

X-Wines: Dados sobre Vinhos para Ampla Utilização

Rogério Xavier de Azambuja^{1,2,3}, A. Jorge Morais^{2,4}, Vítor Filipe^{3,4}

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS)
95.174-274 – Farroupilha, RS – Brazil

²Departamento de Ciência e Tecnologia (DCeT)
Universidade Aberta (UAb) – 1269-001 – Lisboa – Portugal

³Escola de Ciência e Tecnologia (ECT)
Universidade de Trás-os-Montes e Alto Douro (UTAD)–5000-801–Vila Real – Portugal

⁴INESC Tecnologia e Ciência (INESC TEC) – 4200-465 – Porto – Portugal

rogerio.xavier@farroupilha.ifrs.edu.br, jorge.morais@uab.pt,
vfilipe@utad.pt

Abstract. *In the current scenario of technological progress, similar to most agricultural products, wine has a very small volume of data available or only a few attributes, which restricts scientific exploration. This also applies to recommender systems. This paper presents and evaluates a new database called X-Wines in its first year of publication. It consists of 100,646 wine labels produced in 62 countries, with 21 million real consumer ratings sourced from the open Web in the year 2022. X-Wines is available for wider free use in recommender systems, machine learning, and general purposes, contributing to data science.*

Resumo. *No atual cenário de crescimento tecnológico, à semelhança da maioria dos produtos agrícolas, o vinho apresenta um volume de dados disponibilizado muito reduzido ou com poucos elementos, o que limita a exploração científica, como é o caso nos sistemas de recomendação. Este artigo apresenta e avalia uma nova base de dados denominada X-Wines no seu primeiro ano de publicação. Ela é constituída por 100.646 rótulos de vinhos produzidos em 62 países e 21 milhões de classificações reais dos consumidores encontrados na Web aberta em 2022. X-Wines é disponibilizada para ser livremente utilizada em sistemas de recomendação, aprendizado de máquina e uso geral, como uma contribuição à ciência de dados.*

1. Introdução

Na atualidade é encontrado um volume cada vez maior de informações on-line, dado o aumento do interesse por aplicações Web e a oferta de produtos e serviços em escala mundial, como por exemplo em: *e-commerce*, *e-business*, *e-learning*, *e-tourism*, etc. [Zhang et al. 2021]. Utilizando dados gerados em diferentes domínios, os sistemas de recomendação (*recommender systems* – RSs) têm se apresentado como uma estratégia eficaz para mitigar a sobrecarga de informações ao identificar itens relevantes oferecidos aos utilizadores [Shao et al. 2021]. A recomendação é identificada na literatura como um problema e os comportamentos personalizados dos utilizadores

podem ser úteis, a fim de prever suas preferências individuais para atender demandas.

No âmbito do projeto de pesquisa WD4WU – *Wine Data for Wider Use* (disponível em: <https://sites.google.com/farroupilha.ifrs.edu.br/xwines>. Acesso em: 25/02/2024), desenvolvido no Instituto Federal do Rio Grande do Sul (IFRS) com a colaboração de universidades portuguesas, os RSs são aplicados num caso de estudo do setor vinícola para recomendar vinhos para os utilizadores em ambiente Web. O domínio de vinhos foi escolhido por se tratar de um produto importante para a economia de Portugal e do Brasil, especialmente para a região da Serra Gaúcha no Rio Grande do Sul. Entretanto, poucos conjuntos de dados de vinhos estão disponíveis para uso em sistemas de recomendação, aprendizado de máquina e para uso geral. Atualmente são encontradas diferentes bases de dados disponibilizadas em repositórios on-line, tais como as usuais MovieLens¹, Amazon², Diginetica³, Yoochoose³, Github⁴, Pinterest⁵, Gowalla⁶, Taobao⁷, entre outras. Estas são amplamente conhecidas e utilizadas como referência em execuções de modelos computacionais de recomendação e, pela escassez de dados identificada, poucos trabalhos são encontrados no domínio de vinhos.

Para ajudar a superar esta lacuna encontrada na carência de dados no domínio de vinhos disponibilizados de forma ampla e aberta, o projeto WD4WU apresentou uma nova e consistente base de dados de vinhos denominada X-Wines [de Azambuja et al. 2023]. Ela é composta por mais de 100 mil instâncias e 21 milhões de avaliações reais realizadas por 1 milhão de utilizadores. Os dados foram coletados na Web aberta em 2022 e pré-processados para a distribuição gratuita mais ampla. Os conjuntos de dados referem-se às características encontradas nos rótulos dos vinhos e nas avaliações dos consumidores, classificadas em escala 5 estrelas (1 a 5) ao longo de um período de 10 anos (2012-2021) para vinhos produzidos em 62 países.

Como organização deste artigo, a base de dados X-Wines é caracterizada nas próximas seções. Primeiramente são apresentados aspectos de sua construção e de sua usabilidade em sistemas de recomendação e, em seguida, alguns resultados evidenciados no primeiro ano de sua publicação são explicitados por meio das citações recebidas em 2023. Por fim, o artigo é concluído em observância aos resultados obtidos.

2. Motivação para Uma Nova Base de Dados no Domínio de Vinhos

O vinho é, sem dúvida, um produto de grande requinte e com uma cadeia produtiva que exige conhecimento, talento e criatividade. Há quem o considere ser um produto de arte. De acordo com a *International Organization of Vine and Wine* (OIV) [Juban 2022:4], "*O vinho é exclusivamente a bebida resultante da fermentação alcoólica total ou parcial de uvas frescas, esmagadas ou não, e do mosto de uvas*". Adquire estrutura, aroma e sabor próprios ligados às características da produção, sendo um produto alcoólico muito apreciado pelos adultos e de grande importância econômica global.

No desenvolvimento do projeto de pesquisa WD4WU, para além da construção

¹ Disponível em: <https://grouplens.org/datasets/movielens>. Acesso em: 27/03/2023.

² Disponível em: <http://jmcauley.ucsd.edu/data/amazon/>. Acesso em: 27/03/2023.

³ Disponível em: <https://beta-recsys.readthedocs.io/en/latest/notes/datasets.html>. Acesso em: 27/03/2023.

⁴ Disponível em: <https://github.com/John-K92>. Acesso em: 27/03/2023.

⁵ Disponível em: <https://sites.google.com/site/xueatalphabeta/academic-projects>. Acesso em: 27/03/2023.

⁶ Disponível em: <http://snap.stanford.edu/data/loc-gowalla.html>. Acesso em: 27/03/2023.

⁷ Disponível em: <https://tianchi.aliyun.com/datalab/dataSet.html?dataId=649>. Acesso em: 27/03/2023.

de um novo modelo de recomendação no domínio de vinhos, escolhido como caso de estudo, encontrou-se a necessidade da construção prévia de uma ampla base de dados. Também foi construída, de forma complementar, uma plataforma colaborativa Web para suportar a recomendação no ambiente on-line, estimulando a colaboração entre utilizadores. A Figura 1 ilustra os três artefatos construídos: a base de dados X-Wines, o modelo X-Model4Rec e a plataforma colaborativa Web apresentando os seus principais elementos e interligações utilizadas para gerar a recomendação adaptativa de vinhos aos utilizadores Web (à direita na figura). A geração da recomendação adaptativa proposta acontece em observância aos conjuntos de dados disponíveis em momentos contextuais diferentes, ilustrados pelos intervalos temporais (t) [de Azambuja et al. 2022]. Na construção, os algoritmos foram codificados na linguagem de programação Python (Python Software Foundation, Wilmington, DE, USA) por meio de suas bibliotecas.

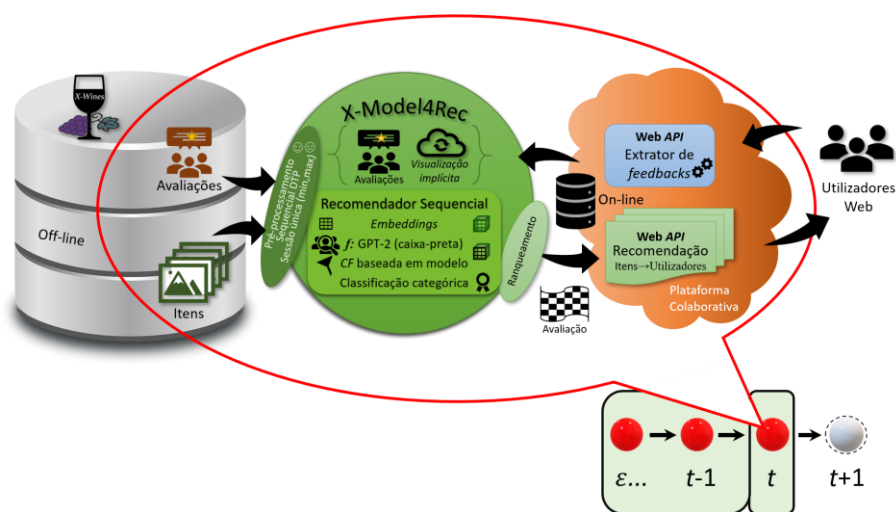


Figura 1. Artefatos construídos para suportar a recomendação adaptativa on-line.

Recentemente, a investigação sobre aplicações de RSs vem aumentando no comércio eletrônico [Geuens et al. 2018; Schafer et al. 2001], no domínio aberto das redes sociais [Yang et al. 2020; Sperli et al. 2018], multimídia [Baek and Chung 2021; Yang et al. 2017], saúde e comportamento saudável [Sahoo et al. 2019; Iwendi et al. 2020], *e-tourism* [Fararni et al. 2021; Artemenko et al. 2017] e outros. No campo agrícola, trabalhos em domínios amplos são encontrados, tais como a sugestão do cultivo de culturas considerando uma base de dados de solos [Kulkarni et al. 2018] e a assistência aos pedidos de informação dos agricultores através de um sistema de recomendação colaborativo [Jaiswal et al. 2020]. Estudos em domínios específicos também são encontrados, tais como a recomendação de fertilizantes aos agricultores [Archana and Saranya 2020] e a recomendação de tratamentos para pragas [Lacasta et al. 2018], contudo encontram-se poucos trabalhos envolvendo o domínio de vinhos, uma lacuna identificada na revisão da literatura realizada.

3. Coleta de Dados

O processo de coleta de dados foi realizado em vários websites especializados em vinhos na Web aberta entre os meses de fevereiro e março de 2022. Não foram utilizados dados privados, respeitando as leis de proteção de dados e privacidade e nenhum sistema teve a sua segurança verificada ou violada. Os portais foram oficialmente contatados e informados sobre esta pesquisa, não havendo objeções na

coleta de dados, realizada em períodos definidos. Nos contatos efetuados, algumas vinícolas agradeceram formalmente pela exposição e divulgação dos seus produtos.

Primeiramente, foi realizada uma pesquisa exploratória de candidatos a itens de vinhos e websites de vinícolas. Basicamente, os algoritmos procuraram detectar palavras-chave em metadados, pontuações atribuídas, URIs candidatos e, assim, um conjunto de referências eletrônicas no domínio de vinhos foi montada. Posteriormente foram exploradas por algoritmos de extração as principais fontes de referência encontradas; entre elas, mais de 350 foram consideradas úteis para fornecer dados sistemáticos: atributos identificados como objetos de vinhos e suas instâncias reais, além de imagens não marcadas ilustrando os rótulos dos vinhos e outros arquivos com informações técnicas. Várias adaptações específicas nos algoritmos exploratórios e de extração de conteúdo foram necessárias. As avaliações realizadas por utilizadores reais cadastrados nas plataformas Web seguiram o protocolo de processamento dos dados extraídos somente após a identificação da autorização destes via website para publicar suas avaliações e, assim, ajudar outras pessoas em um formato de rede social aberta.

A maioria dos dados coletados foram retirados diretamente de websites no formato textual, ainda alguns atributos foram encontrados em abundância nas imagens de rótulos dos vinhos e em arquivos com as especificações técnicas dos produtos. Também foram encontradas, em menor quantidade, informações disponíveis na Web em notas de avaliações produzidas pelos utilizadores. Quando a informação não era encontrada explicitamente nos websites, incluindo os oficiais dos produtores, que por vezes não forneciam informações importantes sobre os seus produtos, foram aplicados nestes casos algoritmos específicos de extração de texto, processamento de linguagem natural (traduções entre idiomas) e seleção para encontrar relevância nos dados. O reconhecimento óptico de caracteres (OCR) foi utilizado para explorar o cisalhamento da imagem do rótulo na garrafa de vinho. Muitas vezes a imagem apresentada era curva e em várias posições diferentes, fatores que dificultaram a padronização do processo de extração de dados em imagens. É importante referir que muitos rótulos não apresentam informações relevantes aos consumidores, tais como sobre o teor alcoólico ou a região vitivinícola, por exemplo, devido a legislação ser diferente entre os países produtores.

As fontes utilizadas na extração de dados e as imagens não marcadas selecionadas foram organizadas em formato *jpeg* no tamanho de resolução de 480×640 pixels e publicadas no repositório oficial do projeto WD4WU.

4. Verificação e Validação dos Dados

Um extenso trabalho de pré-processamento foi realizado ao longo de quatro meses sobre os dados coletados, o que produziu mais de 3.000 linhas de código-fonte e levou ao descarte de milhões de dados não aprovados. Diferentes formatos foram verificados e validados para aproximar ao máximo da realidade os conjuntos de dados produzidos.

Para aumentar a consistência entre os dados e permitir a construção de sequências que possam ser exploradas na experimentação científica, procurou-se limitar o número de utilizadores que avaliaram apenas um ou alguns poucos vinhos e os itens que receberam pouquíssimas avaliações. A formação de sequências de dados pode ser particularmente útil para sistemas de recomendação, por exemplo, no tratamento de dados históricos [Herlocker et al. 2004; Koren et al. 2009; Smith and Linden 2017]. Os dados coletados foram pré-processados com base em duas limitações quantitativas:

- a) Apenas itens de vinhos que apresentaram cinco ou mais avaliações;
- b) Apenas avaliações realizadas por utilizadores com cinco ou mais avaliações de vinhos entre os vinhos selecionados.

Após o corte quantitativo mínimo, colunas de controle adicionais foram criadas e diferentes regras de negócios baseadas em conhecimentos específicos no domínio do vinho foram exaustivamente executadas para validar os atributos. Também foram utilizados dicionários com palavras-chave do domínio do vinho. Isso envolveu pesquisa específica e processamento intensivos. Somente quando uma instância alcançava 100% positivo em todas as colunas de controle é que ela era finalmente validada como acionável por máquina [Dumontier et al. 2016]. As principais fontes de conhecimento de referência sobre o domínio de vinhos utilizadas foram: Wine Folly [Puckette and Hammack 2015], A Bíblia do Vinho [Macneil 2015], Lexicon Enciclopédia do Vinho [Wine Lexicon 2022] e *OIV-International Organization of Vine and Wine* [Juban 2022].

As avaliações no formato 5 estrelas foram validadas no período de 10 anos entre 2012 e 2021, pois poucas avaliações foram encontradas antes do ano de 2012. Em relação às safras avaliadas, foram consideradas apenas aquelas encontradas na faixa de produção dos últimos 70 anos aproximadamente (entre 1950 e 2021). Foram também validados vinhos que, para além das colheitas anuais regulares, utilizaram variedades colhidas em diferentes anos, identificados com o acrônimo “N.V.” para “*non-vintage*”.

Os utilizadores validados na base de dados X-Wines foram anonimizados por uma sequência numérica aleatória para não serem explicitamente identificados, preservando a sua identidade, uma vez que se torna mais importante para a investigação registar as suas classificações. As principais atividades de pré-processamento realizadas eletronicamente pelos algoritmos de classificação e consistência foram:

- a) Remoção de registros duplicados;
- b) Remoção de registros com alguma inconsistência de data na relação entre vinhos-avaliações-utilizadores. Assim, o erro de calendário foi eliminado;
- c) Adaptação do tipo de vinho à classificação internacional [Juban 2022];
- d) Ajuste estatístico de arredondamento em valores numéricos para a classificação 5 estrelas mais utilizada. Por exemplo, 3,9 para 4,0 e 4,6 para 4,5 na coluna de avaliação. Isso foi necessário em menos de 0,05% das instâncias encontradas, com apenas 10.298 ajustes referentes a 463 vinhos;
- e) Opcionalmente, os URLs obtidos referentes aos websites de vinícolas ou relacionados a elas foram mantidos quando sua origem foi validada.

Ao verificar a origem dos dados obtidos, os websites das vinícolas e as fichas técnicas dos produtos receberam a pontuação mais alta num *ranking* de origem criado internamente, seguidos pelos dados obtidos nas plataformas especializadas em vinhos na Web aberta. Websites de terceiros, tais como o *e-commerce*, foram usados com pontuações mais baixas no *ranking* de origem próprio. Os *blogs* foram evitados.

5. Conjuntos de Dados Produzidos

A base de dados X-Wines foi construída por processos em conformidade com os princípios orientadores FAIR: localizáveis, acessíveis, interoperáveis e reutilizáveis [Dumontier et al. 2016], permitindo que os recursos computacionais localizem e utilizem automaticamente os dados. Neste sentido, foram utilizados metadados com um identificador único e persistente para aumentar a rastreabilidade. Foi utilizada uma

linguagem comum e aplicável para representar o conhecimento no domínio do vinho em língua Inglesa, a fim de estabelecer a interoperabilidade dos dados e atributos relevantes do vinho. As sequências de avaliações de preferência realizadas por utilizadores anonimizados foram validadas com consistência para a melhor reutilização dos dados.

X-Wines está disponível abertamente e publicada sob uma licença livre no repositório oficial do projeto WD4WU em três versões: **Full**, contendo todos os registros; **Slim**, contendo um por cento da amostra aleatória de instâncias e uma versão **Test** contendo apenas os cem vinhos verificados por amostragem com mil avaliações de preferência para a experimentação da base de dados produzida nesta investigação. A Tabela 1 apresenta o comparativo entre as diferentes versões.

Tabela 1. Detalhamento das versões distribuídas da base de dados X-Wines.

Versão	Vinhos	Tipos de vinhos	Países produtores	Utilizadores	Avaliações	Múltiplas avaliações Utilizador-vinho
Test	100	6	17	636	1.000	Não
Slim	1.007	6	31	10.561	150.000	Não
Full	100.646	6	62	1.056.079	21.013.536	Sim

Os dois arquivos que compõem a base de dados X-Wines são detalhados na publicação original. Eles mantêm a mesma estrutura de atributos listada abaixo e apenas o número de instâncias varia entre as suas diferentes versões.

Arquivo *XWines_100K_wines.csv*:

{*WineID, WineName, Type, Elaborate (Varietal ou Assemblage/Blend), Grapes, Harmonize* (harmonização com alimentos), *ABV* (% de álcool por volume), *Body, Acidity* (alta, média ou baixa), *Code* (acrônimo internacional do país ISO-3166), *Country, RegionID, RegionName, WineryID, WineryName, Website, Vintages*}.

Arquivo *XWines_21M_ratings.csv*:

{*RatingID, UserID* (número aleatório), *WineID, Vintage, Rating* (1-5), *Date*}

6. Teste de Acurácia da Base de Dados X-Wines

Para estimar a confiança dos dados obtidos na Web, um teste estatístico foi realizado. Uma amostra aleatória simples contendo 100 vinhos foi sorteada entre os vinhos validados com seus respectivos atributos. O sorteio foi organizado de forma estratificada considerando todos os seis tipos classificados e realizado por algoritmo em duas etapas:

- Cinco vinhos foram selecionados aleatoriamente para cada um dos seis tipos de vinhos encontrados, constituindo 30% da amostra;
- Setenta vinhos foram selecionados aleatoriamente, independentemente de qualquer atributo, constituindo 70% da amostra.

Dez atributos para cada vinho selecionado foram verificados individualmente num *benchmark* baseado em documento, em que um total de 1.000 informações foram confrontadas com dados apresentados oficialmente pelas vinícolas ou seus representantes. A cada verificação foi atribuída uma pontuação entre três condições possíveis: valor 1 se houve correspondência, 0,5 se estava parcialmente correto com a necessidade de algum ajuste ou 0 se não correspondeu aos dados encontrados. Os documentos, imagens e websites utilizados no *benchmark* baseado em documento estão disponíveis no repositório oficial. A Figura 2 ilustra os resultados analisados a seguir.

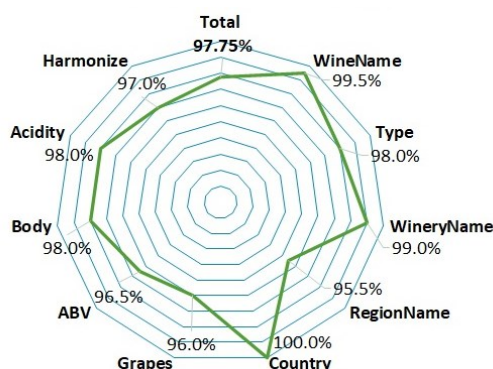


Figura 2. Resultado do benchmark baseado em documento nos 100 vinhos selecionados.

O resultado global obteve uma coincidência de 97,75% e desvio padrão de 4,15%. Este resultado foi encontrado através da média de todas as proporções de dados obtidos a partir da Web quando comparado com a ficha técnica oficial informada pelos produtores. Assim, com um intervalo de confiança de 95%, a assertividade média real situa-se entre 96,94% e 98,56%. Especificamente, os processos seguidos para a construção da base de dados X-Wines foram capazes de alcançar 100% de precisão no nome dos países de origem (atributo categórico *Country*). Resultados semelhantes com precisão ainda alta foram encontrados ao verificar os demais atributos. De acordo com [Juban 2022; Puckette and Hammack 2015], no processo de vinificação, as variedades e os valores de ABV podem variar consoante às colheitas, explicando algumas das discrepâncias encontradas nestes atributos (*Grapes* e *ABV*) obtidos na Web, possivelmente por se tratar de diferentes vintage. Os atributos corpo e acidez foram constituídos por uma grande quantidade de informações encontradas nas avaliações dos utilizadores e obtiveram um alto percentual de precisão. A harmonização com alimentos representada no atributo *Harmonize* é uma questão muito particular a ser avaliada, e foi considerado correspondente quando pelo menos uma indicação de paridade coincidiu. Os maiores ajustes foram necessários no atributo região vitivinícola, contendo o nome da região de origem do vinho. A região é geralmente apresentada como encontrada no rótulo (região de denominação), contudo os rótulos também podem exibir a região geográfica ou sub-região do país produtor.

7. Citações da Base de Dados X-Wines

No primeiro ano de sua publicação, a base de dados X-Wines [de Azambuja et al. 2023] tem recebido milhares de acessos e 1.253 *downloads* foram realizados nos arquivos que compõem as três diferentes versões. Conforme ilustrado na Figura 3, X-Wines foi

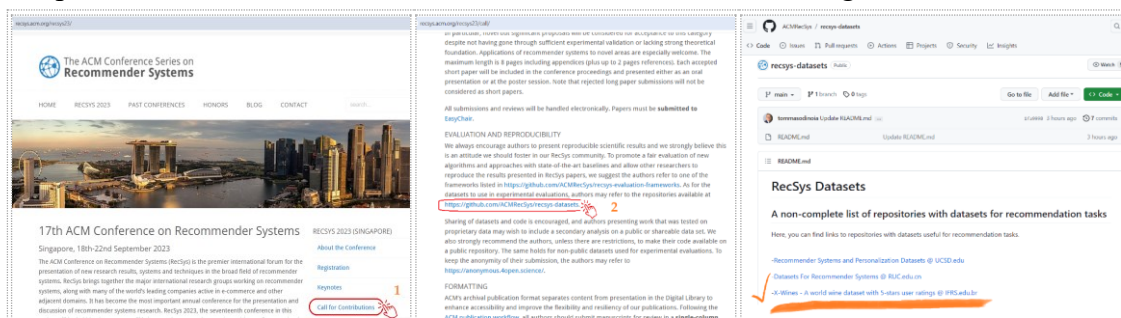


Figura 3. X-Wines indicada para experimentação científica na ACM/RecSys'2023 (disponível em: <https://recsys.acm.org/recsys23>. Acesso em: 28/11/2023)

selecionada por especialistas e recomendada para investigações e produção de trabalhos científicos no tema de sistemas de recomendação na ACM/RecSys'2023 – *The ACM Conference Series on Recommender Systems*, a maior conferência mundial no tema. Outras publicações com indexação e *citescore* relevantes também fizeram uso da base de dados X-Wines. As três principais são apresentadas abaixo, sendo duas em revista científica com fator de impacto (2022) de 3,1 e 2,6 e uma em conferência internacional:

- Com 4 autores gregos, em Gkikas et al. [2023] é apresentado um estudo que procura uma solução para o problema de *overfitting* ao classificar grandes conjuntos de dados. Combinando algoritmos de árvore de decisão e genético, o número de atributos selecionados é minimizado no treinamento do aprendizado de máquina para reduzir o risco de *overfitting*. São utilizados nos experimentos realizados dados filtrados de vinhos cultivados numa mesma região e derivados de, pelo menos, três cultivares diferentes de uvas.
- Com 15 autores gregos, em Papadakis et al. [2023] é apresentado um ecossistema *Message-in-a-Bottle* (MiB), que explora dados multidimensionais de múltiplas fontes para criar histórias envolventes e interativas em torno dos rótulos de vinhos. A plataforma MiB permite aos vitivinicultores fazer a curadoria de histórias sobre os seus vinhos com um nível extra de controle, especialmente para receber informações sobre caixas de garrafas não genuínas.
- Com 3 autores chineses, em He et al. [2023] é proposto um paradigma de aprendizado interpretável para prever as classificações de vinhos utilizando suas múltiplas características oriundas dos atributos. O estudo analisou a contribuição das características dos vinhos nas classificações realizadas por utilizadores, descobrindo que os fatores relacionados às condições climáticas tornaram-se menos influentes nas classificações realizadas nos vinhos nos últimos anos.

Para além das citações recebidas, muitos contatos com outros pesquisadores interessados no tema em várias áreas do conhecimento foram mantidos, tais como: computação e ciência de dados, engenharia de alimentos e ambiental, química, enologia, gastronomia, e outras áreas a partir de países como: Estados Unidos, Inglaterra, China, Índia, Áustria, França, Itália, Coreia do Sul, e outros além de Brasil e Portugal.

8. Conclusão

No atual cenário tecnológico de crescimento da inteligência artificial, especialmente com o uso do aprendizado de máquina, grandes volumes de dados tornam-se necessários. Este artigo apresentou uma breve caracterização da base de dados X-Wines no domínio de vinhos produzida no âmbito do projeto de pesquisa *Wine Data for Wider Use*. Foi verificada a qualidade dos dados publicados por meio de um *benchmark* baseado em documento e verificado o impacto na aceitação desta nova base de dados pela comunidade científica durante o primeiro ano de sua publicação oficial.

Preencher uma lacuna detectada na carência de dados pré-processados foi a principal motivação à construção de uma inédita base de dados de vinhos internacionais. Princípios éticos e científicos foram seguidos na construção desta base de dados especialmente voltada para sistemas de recomendação e aprendizado de máquina. Atualmente, X-Wines destaca-se no cenário internacional como uma base de dados de referência, sendo uma contribuição para a ciência de dados utilizada por milhares de

interessados e citada por outros pesquisadores na literatura, estando disponível abertamente para a livre utilização mais ampla.

Por fim, a maior contribuição verificada reside no trabalho de pré-processamento e na confirmação de utilidade destes dados para a ciência. Este trabalho é de suma importância na preparação de dados consistentes em volume massivo, sendo muitas vezes difícil, oneroso e dispendioso. Como trabalho futuro, a publicação de uma nova e inédita base de dados formada por sessões geradas por *feedbacks* implícitos obtidos na plataforma colaborativa Web poderá contribuir ainda mais em investigações.

Agradecimentos. Este trabalho foi apoiado pelo Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) – Brasil.

Referências

- Archana, K. and Saranya, K.G. (2020). “Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier”, In: *International Journal of Computer Science and Engineering* 7, 5, p. 1–4. [DOI](#).
- Artemenko, O., Kunanets, O., and Pasichnyk, V. (2017). “E-tourism recommender systems: a survey and development perspectives”, In: *ECONTECHMOD International Journal on Economics of Technology and Modelling Processes* 6, 2, p. 91–95. [DOI](#).
- de Azambuja, R.X., Morais, A.J., and Filipe, V. (2022). “Adaptive Recommendation in Online Environments”, In: *Distributed Computing and Artificial Intelligence, Vol. 2: 18th International Conference*, Springer International Publishing, p. 185–189. [DOI](#).
- de Azambuja, R.X., Morais, A.J., and Filipe, V. (2023). “X-Wines: A Wine Dataset for Recommender Systems and Machine Learning”, In: *Big Data and Cognitive Computing* 7, 1. [DOI](#).
- Baek, J.-W. and Chung, K.-Y. (2021). “Multimedia recommendation using Word2Vec-based social relationship mining”, In: *Multimedia Tools and Applications* 80, 26, p. 34499–34515. [DOI](#).
- Dumontier, M., Aalbersberg, I.J., Appleton, G., et al. (2016). “The FAIR Guiding Principles for scientific data management and stewardship”, In: *Scientific Data* 3, 1, 160018. [DOI](#).
- Fararni, K.A., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., and Sabri, A. (2021). “Hybrid recommender system for tourism based on big data and AI: A conceptual framework”, In: *Big Data Mining and Analytics* 4, 1, p. 47–55. [DOI](#).
- Geuens, S., Coussement, K., and De Bock, K.W. (2018). “A framework for configuring collaborative filtering-based recommendations derived from purchase data”, In: *European Journal of Operational Research* 265, 1, p. 208–218. [DOI](#).
- Gkikas, D.C., Theodoridis, P.K., Theodoridis, T., and Gkikas, M.C. (2023). “Finding Good Attribute Subsets for Improved Decision Trees Using a Genetic Algorithm Wrapper; a Supervised Learning Application in the Food Business Sector for Wine Type Classification”, In: *Informatics* 10, 3. [DOI](#).
- He, M., Jiang, W., and Gu, W. (2023). “Making Wines Smarter: Evidence from an Interpretable Learning Paradigm”, In: *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, p. 60–65. [DOI](#).
- Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. (2004). “Evaluating Collaborative Filtering Recommender Systems”, In: *ACM Transactions on Information Systems* 22, 1, p. 5–53. [DOI](#).

- Iwendi, C., Khan, S., Anajemba, J.H., Bashir, A.K., and Noor, F. (2020). “Realizing an Efficient IoMT-Assisted Patient Diet Recommendation System Through Machine Learning Model”, In: *IEEE Access* 8, p. 28462–28474. [DOI](#).
- Jaiswal, S., Kharade, T., Kotambe, N., and Shinde, S. (2020). “Collaborative Recommendation System For Agriculture Sector”, *ITM Web Conf.* 32, 03034. [DOI](#).
- Juban, Y. (2022). “*International Standard for the Labelling of Wines*”, OIV-International Organization of Vine and Wine, Paris.
- Koren, Y., Bell, R., and Volinsky, C. (2009). “Matrix Factorization Techniques for Recommender Systems”, In: *Computer* 42, 8, p. 30–37. [DOI](#).
- Kulkarni, N.H., Srinivasan, G.N., Sagar, B.M., and Cauvery, N.K. (2018). “Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique”, In: *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, p. 114–119. [DOI](#)
- Lacasta, J., Lopez-Pellicer, F.J., Espejo-García, B., Nogueras-Iso, J., and Zarazaga-Soria, F.J. (2018). “Agricultural recommendation system for crop protection”, In: *Computers and Electronics in Agriculture* 152, p. 82–89. [DOI](#).
- Macneil, K. (2015). “*The Wine Bible*”, Workman Publishing, New York.
- Papadakos, P., Chrysakis, I., Patkos, T., et al. (2023). “Message-in-a-bottle: engaging stories around sustainable and safe wine products”, In: *Discover Sustainability* 4, 1, 43. [DOI](#).
- Puckette, M. and Hammack, J. (2015). “*Wine Folly: The Essential Guide to Wine*”, Wine Folly LLC, New York.
- Sahoo, A.K., Pradhan, C., Barik, R.K., and Dubey, H. (2019). “DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering”, In: *Computation* 7, 2, 25. [DOI](#).
- Schafer, J.B., Konstan, J.A., and Riedl, J. (2001). “E-Commerce Recommendation Applications”, In *Data Mining and Knowledge Discovery* 5, 1, p. 115–153. [DOI](#).
- Shao, B., Li, X., and Bian, G. (2021). “A Survey of Research Hotspots and Frontier Trends of Recommendation Systems from the Perspective of Knowledge Graph”, In: *Expert Systems with Applications* 165, 113764. [DOI](#).
- Smith, B. and Linden, G. (2017). “Two Decades of Recommender Systems at Amazon.com”, In: *IEEE Internet Computing* 21, 3, p. 12–18. [DOI](#).
- Sperli, G., Amato, F., Mercurio, F., Mezzanzanica, M., Moscato, V., and Picariello, A. (2018). “A Social Media Recommender System”, In: *International Journal of Multimedia Data Engineering and Management (IJMDEM)* 9, 1, 36–50. [DOI](#).
- Wine Encyclopedia Lexicon in the World. (2022). Disponível em: <https://glossary.wein.plus>. Acesso em 26/01/2024.
- Yang, J., Wang, H., Lv, Z., et al. (2017). “Multimedia recommendation and transmission system based on cloud platform”, In: *Future Generation Computer Systems* 70, p. 94–103. [DOI](#).
- Yang, X., Dong, M., Chen, X., and Ota, K. (2020). “Recommender System-Based Diffusion Inferring for Open Social Networks”, In: *IEEE Transactions on Computational Social Systems* 7, 1, p. 24–34. [DOI](#).
- Zhang, Q., Lu, J., and Jin, Y. (2021). “Artificial Intelligence in Recommender Systems”, In: *Complex & Intelligent Systems* 7, 1, p. 439–457. [DOI](#).