ALGORITMOS DE APRENDIZADO DE MÁQUINA NAS HUMANIDADES DIGITAIS: UM MAPEAMENTO SUPORTE PARA REVISÃO DE LITERATURA.

João Victor Gontijo[1], Alan Keller Gomes[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia de Goiás – Câmpus Inhumas
Av. Universitária, s/nº, Vale das Goiabeiras, Inhumas-GO. CEP: 75400-000

jvgontijo16@gmail.com, alan.gomes@ifg.edu.br

Abstract. *The advent of the Big Data era, Digital Humanities (DH) have emerged as new area of research, transverse to the humanities (sociology, history, philosophy, linguistics, literature, art, etc.) and the computer sciences. In this paper we map the use of Machine Learning (ML) algorithms in DH from the search for papers in Google Scholar and Capes Portal. In order to provide support for the organization of systematic literature reviews, it is outlining an overview of the use of ML algorithms, models and techniques in DH. We found that the most commonly used ML algorithms used in DH are Clustering and Neural Networks, and, the least used are Association Rules and Classification Rules.*

Resumo. *Com o advento da era Big Data, as Humanidades Digitais (HD) surgem como nova área de investigação, transversal às humanidades tradicionais (sociologia, história, filosofia, linguística, literatura, arte, etc.) e as ciências da computação. Neste trabalho é realizado um mapeamento da utilização de algoritmos de Aprendizado de Máquina (AM) nas HD, a partir da busca de trabalhos nas bases científicas Google Acadêmico e Periódicos Capes. Com vistas a fornecer suporte para organização de revisões sistemáticas de literatura, é delineando um panorama da utilização de algoritmos, modelos e técnicas de AM nas HD. Foi possível constatar que os algoritmos de AM mais utilizados nas HD são Clusterização e Redes Neurais, e os menos utilizados são Regras de Associação e Regras de Classificação.*

1. Introdução

As epistemologias estabelecidas em todas as ciências são desafiadas pelo advento da era *Big Data*, trazendo consigo uma gama de oportunidades e desafios [Kitchin 2014]. As oportunidades surgem à medida que crescem as possibilidades de acesso a dados em larga escala, o que impulsiona o aprimoramento contínuo dos mais variados tipos de tecnologias da informação e seus usos nos mais variados espectros da vida humana [Demchenko 2013].

Sob a égide dessa avalanche de dados da era *Big Data*, surge a área de investigação denominada Humanidades Digitais [Cambridge 2017]. Essa área de pesquisa é transversal às humanidades tradicionais (como história, filosofia, linguística, literatura, arte, arqueologia, música e cultura) e às ciências da computação [Borgman 2009] [Evans and Rees 2012].

Na era *Big Data*, grandes desafios ainda devem ser enfrentados pelas HD: de um lado, tem-se a abundância de dados que retratam as realidades humanas, do outro lado, uma variedade de algoritmos, métodos e técnicas computacionais que podem ser usadas no processamento desses dados. Métodos, modelos e técnicas que implementam algoritmos com origem no Aprendizado de Máquina (AM) são capazes de apoiar as HD a superar esses desafios [Cambridge 2017].

Dentro da Ciência da Computação, o Aprendizado de Máquina é um subcampo da Inteligência Artificial que investiga técnicas computacionais capazes de adquirir automaticamente novas habilidades e conhecimentos. A premissa básica do AM é a construção de algoritmos que possam receber dados de entrada e extrair desses dados algum tipo de modelo ou padrão [Norvig and Russell 2014].

Os algoritmos de AM podem ser divididos em duas principais categorias, de acordo com o tipo da tarefa de aprendizado. Nas tarefas preditivas, o objetivo é predizer novos atributos ou classificar novos dados. Nas tarefas descritivas, o objetivo é aprender um modelo ou padrão que sumariza alguma relação entre os dados [Tan 2013] [Han 2011].

Quando categorizados de acordo com a técnica de aprendizado que utilizam, existem grupos de algoritmos de AM relativamente conhecidos que realizam as tarefas de predição (classificação) e descrição. Em alguns casos, dependendo da técnica, ambas tarefas de aprendizado podem ser realizadas [Tan 2013] [Han 2011].

O objetivo da presente pesquisa é realizar um mapeamento da utilização de algoritmos de Aprendizado de Máquina nas Humanidades Digitais, de forma específica, pretende-se identificar quantitativos de trabalhos que utilizam algoritmos, modelos e técnicas de Aprendizado de Máquina utilizados nas Humanidades Digitais.

Realizando uma busca nas bases de dados Google Acadêmico e Periódicos Capes com os termos (*strings* de busca) *"Digital Humanities" AND "Machine Learning"*, é retornado como resultado, respectivamente, 5.350 e 570 trabalhos científicos tratando do tema. É necessário pontuar que pesquisas que exploram temáticas ligadas a utilização do Aprendizado de Máquina nas Humanidades Digitais são relativamente recentes e tem um caráter altamente inovador e multidisciplinar [Cambridge, 2017].

Os resultados dessas buscas demonstram que é necessário realizar um mapeamento desses trabalhos, para que questões de pesquisa, como a que segue, seja respondida: *Quais são os algoritmos, modelos e técnicas de AM utilizados nas Humanidades Digitais?*

Para responder essa questão, são apresentados resultados de uma pesquisa exploratória e quantitativa nas bases de dados Google Acadêmico e Periódicos Capes, com o propósito de que mapeamento da utilização dos algoritmos de AM nas HD seja realizado e, além disso, sirva de suporte para a elaboração de revisões sistemáticas de literatura.

Uma contribuição da presente pesquisa é a constatação de que os algoritmos de AM mais utilizados nas HD são Clusterização e Redes Neurais, e os algoritmos menos utilizados Regras de Associação e Regras de Classificação.

A pesquisa aqui apresentada é fundamental para o empoderamento de profissionais e pesquisadores que atuam tanto nas Ciências da Computação quanto nas Humanidades Digitais. De forma específica, espera-se que esse empoderamento se dê por meio da compreensão do cenário utilização dos algoritmos de AM nas HD, e posteriormente, com a apropriação das formas de utilização desses algoritmos por parte desses profissionais e pesquisadores.

2. Referencial Teórico

Uma nova área de estudo e pesquisa denominada Humanidades Digitais surge com a evolução de Big Data. Essa nova área de investigação é transversal a áreas como as humanidades tradicionais (como história, filosófica, linguística, literatura, arte, arqueologia, música e cultura), ciências da computação [Borgman 2009].

Dentro das Ciências da Computação, o Aprendizado de Máquina é um subcampo da Inteligência Artificial que investiga técnicas computacionais capazes de adquirir automaticamente novas habilidades e conhecimentos. A premissa básica do Aprendizado de Máquina é a constrição de algoritmos que possam receber dados de entrada e extrair desses dados algum tipo de modelo ou padrão [Norvig and Russel, 2014].

Os algoritmos de Aprendizado de Máquina podem ser classificados em duas principais categorias, de acordo com o tipo da tarefa de aprendizado. Nas tarefas preditivas ou de classificação, o objetivo é predizer novos atributos ou classificar novos dados. Nas tarefas descritivas, o objetivo é derivar um modelo ou padrão que sumariza algum tipo de relação entre os dados [Tan 2013].

Nas tarefas de predição podem ser utilizados algoritmos que empregam como técnicas de aprendizado Redes Neurais [Haykin 2001], Máquinas de Suporte Vetorial [Hwanjo 2011], Naive Bayes [Larsen 2005], Regressão Linear [Lou 2012], Algoritmos Genéticos [Linden 2008]. Em tarefas de descrição, podem ser empregadas técnicas como Regras de Associação e Extração de Sequências [Adamo 2008], e Clusterização (Agrupamento de Dados) [Arabie and Soete, 1996].

Em alguns casos, dependendo da técnica empregada no algoritmo, ambas tarefas de aprendizado podem ser realizadas [Han et. al, 2011]. Algoritmos que empregam técnicas de aprendizado como Árvores de Decisão (Rezende, 2003), Regras de Classificação [Furnkranz et. al. 2012] e K-NN (K-*Nearest Neighbours*) [Fonseca 2008] realizam ambas tarefas de aprendizado (Han et. al, 2011; Tan et al., 2013).

Tomando como referência os trabalhos de Foster (2016) e Amaral (2016), foram considerados na presente pesquisa Redes Neurais, Máquina de Suporte Vetorial, Regras de Classificação, Regras de Associação e Clusterização (Agrupamento de Dados) em estudos das Humanidades Digitais.

3. Materiais e Métodos

Os materiais utilizados na pesquisa são bases de dados de publicações científicas Google Acadêmico e Periódicos Capes, os trabalhos retornados nas buscas realizadas nessas bases, além de editores de textos, dentre eles LibreOffice[1], Microsoft Word[2] e Latex[3], utilizados na elaboração das comunicações. As bases de publicações científicas mencionadas foram escolhidas devido a sua capacidade de indexar outras bases tais como ACM DL, Science Direct, IEEExplorer e Scorpus.

[1] https://pt-br.libreoffice.org/
[2] https://support.office.com/pt-BR/word
[3] https://www.latex-project.org/

A pesquisa realizada neste trabalho é do tipo exploratória e quantitativa. O mapeamento é realizado a partir da busca de trabalhos que constam nas bases de publicações científicas com vistas a organizar revisões sistemáticas de literatura, tomando como referência o método apresentado por [Sampaio and Mancini 2007].

A partir da constatação de uma grande variedade de trabalhos resultantes da busca com os termos *"Digital Humanities" AND "Machine Learning"*, novos termos foram elaborados para refinar os resultados da busca. As buscas e os acessos aos trabalhos retornados nas buscas foram realizados com identificação, por meio do acesso remoto CAFé[4], disponível no portal de periódicos CAPES.

4. Resultados e Discussão

4.1 Refinamento dos termos utilizados nas buscas

Na primeira etapa de buscas foram utilizados termos em português e sua respectiva tradução em inglês. Na segunda etapa, foi feito um refinamento apenas dos termos em inglês.

Tabela 1 – Termos de Busca e seus Refinamentos

Termos de Busca	Refinamento dos Termos (em inglês)
"Humanidades Digitais" AND "Redes Neurais"	*"Digital Humanities" AND "Neural Network Classifier"*
	"Digital Humanities" AND "Neural Network Model"
	"Digital Humanities" AND "Neural Network Algorithm
Humanidades Digitais" AND "Máquina de Suporte Vetorial"	*"Digital Humanities" AND "SVM[5] Classifier"*
	"Digital Humanities" AND "SVM Model"
	"Digital Humanities" AND "SVM Algorithm"
Humanidades Digitais" AND "Clusterização"	*"Digital Humanities" AND "Clustering Technique"*
	"Digital Humanities" AND "Clustering Model"
	"Digital Humanities" AND "Clustering Algorithm"

Na Tabela 1 são apresentados os termos de busca em português e o respectivo refinamento dos termos em inglês. Além desses termos, foram realizadas buscas com "Humanidades Digitais" AND "Regras de Classificação" e "Humanidades Digitais" AND "Regras de Associação", em português e inglês. Como o número de trabalhos retornados para os termos em português foi 1 e 0, respectivamente, no Google Acadêmico e no Periódicos Capes, optou-se por não refinar esses termos de busca.

4.2 Resultados das buscas com termos Humanidades Digitais e Redes Neurais

Tabela 2 – Resultados das buscas com termos Humanidades Digitais e Redes Neurais

String de busca	Site	Quantidade de resultados
"Humanidades Digitais" AND "Redes Neurais"	Google Acadêmico	7 resultados
"Digital Humanities" AND "Neural Networks"	Google Acadêmico	1.070 resultados
"Humanidades Digitais" AND "Redes Neurais"	Periódicos Capes	16 resultados
"Digital Humanities" AND "Neural Networks"	Periódicos Capes	391 resultados

[4] Disponível em https://www.periodicos.capes.gov.br/?option=com_plogin&ym=3&pds handle=&calling_ system=primo&institute=CAPES&targetUrl=http://www.periodicos.capes.gov.br&Itemid=155&pagina=CAFe acesso em 12/08/2019.

[5] Sigla para *Support Vector Machine*

Foram realizadas buscas com os termos "Humanidades Digitais" AND "Redes Neurais" e *"Digital Humanities" AND "Neural Networks"* tanto na base Google Acadêmico quanto Periódicos Capes. Os resultados são apresentados na Tabela 2.

É possível observar que o número de trabalhos publicados em português é bem menor que o número de trabalhos publicados em inglês. Em português foram encontrados 7 trabalhos no Google Acadêmico e 16 trabalhos nos Periódicos Capes. Como o número de trabalhos em inglês é grande, foi necessário refinar a busca com os termos *"Digital Humanities" AND "Neural Network Classifier"*, *"Digital Humanities" AND "Neural Network Model"* e *"Digital Humanities" AND "Neural Network Algorithm"*. Na Tabela 3 são apresentados os resultados dessas buscas.

Tabela 3 – Resultados do Refinamento das buscas com termos Humanidades Digitais e Redes Neurais

String de busca	Site	Quantidade de resultados
"Digital Humanities" AND "Neural Network Classifier"	Google Acadêmico	32 resultados
"Digital Humanities" AND "Neural Network Model"	Google Acadêmico	106 resultados
"Digital Humanities" AND "Neural Network Algorithm"	Google Acadêmico	8 resultados
"Digital Humanities" AND "Neural Network Classifier"	Periódicos Capes	1 resultado
"Digital Humanities" AND "Neural Network Model"	Periódicos Capes	4 resultados
"Digital Humanities" AND "Neural Network Algorithm"	Periódicos Capes	0 resultados

Comparando os quantitativos da Tabela 2 com a Tabela 3, de 1.070 resultados da busca no Google Acadêmico com termos *"Digital Humanities" AND "Neural Networks",* o refinamento dos termos para*"Digital Humanities" AND "Neural Network Classifier"*, *"Digital Humanities" AND "Neural Network Model"* e *"Digital Humanities" AND "Neural Network Algorithm"* fez o número de trabalhos retornados cair para 146. Essa diferença aponta que 924 trabalhos não foram alcançados pela busca com o refinamento dos termos.

Ainda comparando os quantitativos da Tabela 2 com a Tabela 3, para as buscas realizadas na base Periódicos Capes, é possível observar que de 391 trabalhos retornados pela busca com termos *"Digital Humanities" AND "Neural Networks*", apenas 5 trabalhos foram retornados pela busca com refinamento dos termos, ou seja, 386 trabalhos não foram alcançados com o uso dos termos *"Digital Humanities" AND "Neural Network Classifier"*, *"Digital Humanities" AND "Neural Network Model"* e *"Digital Humanities" AND "Neural Network Algorithm".*

Os resultados apresentados na Tabela 3 apontam para uma predominância de trabalhos que usam o termo "*model*" para indicar a utilização de Redes Neurais nas HD.

4.3 Resultados das buscas com Humanidades Digitais e Máquina de Suporte Vetorial

Tabela 4 – Resultados das buscas com termos Humanidades Digitais e SVM

String de busca	Site	Quantidade de resultados
"Humanidades Digitais" AND "Máquina de Suporte Vetorial"	Google Acadêmico	0 resultados
"Digital Humanities" AND "Support Vector Machine"	Google Acadêmico	765 resultados
"Humanidades Digitais" AND "Máquina de Suporte Vetorial"	Periódicos Capes	0 resultados
"Digital Humanities" AND "Support Vector Machine"	Periódicos Capes	67 resultados

Na Tabela 4 são apresentados os resultados das buscas com os termos "Humanidades Digitais" AND "Máquina de Suporte Vetorial" e *"Digital Humanities" AND "Suport Vector Machine"* tanto na base Google Acadêmico quanto Periódicos Capes. É possível observar que não há publicações que empregam os termos utilizados nas buscas em português em ambas bases de trabalhos científicos.

Como o número de trabalhos em inglês é grande, foi necessário refinar a busca utilizando os termos *"Digital Humanities" AND "SVM Classifier"*, *"Digital Humanities" AND "SVM Model"* e *"Digital Humanities" AND "SVM Algorithm"*. Os resultados para essas buscas são apresentados na Tabela 5.

Tabela 5 – Resultados do Refinamento das buscas com termos Humanidades Digitais e SVM

String de busca	Site	Quantidade de resultados
"Digital Humanities" AND "SVM Classifier"	Google Acadêmico	197 resultados
"Digital Humanities" AND "SVM Model"	Google Acadêmico	70 resultados
"Digital Humanities" AND "SVM Algorithm"	Google Acadêmico	41 resultados
"Digital Humanities" AND "SVM Classifier"	Periódicos Capes	20 resultados
"Digital Humanities" AND "SVM Model"	Periódicos Capes	6 resultados
"Digital Humanities" AND "SVM Algorithm"	Periódicos Capes	3 resultados

Comparando os quantitativos da Tabela 4 com a Tabela 5, de 765 resultados da busca no Google Acadêmico com termos *"Digital Humanities" AND "SVM",* o refinamento dos termos para *"Digital Humanities" AND "SVM Classifier"*, *"Digital Humanities" AND "SVM Model"* e *"Digital Humanities" AND "SVM Algorithm"* fez o número de trabalhos retornados cair para 308 resultados. Essa diferença aponta que 457 trabalhos não foram alcançados pela busca com o refinamento dos termos.

Considerando as buscas realizadas na base Periódicos Capes, ainda comparando os quantitativos da Tabela 4 e Tabela 5, é possível observar que de 67 trabalhos retornados pela busca com termos *"Digital Humanities" AND "Neural Networks*", 29 trabalhos foram retornados pela busca com refinamento dos termos, ou seja, 38 trabalhos não foram alcançados com o uso dos termos *Digital Humanities" AND "SVM Classifier"*, *"Digital Humanities" AND "SVM Model"* e *"Digital Humanities" AND "SVM Algorithm".*

Os resultados apresentados na Tabela 5 apontam para uma predominância de trabalhos que usam o termo "*classifier*" para indicar a utilização de Máquinas de Suporte Vetorial nas HD.

4.4 Resultados das buscas com termos Humanidades Digitais e Regras de Classificação

Os resultados das buscas com os termos "Humanidades Digitais" AND "Regras de Classificação" e *"Digital Humanities" AND "Classification Rules*" tanto na base Google Acadêmico quanto Periódicos Capes, são apresentados na Tabela 6.

Tabela 6 – Resultados das buscas com termos Humanidades Digitais e Regras de Classificação

String de busca	Site	Quantidade de resultados
"Humanidades Digitais" AND "Regras de Classificação"	Google Acadêmico	1 resultado
"Digital Humanities" AND "Classification Rules"	Google Acadêmico	56 resultados
"Humanidades Digitais" AND "Regras de Classificação"	Periódicos Capes	0 resultados
"Digital Humanities" AND "Classification Rules"	Periódicos Capes	8 resultados

É possível observar que apenas 1 publicação emprega os termos utilizados nas buscas em português no Google Acadêmico e, na base Periódicos Capes, nenhum trabalho foi retornado. Sendo assim, os resultados apresentados na Tabela 6 apontam para uma predominância de trabalhos que usam o termo em inglês na utilização de Regras de Classificação nas HD.

4.5 Resultados das buscas com termos Humanidades Digitais e Regras de Associação

Os resultados das buscas com os termos "Humanidades Digitais" AND "Regras de Associação" e *"Digital Humanities" AND "Association Rules*" tanto na base Google Acadêmico quanto Periódicos Capes, são apresentados na Tabela 7.

Tabela 7 – Resultados das buscas com termos Humanidades Digitais e Regras de Associação

String de busca	Site	Quantidade de resultados
"Humanidades Digitais" AND "Regras de Associação"	Google Acadêmico	1 resultado
"*Digital Humanities*" AND "*Association Rules*"	Google Acadêmico	154 resultados
"Humanidades Digitais" AND "Regras de Associação"	Periódicos Capes	0 resultados
"*Digital Humanities*" AND "*Association Rules*"	Periódicos Capes	15 resultados

É possível observar que apenas 1 publicação emprega os termos utilizados nas buscas em português no Google Acadêmico e, na base Periódicos Capes, nenhum trabalho foi retornado. Portanto, os resultados apresentados na Tabela 7 apontam para uma predominância de trabalhos que usam o termo em inglês para indicar a utilização de Regras de Associação nas HD.

4.6 Resultados das buscas com termos Humanidades Digitais e Clusterização

Os resultados das buscas com os termos "Humanidades Digitais" AND "Clusterização" e *"Digital Humanities" AND "Clustering"* tanto na base Google Acadêmico quanto Periódicos Capes são apresentados na Tabela 8. Na presente pesquisa optou-se por não usar o termo "Agrupamento" no lugar de "Clusterização", o que pode ser feito em trabalhos futuros.

Tabela 8 – Resultados das buscas com termos Humanidades Digitais e Clusterização

String de busca	Site	Quantidade de resultados
"Humanidades Digitais" AND "Clusterização"	Google Acadêmico	4 resultados
"*Digital Humanities*" AND "*Clustering*"	Google Acadêmico	3.950 resultados
"Humanidades Digitais" AND "Clusterização"	Periódicos Capes	0 resultados
"*Digital Humanities*" AND "*Clustering*"	Periódicos Capes	88 resultados

É possível observar que foi retornado na busca 4 publicações em português no Google Acadêmico e, na base Periódicos Capes, nenhum trabalho foi retornado. Sendo assim, os resultados apresentados na Tabela 8 apontam para uma predominância de trabalhos publicados em inglês que utilizam Clusterização nas HD.

Tabela 9 – Resultados do Refinamento das buscas com termos Humanidades Digitais e Clusterização

String de busca	Site	Quantidade de resultados
"*Digital Humanities*" AND "*Clustering Technique*"	Google Acadêmico	110 resultados
"*Digital Humanities*"" AND "*Clustering Model*"	Google Acadêmico	27 resultados
"*Digital Humanities*" AND "*Clustering Algorithm*"	Google Acadêmico	374 resultados
"*Digital Humanities*" AND "*Clustering Technique*"	Periódicos Capes	13 resultados
"*Digital Humanities*" AND "*Clustering Model*"	Periódicos Capes	0 resultados
"*Digital Humanities*" AND "*Clustering Algorithm*"	Periódicos Capes	62 resultados

Como o número de trabalhos em inglês é expressivo, foi necessário refinar a busca utilizando os termos *"Digital Humanities" AND "Clustering Technique"*, *"Digital Humanities" AND "Clustering Model"* e *"Digital Humanities" AND "Clustering Algorithm"*. Os resultados para essas buscas são apresentados na Tabela 9.

Comparando os quantitativos da Tabela 8 com a Tabela 9, de 3.950 resultados da busca no Google Acadêmico com termos *"Digital Humanities" AND "Clustering",* o refinamento dos termos para *"Digital Humanities" AND "Clustering Technique"*, *"Digital Humanities" AND "Clustering Model"* e *"Digital Humanities" AND "Clustering Algorithm"* fez o número de trabalhos retornados cair para 511 resultados. Essa diferença aponta que 3.439 trabalhos não foram alcançados pela busca com o refinamento dos termos.

Considerando as buscas realizadas na base Periódicos Capes, ainda comparando os quantitativos da Tabela 8 e Tabela 9, é possível observar que de 88 trabalhos retornados pela busca com termos *"Digital Humanities" AND "Clustering*", 78 trabalhos foram retornados pela busca com refinamento dos termos, ou seja, 10 trabalhos não foram alcançados com o uso dos termos *"Digital Humanities" AND "Clustering Technique"*, *"Digital Humanities" AND "Clustering Model"* e *"Digital Humanities" AND "Clustering Algorithm"*.

Os resultados apresentados na Tabela 9 apontam para uma predominância de trabalhos que usam o termo "*Algorithm*" para indicar a utilização de Clusterização nas HD.

4.7 Discussão dos Resultados

Os resultados apresentados mostram que há uma carência de publicações em português e evidenciam que as comunicações científicas escritas em inglês são predominantes.

Tabela 10 – Sumarização dos Resultados obtidos via Google Acadêmico

Aprendizado de Máquina	**Resultados sem Refinamento (termos em inglês)**	**Termo Predominante (refinamento em inglês)**	**Resultados com Refinamento**
Redes Neurais	1.070	*Model*	106
Máquina de Suporte Vetorial	765	*Classifier*	197
Regras de Classificação	56	-	-
Regras de Associação	154	-	-
Clusterização	3.950	*Algorithm*	374

Na Tabela 10 é apresentada uma sumarização dos resultados obtidos nas buscas com termos em inglês para a utilização dos algoritmos de AM nas HD via Google Acadêmico. É possível observar que, para tarefas de predição, Redes Neurais são as mais utilizadas; e para tarefas descritivas, Clusterização é mais utilizada.

É interessante observar que o termo "*Model*" é predominante para Redes Neurais, enquanto o termo "*Classifier*" é predominante para Máquina de Suporte Vetorial, e o termo "*Algorithm*" é predominante para Clusterização. Esses resultados podem contribuir significativamente na definição dos termos adequados em buscas de trabalhos para uma revisão sistemática de literatura.

Tabela 11 – Sumarização dos Resultados obtidos via Periódicos Capes

Aprendizado de Máquina	**Resultados sem refinamento (termos em inglês)**	**Termo Predominante (refinamento em inglês)**	**Resultados com Refinamento**
Redes Neurais	391	*Model*	4
Máquina de Suporte Vetorial	67	*Classifier*	20

Regras de Classificação	8	-	-
Regras de Associação	15	-	-
Clusterização	88	*Algorithm*	62

É possível constatar, ainda na Tabela 10, que os algoritmos de Clusterização são mais utilizados nas HD que os demais. Em seguida, os mais utilizados são os algoritmos de Redes Neurais. Além disso, é possível ainda observar os algoritmos menos utilizados destacando-se Regras de Associação e Regras de Classificação.

Na Tabela 11 é apresentada uma sumarização dos resultados obtidos nas buscas com termos em inglês via Periódicos Capes, para mapear a utilização dos algoritmos de AM nas HD. Pode ser observado que, para tarefas de predição, Redes Neurais são as mais utilizadas nas HD; e para tarefas descritivas, Clusterização é mais utilizada nas HD, corroborando os resultados obtidos nas buscas via Google Acadêmico. Observando-se os algoritmos menos utilizados, destacam-se novamente Regras de Associação e Regras de Classificação.

É interessante observar novamente que o termo "*Model*" é predominante para Redes Neurais, enquanto o termo "*Classifier*" é predominante para Máquina de Suporte Vetorial, e o termo "A*lgorithm*" é predominante para Clusterização.

5. Conclusão

Pesquisas que exploram temáticas ligadas a utilização do Aprendizado de Máquina nas Humanidades Digitais são recentes e tem um caráter altamente inovador e multidisciplinar. O escopo do presente trabalho se limita em compreender quais algoritmos de AM são mais utilizados nas HD, não englobando o funcionamento desses algoritmos. Não foram consultadas diretamente bases como ACM DL, Science Direct, IEEExplorer e Scorpus pois, tanto Google Acadêmico quanto Periódicos Capes indexam trabalhos dessas bases.

Uma importante contribuição do mapeamento realizado nesta pesquisa é a constatação de que os algoritmos de AM mais utilizados nas HD são Clusterização e Redes Neurais, com a predominância de trabalhos publicados em inglês. Os algoritmos menos utilizados são Regras de Associação e Regras de Classificação, também com predominância de publicações em inglês.

O mapeamento produzido mostra uma extensa utilização dos principais algoritmos de AM nas HD, tais como, Redes Neurais, Máquina de Suporte Vetorial e Clusterização (Agrupamento de Dados). Destaca-se que, nem sempre, os trabalhos reportados utilizam o termo "A*lgorithm*" para se referir ao Aprendizado de Máquina empregado no estudo das HD. Termos como "*Model*" e "*Classifier*" também são empregados.

É importante observar que a base de trabalhos científicos Google Acadêmico é mais abrangente na indexação dos trabalhos que utilizam AM nas HD, em detrimento ao portal Periódicos Capes. Existe a possibilidade de que o Google Acadêmico aponte para os mesmos trabalhos que Periódicos Capes, e vice-versa. Essa hipótese pode ser investigada no futuro.

Referências Bibliográficas

ADAMO, J. (2012). **Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms**. Springer Science & Bus. Media ed., ISBN 1461300851, 2012.

AMARAL, F. (2016). **Introdução à Ciência de Dados: mineração de dados e big data**. Elsevier / Alta books. 2016.

ARABIE, P. and SOETE, G. (1996). **Clustering and Classification**. World Scientific ed., ISBN 9810212879, 1996.

BORGMAN, C. L. (2009). **The digital future is now: a call to action for the humanities.** Digital Humanities Quarterly (DHQ), *3*(4), 1–30.

CAMBRIDGE U. (2017). **CDH: Cambridge Digital Humanities.** University of Cambridge, 2017. Disponível em https://www.digitalhumanities.cam.ac.uk/, acesso em 08/05/2018.

DEMCHENKO, Y. (2013). **Addressing Big Data issues in scientific data infrastructure**. International Conference on Collaboration Technologies and Systems (CTS) 2013.

EVANS, L. and REES, S. (2012). **An interpretation of digital humanities.** In D. M. Berry (Ed.), Understanding digital humanities (pp. 21–41). Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2012.

FONSECA, E. (2008). **k-RNN: k-relational nearest neighbour algorithm**. In Proceedings of the 2008 ACM Symp. on Applied Computing (SAC '08). ACM, NY, USA, 944-948.

FOSTER, I. (2016). **Big Data and Social Science: A Practical Guide to Methods and Tools**, Chapman and Hall/CRC Press, 2016. ISBN 9781498751407.

FURNKRANZ, J. (2012) **Foundations of Rule Learning**. Cognitive Technologies, Springer Science & Business Media ed., ISBN 3540751971, 2012.

HAN J. (2011). **Data Mining: Concepts and Techniques**. Ed. 3. ISBN 0123814804.

Elsevier, 2011.

HAYKIN (2001). **Redes Neurais** - 2ed, ISBN 8573077182, Bookman Companhia Ed, 2001.

HWANJO, Y. (2011). **Exact indexing for support vector machines**. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (SIGMOD '11). ACM, New York, NY, USA, 709-720.

KITCHIN, R. (2014). **The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences**. London: Sage. ISBN:978-1-4462-8747-7.

LARSEN, K., (2005). **Generalized Naive Bayes Classifiers**. SIGKDD Explor. Newsl. 7, 1 (une 2005, 76-81. DOI=http://dx.doi.org/10.1145/1089815.1089826.

LINDEN, R. (2008). **Algoritmos Genéticos** (2a edição), ISBN 8574523739, Brasport, 2008.

LOU, Y. (2012). **Intelligible models for classification and regression**. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). ACM, New York, NY, USA, 150-158.

NORVIG, P. and RUSSELL, S. (2014). **Inteligência Artificial: Tradução da 3a Edição**. ISBN 8535251413, Elsevier Brasil, 2014.

REZENDE, S. O. (2003), **Sistemas inteligentes: fundamentos e aplicações**, Ed. Manole Ltda, ISBN 8520416837, 2003.

SAMPAIO, R.F. and MANCINI, M.C. (2007) Estudos de revisão sistemática. **Revista Brasileira de Fisioterapia**, ISSN 1413-3555, São Carlos, v. 11, n. 1, p. 83-89, 2007.

TAN, P. (2013). **Introduction to Data Mining:** What's New in Computer Science Series. ISBN 0133128903. Pearson, 2013.