Cardiologia | 8 | 7 | 5 | 13 | 11 |
| ***Msp149*** | Hematologia e Patologia Clínica | 9 | 12 | 14 | 12 | 16 |
| ***Msp244*** | - | 10 | 13 | 12 | 15 | 10 |

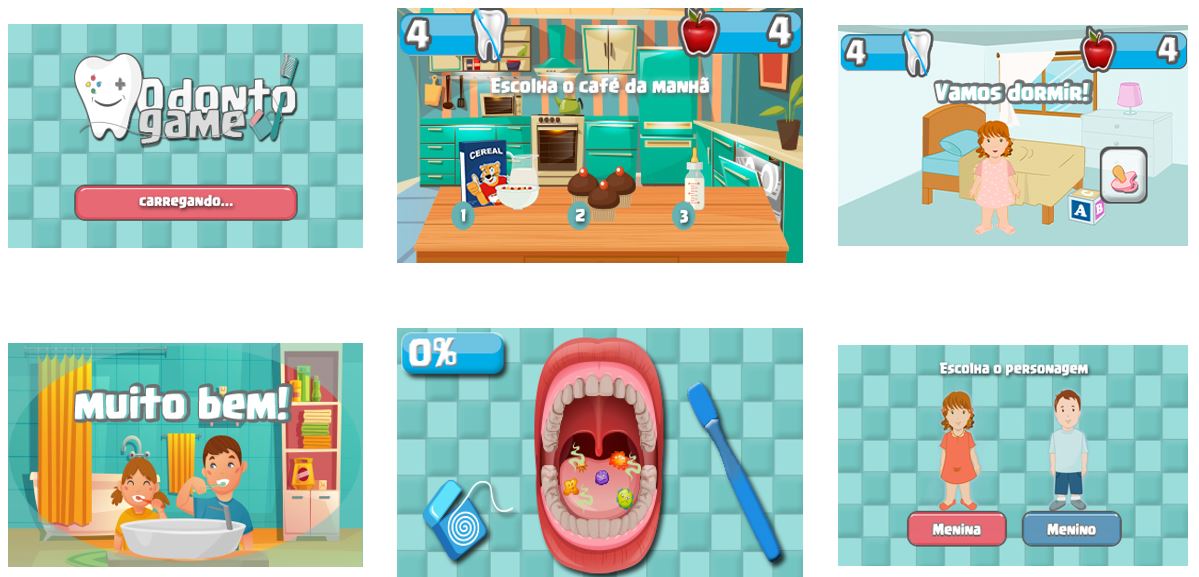
**Quadro 1 –** Exemplo de um quadro (6)

|  |  |  |  |
| --- | --- | --- | --- |
| **Área temática** | **Autores e ano** | **Título do artigo** | **Senso de presença** |
| **Promoção da saúde e Prevenção de doenças e agravos** | Alyssa M. P; Eric D. R. (2017) | Contextualizing construction accident reports in virtual environments for safety education | Imersiva |
| Lee, S; Shin, S (2013). | Effectiveness of virtual reality using video gaming technology in elderly adults with diabetes mellitus | Não imersiva |
| Thomas J.G.; Spitalnick J.S.; Hadley W.; Bond D.S.; Wing R.R. (2015) | Development of and feedback on a fully automated virtual reality system for online training in weight management skills. | Não imersiva |
| Weiner, E ; Trangenstein, P ; McNew, R; Gordon, J. (2016) | Using the Virtual Reality World of Second Life to Promote Patient Engagement | Não imersiva |
| Ge Jin; Shoji Nakayama. (2014) | Virtual reality game for safety education | Não imersiva |
| R.C.W. Webber-Youngman; E.A. van Wyk. (2013) | Incident reconstruction simulations-potential impact on the prevention of future mine incidents | Não imersiva |
| **Reabilitação da saúde** | Voinea, A; Moldoveanu, A; Moldoveanu, F; Ferche, O. (2015) | Motion Detection and Rendering for Upper Limb Post-Stroke Rehabilitation | Imersiva |
| Campelo. A.M ; Hashim, J.A.; Weisberg, A ; Katz, L. (2017) | Virtual Rehabilitation in the Elderly: Benefits, Issues, and Considerations | Não imersiva |
| Monteiro, C.B.M.; Massetti, T; Silva, T.D.; Kamp, J.V.; Abreu, L.C.A.; Leone, C.; Savelsbergh, G.J.P.(2014) | Transfer of motor learning from virtual to natural environments in individuals with cerebral palsy | Não imersiva |
| Kato, N ; Tanaka, T. ; Sugihara, S; Shimizu, K ; Kudo, N. (2016) | Trial operation of a cloud service-based three-dimensional virtual reality tele-rehabilitation system for stroke patients | Não imersiva |
| Lorenzo, S.M.; Braccialli, L.M.P.; Araújo, R.C.T.(2015) | Realidade Virtual como Intervenção na Síndrome de Down: uma Perspectiva de Ação na Interface Saúde e Educação | Não imersiva |

Para as figuras e gráficos aplicam-se as mesmas regras da tabela, com exceção do cabeçalho que aparece abaixo da figura ou gráfico, veja os exemplos na figura 1, figura 2 e gráfico 1. As figuras devem ter qualidade suficiente para impressão.



**Figura 1 –** Exemplo de uma figura (7)



**Figura 2 –** Exemplo de uma figura com telas de um computador (8)

**Gráfico 1 –** Exemplo de um gráfico (9)

## Equações

As equações devem ser numeradas sequencialmente, no lado direito entre parênteses segundo as indicações da equação (1).

 (1)

## Conclusão

Escrever um artigo científico do tipo completo não é tarefa simples e exige que o estilo de escrita científica seja obedecido. Recomendamos que se utilizem as ferramentas de revisão e que principalmente depois de escrever, o documento seja lido várias vezes, para assegurar que o artigo esteja completo, objetivo e claro.

É fundamental que a conclusão seja embasada pelos resultados obtidos. Cuide para não estabelecer conclusões que não sejam fundamentadas nos seus dados. Mesmo que os resultados obtidos tenham sido diferentes dos esperados inicialmente, descreva as suas descobertas e as relacione com o objetivo do seu trabalho. Certifique-se de que haja coerência entre o título, os desfechos escolhidos nos objetivos, a conclusão do resumo e a conclusão final do artigo. (3)

## Agradecimentos

Agradecimentos aos membros do comitê científico do CBIS 2020 que auxiliaram no desenvolvimento deste *template*.

À instituição SBIS, pela realização das medidas ou empréstimo de equipamentos.

À FAPESP (ou CNPq ou outra), pelo apoio financeiro ou por bolsas concedidas.

Ao Professor/Pesquisador/Técnico XXX pela colaboração.

## Referências

1. Wazlawick R. Metodologia de Pesquisa para Ciência da Computação. 2a. ed. São Paulo: GEN LTC; 2017. 150 p.
2. Pereira MG. Dez passos para produzir artigo científico de sucesso [Internet]. Epidemiol. Serv. Saude 2017 jul./set.;26(3) doi: http://dx.doi.org/10.5123/s1679-49742017000300023[cited 2020 Jul 28]. Available from: <http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742017000300661>
3. Caramelli B. Conclusão: como exibir a cereja do bolo [Internet]. Ver. Assoc. Med. Bras. 2012 nov./dec.; 58(6) doi: http://dx.doi.org/10.1590/S0104-42302012000600001 [cited 2020 Jul 28]. Available from: <https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-42302012000600001>
4. Aquino IS. Como escrever artigos científicos: sem Arrodeio e sem medo da ABNT. 8a. ed. São Paulo:Saraiva; 2012. 126 p.
5. Ito M, Moyano LG, Appel AP, de Santana VF. Análise do relacionamento na comunidade de médicos de seguradoras de saúde. Anais do XV Congresso Brasileiro de Informática em Saúde. 2016 Nov 27-30; Goiania. 309-18.
6. Campos Filho AS, Lemos WB, Souza, RC, Lima, LLB. Realidade virtual como ferramenta educacional e assistencial na saúde: uma revisão integrativa.J.Health Inform. 2020 Abril-Junho; 12(2):58-63.
7. Da Silva MA, Santos N, Rosa COCS, Santos M, Ito M, Vieira A, Ziviani A, Oliveira RM. Análise dos atendimentos de gestantes na rede de atenção básica de saúde no município de São Paulo. Anais do XX Simpósio Brasileiro de Computação Aplicada a Saúde. 2020 Sept 15-18; Salvador. (no prelo)
8. Ito C, Haddad AE, Pitta Neto L., Ito M. OdontoGame: Um jogo sério para conscientização de higiene bucal de crianças em idade pré-escolar. Proceedings of XVIII SBGames - Game and Health Workshop.2019 Oct 28; Rio de Janeiro. 1408-9.
9. Saraiva EH, Patrao DFC, Ito M. Analysis of usage of term weighting algorithm for mapping health procedures into the unified terminology of supplemental health (TUSS). Stud Health Technol Inform. Medinfo 2019: Health and Wellbeing e-Networks for All. 2019: 1496-7. doi: 10.3233/SHTI190502

1. Exemplo de Nota de Rodapé [↑](#footnote-ref-2)