

Domestic Violence Data Analysis for Machine Learning Use: A Systematic Mapping

Alfeu Buriti Pereira Júnior

CESAR – Recife Center for Advanced Studies and Systems
Recife, Brazil
abpj@cesar.school

Francisco Icaro do Nascimento Ribeiro

CESAR - Recife Center for Advanced Studies and Systems
Recife, Brazil
finr@cesar.org.br

***Abstract.** There has been a strong interest in machine learning techniques in recent years to help with data analysis and knowledge extraction from various fields. Domestic violence is a social and human health problem, where the use of technology should be very beneficial to support various decision-making. This article is designed to contribute and give an overview of the current state of publications on the subject and what data is most relevant to use in this type of analysis. In the end, 38 publications were selected to be studied from a total of 6,235 in four internationally recognized digital scientific libraries.*

***Resumo.** Existe um interesse bastante acentuado nos últimos anos pelas técnicas de aprendizagem de máquina para ajudar na análise de dados e extração de conhecimento das mais diversas áreas. Violência Doméstica é um problema social e de saúde humana, em que a utilização de tecnologia deve ser muito benéfica para apoio as diversas tomadas de decisões. Este artigo foi elaborado para contribuir e dar uma visão geral do estado atual das publicações realizadas a respeito do assunto e quais dados são mais relevantes para este tipo de análise. Ao final, 38 publicações foram selecionadas para serem estudadas de um total de 6235 em quatro bibliotecas digitais científicas reconhecida internacionalmente.*

1. Introdução

Ferramentas de aprendizagem de máquina e avaliação de risco estão sendo cada vez mais usadas para orientar decisões sobre supervisão e tratamento de agressores de violência doméstica [van der Put et al. 2019]. Um grande número de mulheres relatou a experiência de violência por parceiro íntimo (VPI) e é de extrema importância identificar possíveis fatores que precipitem a VPI incorporando-os nos protocolos policiais para avaliar o risco da VPI [Sanz-Barbero et al. 2018]. Alguns profissionais, como policiais, são obrigados a prevenir comportamentos violentos, como a violência por parceiro íntimo (VPI), para isso, utilizam ferramentas atuariais destinadas a estimar o risco de ocorrência de novas violências após uma queixa prévia (reincidência policial), levando em conta os indicadores de risco e proteção que podem observar, apesar de não serem especialistas em avaliação comportamental [Roy Chowdhury et al. 2018].

A violência doméstica é um grave problema social e de saúde pública que ocorre em todos os países do mundo, com cerca de 30% das mulheres. A violência doméstica inclui uma série de violências físicas e sexuais, abusos emocionais e controle de comportamentos por parceiros íntimos [Matud, 2007]. Uma combinação de aprendizado de máquina e análise de dados disponíveis no processo pode ser capaz de fornecer previsões de risco de violência doméstica, oportunas e úteis. [Berk et al, 2016].

Políticas proativas de prisão e sentenciamento resultaram em um crescente e, em alguns casos, em um número esmagador de espancadores de cônjuges sendo processados através do sistema de justiça. Escassos recursos correcionais e de tratamento exigem que decisões difíceis sejam tomadas sobre o manejo de autores de violência doméstica. A avaliação do risco de agressão do cônjuge agora serve para informar os que tomam decisões sobre a sentença (por exemplo, liberação comunitária versus encarceramento), adequação de tratamento e intensidade de supervisão. Com esses avanços, pesquisadores e clínicos começaram a discutir o conteúdo apropriado e o processo de avaliação do risco do cônjuge. Nos últimos anos, tem havido um grande esforço para desenvolver instrumentos e procedimentos de avaliação de risco, teoricamente e cientificamente sólidos a este respeito [DUTTON e KROPP, 2000].

A temática de violência doméstica e a necessidade de que sejam realizadas análises dos dados para mensuração dos riscos de potencial agressor existe há bastante tempo, e somente nos últimos anos a tecnologia nos possibilitou buscar a resolução deste problema de forma automática, baseada no aprendizado de máquina. É muito importante um mapeamento dos dados que estão sendo utilizadas no uso do aprendizado de máquina para violência doméstica.

2. Protocolo Aplicado

Um mapa sistemático é um método definido de construir um esquema de classificação e estrutura em um campo de interesse [Petersen et al, 2008]. O estágio de análise de um estudo de mapeamento trata de sumarizar dados para responder à questão de pesquisa colocada. É improvável incluir técnicas de análises profundas, tais como meta-análise e síntese narrativa, mas sim totais e sumários baseados nos passos definidos por [Petersen et al, 2008].

O protocolo utilizado neste mapeamento foi dividido nas seguintes fases: Definição da questão de pesquisa, condução da revisão, triagem das publicações pelos critérios de inclusão e exclusão, classificação e extração dos dados, mapeando assim os estudos. O objetivo dessa revisão é identificar estudos relacionados aos dados utilizados para predição de violência doméstica em mulheres a fim de utilizá-los em ferramentas de aprendizado de máquina. A nossa questão principal a ser respondida com essa pesquisa é: Quais são os dados mais relevantes para realizar uma análise preditiva de violência doméstica que podem ser utilizadas em ferramentas de aprendizado de máquina?

2.1. Critérios de Inclusão e de Exclusão

Para a pesquisa, estudos que trataram do uso de aprendizado de máquina e de análise de dados no campo de violência doméstica contra mulheres foram considerados. Para nosso mapeamento todos os estudos retornados nas bases selecionadas passaram pela avaliação por título e por resumo. Não excluímos estudos baseados em datas, apesar de notarmos que nos anos mais recentes estudos mais relevantes foram obtidos.

Os seguintes critérios de exclusão foram utilizados: (a) Estudos não listados pelas palavras-chaves (termos de busca); (b) Estudos não focados na pergunta de pesquisa ou áreas afins (considerar perguntas secundárias); (c) Textos que não sejam estudos como: Notícias, editoriais, comentários, análises não empíricas, discussões, sumários, prefácios, mensagens (emails, cartas), tutoriais, pôsteres, painéis, teses, dissertações, apresentações (keynotes), vídeos, áudios; (d) Estudos que exibam apenas opiniões pessoais; (e) Estudos não acessíveis na Web (considerando possuir acesso as bases acadêmicas de pesquisas escolhidas). Os seguintes critérios de inclusão foram utilizados: (a) Estudos não removidos pelos critérios de exclusão; (b) Estudos que respondam uma ou mais das perguntas de pesquisa (considerar perguntas secundárias).

2.2 Estratégias de busca

Como estratégia de busca, procuramos as palavras mais utilizadas pela comunidade para definir a área de atuação que desejamos mapear. Seguindo esta premissa, chegamos a duas palavras-chaves de busca que compreendemos serem as mais interessantes para utilização na pesquisa do tema nas bases selecionadas. Segue tabela com os resultados:

Tabela 1. Quantidade de estudos encontrados em cada base

Base de dados	Total de Publicações "Domestic Violence" e "Predictive Analysis"	Total de Publicações "Domestic Violence" e "Machine Learning"	Total
ACM Digital Library	117	0	117
IEEE Xplore	1	5	6
SpringerLink	139	0	156
Science Direct	5913	43	5956
Total	6170	48	6235

No processo de extração de informações das bases, as palavras-chaves de busca foram usadas em duas etapas da pesquisa. Primeiro, foi utilizada as palavras de busca "**Domestic Violence**" e "**Predictive Analysis**" e os resultados foram armazenados. Em seguida, foi realizada a busca utilizando "**Domestic Violence**" e "**Machine Learning**" em todas as bases selecionadas. A busca foi realizada entre maio de 2019 e julho de 2019. Na tabela 1 temos a sumarização dos resultados encontrados.

2.3. Processo de seleção

Para seleção dos estudos, após as buscas realizadas nas bases escolhidas e com as palavras-chaves de busca formadas, obtemos os resultados descritos na tabela exibida na seção anterior. Com este resultado em mãos, foi realizada a execução da primeira etapa de seleção com a avaliação dos títulos dos estudos retornados para validar se o estudo satisfazia os critérios de inclusão, ou deveriam ser eliminados pelos critérios de exclusão. Nesta primeira etapa, a seleção foi realizada apenas pelo título, sem considerar o resumo ou conteúdo dos trabalhos.

Verificamos que muitos trabalhos avaliados nesta primeira etapa possuíam relação com violência de forma mais geral, outros eram relacionados à avaliação de conseqüências na saúde e/ou efeitos psicológicos nas vítimas de violência, principalmente crianças. Em muitos casos os títulos não eram objetivos, ou não traziam nenhuma menção à violência doméstica ou análise de dados ou ao aprendizado de máquina. Alguns traziam parcialmente o assunto desejado e foram selecionados para serem avaliados na próxima etapa da análise do resumo. Na etapa de avaliação de títulos de um total de 6235 trabalhos, foram selecionados 82 estudos que passaram para próxima etapa. Percebemos que a quantidade de artigos excluídos na primeira etapa foi significativa, uma vez que existia uma grande quantidade de artigos relacionados à saúde e questões sociais referentes à violência em geral ou análises de dados estatísticos a respeito da violência nas cidades, em comunidades, entre outros que não tinham nenhuma relação com a nossa área de interesse.

Na segunda etapa posterior a avaliação do título, os resumos das obras foram lidos, analisados e classificados e verificamos que 38 publicações que poderiam responder todo ou em parte a nossa questão principal ou as questões secundárias da nossa pesquisa.

Tabela 2. Quantidade de estudos selecionados em cada base

Base de dados	Estudos encontrados	Título 1ª Etapa	Resumo 2ª Etapa
ACM Digital Library	117	11	2
IEEE Xplore	6	6	2
SpringerLink	139	6	2
Science Direct	5956	59	32
Total	6235	82	38

A maioria dos estudos foi encontrada na biblioteca digital Elsevier (ScienceDirect) como pode ser observado na tabela 2. As bibliotecas digitais IEEE (IEEE Xplore), Springer e ACM (ACM Digital Library) tiveram ao final apenas dois estudos selecionados cada uma.

2.4. Avaliação de Qualidade

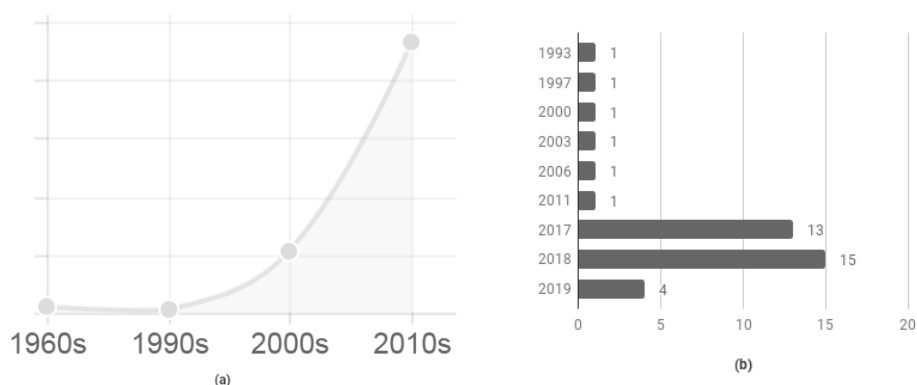
Os estudos analisados passaram por uma avaliação crítica onde os documentos completos foram analisados. Nesta etapa, foi classificada a qualidade do trabalho baseado na sua relevância para o nosso estudo. Para realizar esta classificação foi utilizado método de [Kitchenham et al. 2007] onde perguntas irão verificar a relevância, e o assunto dos artigos estudados. Sendo assim, as questões levantadas a respeito da qualidade do trabalho foram às seguintes: (1) O estudo analisa quais informações podem ser utilizadas como fatores de influência e ou causas da violência doméstica podendo ser utilizados no aprendizado de máquina? (2) O estudo é resultado de uma pesquisa? (3) O contexto do estudo é descrito adequadamente? (4) Os objetivos de pesquisa foram alcançados? (5) Os resultados da pesquisa foram adequadamente validados?

Após uma leitura criteriosa para análise de qualidade, 38 estudos foram selecionados para o estágio de extração e síntese de dados e foram, portanto, considerados como os estudos irão responder as questões levantadas neste mapeamento sistemático.

3. Resultados

Nosso mapa resultou em 38 artigos. Estes trabalhos são de diversas instituições de pesquisa e ensino de várias partes do mundo. Existem estudos de 1993 até 2019. Porém, ficou bastante claro através do gráfico “b” da figura 1 que o assunto, conforme já dito, tem tido um aumento de publicações nos últimos anos. Percebemos também, que o fato da maioria dos artigos serem recentes não foi intencional, pois não consideramos a data de publicação, mas pelo fato de uma maior quantidade de artigos de qualidade recentes a respeito do assunto, podemos comprovar isso pelo gráfico gerado na biblioteca ACM conforme observamos no gráfico “(a)” abaixo. Além disso, as palavras chaves mais utilizadas nos trabalhos foram domestic violence (13), Intimate partner violence (16), health (4), text mining (3), Machine learning (3) e demais também foram encontradas apenas uma ocorrência.

Figura 1. (a) Estudos publicados na área na biblioteca ACM (b) quantidade de publicações por ano



A resposta a questão levantada no nosso estudo é demonstrada em vários artigos. Segundo nosso mapeamento (8) publicações [Raj et al. 2018], [van der Put et al. 2019], [Wright, Hanlon, et al. 2019], [Sanz-Barbero et al. 2018], [Wawrzyniak et al. 2018], [Bowen, 2011], [López-Ossori et al. 2017], [Laeheem e Boonprakarn, 2017] obtiveram um total de 5 pontos, segundo os critérios de qualidade, significa que eles conseguiram responder a primeira questão que possuía mais importância para nosso estudo além de possuir uma boa qualidade em demais aspectos, (14) estudos tiveram 4 pontos e conseguiram ou responder parcialmente nossa questão ou havia algum outro quesito de qualidade não validada; Os demais estudos obtiveram entre 2 e 3 pontos, tendo uma qualidade razoável, mas que possuíam uma pesquisa que atingiu os objetivos desejados apesar de não terem sido aproveitados totalmente em nossa pesquisa.

Em relação ao conteúdo dos estudos podemos observar que tivemos uma grande quantidade de fatores que podem ser usados por ferramentas de aprendizado de máquina para predição de violência doméstica. Nesse sentido, (6) estudos [Roy Chowdhury et al. 2018],[Abbott, 1997], [Spencer et al. 2019], [Gilchrist et al. 2019], [López-Ossorio et al. 2017], demonstraram que o uso frequente ou abuso de bebidas alcoólicas é um fator de risco muito relevante para predição de violência doméstica. Em seguida temos as questões econômicas com (8) estudos (desigualdade econômica, dependência financeira, controle financeiro apenas pelo homem, empoderamento e inclusão financeira pela mulher, e por fim o desemprego e baixa renda) [Cools e Kotsadam, 2017], [Raj et al. 2018], [Willie, Stockman 2018]), [Gilchrist et al. 2019], [van der Put et al. 2019], [Willie et al. 2018], [Pietri e Bonnet, 2017]. As doenças e transtornos mentais também são fatores comuns entre os fatores de risco de VPI com (4) estudos [Gilchrist et al. 2019], [López-Ossorio et al. 2017], [Spencer et al. 2019], a baixa educação ou nível de escolaridade reduzido, apareceu em (4) estudos [Poelmans et al. 2011], [Roy Chowdhury et al. 2018], [Laeheem e Boonprakarn, 2017], [Willie et al. 2018]. Em seguida temos o abuso infantil na família de origem com (3) estudos [Spencer et al. 2019], [López-Ossorio et al. 2017], [Laeheem e Boonprakarn, 2017] além de testemunho de violência doméstica [Gilchrist et al. 2019], [Laeheem e Boonprakarn, 2017], [López-Ossorio et al. 2017].

Também com (3) estudos temos a idade média da mulher como fator de risco [Brignone e Gomez, 2017], [Poelmans et al. 2011], [Amrit, et al. 2017]. O fato do casal não ser casado, ou seja, casos de namoro, ex-parceiros, encontros casuais entre outros, também é fator de risco de violência doméstica e sua reincidência [Brignone e Gomez, 2017], [Sorenson e Spear, 2018], [Connolly, Huzurbazar, e Routh-McGee, 2000], por fim, a cultura patriarcal, onde o homem é o líder da família, além das normas sociais onde a aceitação da diferença de gênero onde o homem é considerado superior e a mulher deve seguir algumas preceitos sociais quem incluem questões de submissão da mulher em relação ao homem. [Gilchrist et al. 2019], [Willie et al. 2018], [Pietri e Bonnet, 2017], [Clark et al. 2018]. Enfim, curiosamente a temperatura do ambiente (clima) é um fator de risco de violência doméstica que deve ser considerado em ferramentas de aprendizado de máquina para predição de violência doméstica. [Sanz-Barbero et al, 2018]

4. Conclusão

O objetivo principal deste mapeamento foi identificar os principais estudos que para ajudar no uso da aprendizagem de máquina para análise preditiva de violência doméstica e seus avanços até o momento. Na fase de busca, 6235 estudos foram encontrados, sendo que 38 foram classificados estudos importantes após serem selecionados e os critérios de qualidade aplicados. A busca foi um tanto complexa, pois existem muitos estudos sobre violência de uma forma geral e muitos estudos sobre aprendizado de máquina em outras áreas.

Nos trabalhos que foram lidos algumas reflexões sobre fatores importantes para predição de violência doméstica são comuns em muitos estudos, como a importância de dados como idade, uso de álcool, maus tratos na infância, cultura local entre outros que são importantes elementos para análise de dados de violência doméstica. Também ficou demonstrado que é um campo importante da área social que possui relevante interesse pela comunidade acadêmica nos últimos anos. Percebemos também que este tipo de fenômeno ocorre em todo mundo e em grande quantidade, já que os textos reforçam a natureza comum deste tipo de violência, pois um terço das mulheres do mundo sofre violência doméstica [Matud, 2007], o que reflete na quantidade de publicações de todas as partes do mundo. Em relação ao trabalho futuro, uma análise de quais algoritmos de aprendizado de máquina é utilizada para dados de violência doméstica é recomendado.

Referências

- Abbott, J. (1997). Injuries and illnesses of domestic violence. *Annals of Emergency Medicine*, 29(6), 781–785.
- Amrit, C., Paauw, T., Aly, R., e Lavric, M. (2017). Identifying child abuse through text mining and machine learning. *Expert Systems with Applications*, 88, 402–418.
- Berk, R. A., Sorenson, S. B., e Barnes, G. (2016). Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions. *Journal of Empirical Legal Studies*, 13(1), 94–115.
- Bowen, E. (2011). An overview of partner violence risk assessment and the potential role of female victim risk appraisals. *Aggression and Violent Behavior*, 16(3).
- Brignone, L., e Gomez, A. M. (2017). Double jeopardy: Predictors of elevated lethality risk among intimate partner violence victims seen in emergency departments. *Preventive Medicine*, 103, 20–25.
- Clark, C. J., Ferguson, G., Shrestha, B., Shrestha, P. N., Oakes, J. M., Gupta, J., ... Yount, K. M. (2018). Social norms and women's risk of intimate partner violence in Nepal. *Social Science and Medicine*, 202, 162–169.
- Connolly, C., Huzurbazar, S., e Routh-McGee, T. (2000). Multiple parties in domestic violence situations and arrest. *Journal of Criminal Justice*, 28(3), 181–188.
- Cools, S., e Kotsadam, A. (2017). Resources and Intimate Partner Violence in Sub-Saharan Africa. *World Development*, 95, 211–230.
- DUTTON, D. G., e KROPP, P. R. (2000). A review of domestic violence risk instruments. *Trauma, Violence and Abuse*, SAGE.

- Gilchrist, G., Dennis, F., Radcliffe, P., Henderson, J., Howard, L. M., e Gadd, D. (2019). The interplay between substance use and intimate partner violence perpetration: A meta-ethnography. *International Journal of Drug Policy*, 65, 8–23.
- Laeheem, K., e Boonprakarn, K. (2017). Factors predicting domestic violence among Thai Muslim married couples in Pattani province. *Kasetsart J. Social Sciences*.
- López-Ossorio, J. J., González Álvarez, J. L., Buquerín Pascual, S., García, L. F., e Buela-Casal, G. (2017). Risk factors related to intimate partner violence police recidivism in Spain. *International Journal of Clinical and Health Psychology*.
- Matud, M. P. (2007). Dating Violence and Domestic Violence. *Journal Adolescent Health*
- Petersen, K., Feldt, R., Mujtaba, S., e Mattsson, M. (2008). *S. Mapping Studies in Software Engineering*.
- Pietri, M., e Bonnet, A. (2017). Analyses des représentations précoces et de la personnalité chez les victimes de violences conjugales. *Revue Européenne de Psychologie Appliquée*, 67(4), 199–206.
- Poelmans, J., Elzinga, P., Viaene, S., e Dedene, G. (2011). Formally analysing the concepts of domestic violence. *Expert Systems with Applications*, 38(4).
- Raj, A., Silverman, J. G., Klugman, J., Saggurti, N., Donta, B., e Shakya, H. B. (2018). Longitudinal analysis of the impact of economic empowerment on risk for intimate partner violence among married women in rural Maharashtra, India. *Social Science and Medicine*, 196(August 2017), 197–203.
- Roy Chowdhury, S., Bohara, A. K., e Horn, B. P. (2018). Balance of Power, Domestic Violence, and Health Injuries. *World Development*, 102, 18–29.
- Sanz-Barbero, B., Linares, C., Vives-Cases, C., González, J. L., López-Ossorio, J. J., e Díaz, J. (2018). Heat wave and the risk of intimate partner violence. *Science of the Total Environment*, 644, 413–419.
- Sorenson, S. B., e Spear, D. (2018). New data on intimate partner violence and intimate relationships: Implications for gun laws and federal data collection. *P. Medicine*
- Spencer, C. M., Stith, S. M., e Cafferky, B. (2019). Risk markers for physical intimate partner violence victimization: A meta-analysis. *Aggression and Violent Behavior*.
- Van der Put, C. E., Gubbels, J., e Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*.
- Wawrzyniak, Z. M., Borowik, G., Szczechla, E., Michalak, P., Pytlak, R., Cichosz, P., ... Perkowski, E. (2018). Relationships between Crime and Everyday Factors.
- Willie, T. C., Stockman, J. K., Perler, R., e Kershaw, T. S. (2018). Associations between intimate partner violence, violence-related policies, and HIV diagnosis rate among women in the United States. *Annals of Epidemiology*, 28(12), 881–885.
- Wright, E. N., Hanlon, A., Lozano, A., e Teitelman, A. M. (2019). The impact of intimate partner violence, depressive symptoms, alcohol dependence, and perceived stress on 30-year cardiovascular disease risk among young adult women.