Análise de dados sobre a vulnerabilidade da mulher brasileira para identificação de padrões

Dírlei R. de Mendonça Júnior[1], **Flávia Megumi Ohara**[1], **Huziel Corrêa Soares**[1], **Nádia Félix Felipe da Silva**[1]

[1]Instituto de Informática - Universidade Federal de Goiás - Goiânia, GO - Brasil

dirleijunior@discente.ufg.br, flavia.ohara@discente.ufg.br,

huzielcorrea@discente.ufg.br, nadia.felix@ufg.br

Abstract. *Data are valuable resources in today's world, they are fundamental knowledge generators for success in any area of knowledge, knowledge generates fast and assertive decision making. In turn, Data Science by definition is a combination of different algorithms, tools and machine learning principles, which are widely used in identifying hidden patterns in raw masses of data. This article aims to identify patterns of criminal occurrences related to violence against Brazilian women. In the pattern identification process, we used the KDD (Knowledge Discovery in Databases) knowledge extraction, in the datasets used in the preparation of this article. In the mining process, we used two datasets, the first refers to the occurrences recorded by Ligue 180, the second refers to the Police Occurrence Bulletins in the state of Goiás, and both datasets refer to the occurrences recorded in 2020. The results achieved, were obtained through the set of frequent items and association rules, with minimum and maximum rule parameters, generating the necessary support and trust indexes. Rules are important as antecedent and consequent attributes increase their precision and bring confidence to the process of generating the most reliable and assertive knowledge of their reality. Before the use of Data Science, data was nothing more than a raw mass or a bunch of data without relationships, making it infeasible for decision making. Thus, one of the main results achieved by this article is the knowledge of all aspects of how Brazilian women are vulnerable today, and how decision-makers (Government and Organized Society) in possession of this knowledge, can change or transform the reality of these victims, who are often left to their own devices.*

Resumo. *Os dados são recursos valiosos no mundo atual, são geradores de conhecimentos fundamentais para o sucesso em qualquer área do conhecimento, o conhecimento gera tomadas de decisões rápidas e assertivas. Por sua vez, a Ciência de Dados por definição é uma combinação de diferentes algoritmos, ferramentas e princípios de aprendizado de máquina, que são amplamente utilizadas na identificação de padrões ocultos em massas brutas de dados. O presente artigo possui o objetivo de identificar os padrões de ocorrências criminais relacionadas à violência contra a mulher brasileira. No processo de identificações de padrões utilizamos a extração de conhecimento KDD (Knowledge Discovery in Databases), nos datasets utilizados na elaboração deste artigo. No processo de mineração utilizamos dois datasets, o primeiro é referente às ocorrências*

registrada pelo Ligue 180, o segundo refere-se aos Boletins de Ocorrências Policiais no estado de Goiás, e ambos os datasets são referentes às ocorrências registradas ano de 2020. Os resultados alcançados foram obtidos através do conjunto de itens frequentes e regras de associações, com parâmetros de regras mínima e máxima, gerando os índices de suporte e confiança necessários. As regras são importantes à medida que atributos antecedentes e consequentes aumentam a sua precisão e trazem confiança para o processo de geração do conhecimento mais fidedigno e assertivo à sua realidade. Antes do emprego da Ciência de Dados, os dados não passavam de uma massa bruta ou de um amontoado de dados sem relações, tornando-se inviáveis na tomada de decisões. Assim, um dos principais resultados alcançados por esse artigo é o conhecimento de todas as vertentes de como as mulheres brasileiras nos dias atuais são vulneráveis, e como os tomadores de decisões (Governo e Sociedade Organizada) em posse desse conhecimento, poderão mudar ou transformar a realidade dessas vítimas, que na maioria das vezes estão relegadas à sua própria sorte.

1. Introdução

A história das brasileiras é marcada por décadas de violência, os tipos de violência sofrida por elas é diversificada, e neste estudo percebemos que as ameaças como assédios moral e emocional, e as violências física e financeira são os tipos mais comuns no dia a dia das brasileiras. Ao longo da nossa história, as mulheres foram denominadas como o sexo frágil, tornando a percepção de posse um fato comum, e a violência contra elas culturalmente aceita, principalmente porque vivemos em uma sociedade de cultura machista, e essa aculturação faz com que o problema seja relativizado. Essa relativização pela sociedade, faz com que o problema perdure por décadas sem uma solução a curto, a médio, ou a longo prazo. Diante desse cenário, podemos afirmar que não há um esforço concreto por parte das autoridades, ou da sociedade organizada para sanar, corrigir ou inibir a prática de violências contra as brasileiras, uma vez que a sociedade costuma tornar as vítimas de estupros por exemplo, como as responsáveis direta pelo ato de violência que sofreram, seja porque as vítimas vestiam saia ou short curta(o), ou faziam uso de blusa decotada. Segundo os dados do Painel de Violência contra as mulheres do Senado Federal Brasileiro [24], no ano de 2019, ocorreram 3,4 homicídios por grupo de 100 mil mulheres.

Durante o processo de obtenção de dados para realização deste estudo, percebe-se que a violência contra a mulher vai muito além do contexto doméstico e familiar, a violência abrange todas as camadas sociais da população brasileira, da favela ao palácio da justiça, as brasileiras são vítimas em potencial. O assassinato da juíza Patrícia Acióli, na cidade de Niterói/RJ em Agosto de 2011, e a impunidade dos envolvidos no crime dez anos após o ocorrido [12] é um exemplo de como a violência contra a mulher é tratada no Brasil. Infelizmente, os noticiários de violência contra as mulheres no Brasil não se resumiram ao caso da juíza Patrícia Acióli, outras brasileiras foram vítimas do médico ginecologista na cidade de Anápolis/GO [17], e do médico geneticista Roger Abdelmassih [13], quando ambos os médicos aproveitavam-se de suas condições médicas para violar a intimidade de suas pacientes.

A fim de oferecer um suporte mínimo às vítimas, no ano de 2003, a Câmara dos Deputados Brasileira, aprovou a Lei nº 10.714 que cria "A Central de Atendimento à Mulher", popularmente conhecida como Ligue 180 [16], o Ligue 180 foi implantado no

ano de 2005, e é utilizado como uma das fontes primárias de dados sobre o tema. O Ligue 180 presta uma escuta e acolhida qualificada para as brasileiras em situação de violência, o serviço registra e encaminha as denúncias de violência para os órgãos competentes. Buscamos identificar os padrões de violência com base nos registros referentes ao ano de 2020 do Ligue 180 e da SSP-GO (Secretaria de Segurança Pública do Estado de Goiás), os registros de ambos órgãos são as fontes primárias dos nossos estudos.

Em hipótese, podemos traçar um perfil que indicará se mulheres com baixa ou alta renda, com mais ou menos anos de escolaridade, brancas ou negras, estarão mais ou menos suscetíveis ao ato de violência, ou, se os seus agressores são pessoas do seu ciclo de convivência. Para traçar o perfil das vítimas, utilizamos neste estudo Regras de Associações em conjunto com Dados Frequentes. Através das técnicas aplicadas, conseguimos obter o perfil das principais vítimas: elas são mulheres entre 18 (Dezoito) e 58 (Cinquenta e oito) anos de idades, com 12 (Doze) anos ou menos de escolaridade, negras ou pardas, vivendo nos grandes centros urbanos, com uma renda familiar entre um e três salários mínimos mensais. Ao chegarmos neste perfil, não significará que as mulheres brancas ou indígenas por exemplo, com renda superior a três salários mínimos mensais, e com mais de 12 anos escolaridade não sejam vítimas em potencial. Essas mulheres estão sujeitas aos mesmos atos de violências, assim como aquelas inseridas dentro do perfil, mas, nos registros de ocorrências do Ligue 180 e da SSP-GO as ocorrências para esse grupo de mulheres é uma minoria. A fim de compreendemos melhor a situação de violência contra as mulheres, realizamos neste estudo uma análise dos registros de ocorrências da SSP-GO, esse olhar local visa identificar se aqui no estado de Goiás os resultados serão diferentes ou iguais ao obtido nacionalmente. Ressaltamos que os registros do Ligue 180 são denúncias, e os registros obtidos junto à SSP-GO são ocorrências materializadas, e para a nossa infelicidade, a realidade da terra do sertanejo não é muito diferente da realidade nacional. Quando obtivemos os padrões espaço-temporais destes crimes, foi possível observar que as mulheres goianas com idade entre 35 (Trinta e cinco) a 64 (Sessenta e quatro) anos, no intervalo de horários das 18h às 20h:59, no dia de Domingo, na área de atuação da RISP Goiânia (Região Integrada de Segurança Pública Goiânia), estão mais vulneráveis ao crime de ameaça física. Diante dos resultados obtidos, concluímos que o investimento em políticas públicas, principalmente aquelas voltadas para a área de educação, e a consequente obtenção de uma fonte de renda maior, tornará as mulheres menos vulneráveis aos atos de violências, principalmente no quesito da violência doméstica. A violência doméstica é o tipo de ocorrência que possui uma liderança absoluta das ocorrências registradas no Ligue 180 e SSP-GO no ano de 2020, ou seja, as mulheres com uma dependência financeira do pai, namorado ou cônjuge agressor, estarão expostas por mais tempo aos atos de agressões.

2. Trabalhos relacionados

Estudos sobre o uso de mineração de dados para reconhecimentos de padrões relacionados como metodologia são muito comuns e podem ser aplicados em diversas áreas onde se haja uma grande quantidade de dados, ou seja, onde os dados analisados façam parte de servidores de banco de dados com *Big Data*.

Porto (2019) [23] tenta descobrir os fatores que têm impacto no desempenho do Enem por meio do processo de mineração de dados e análise estatística. Os dados foram retirados do portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio

Teixeira. Foi possível retirar informações acerca do nível de conhecimento dos estudantes através da avaliação de desempenho de questão por questão em partes da prova do Enem, além de realizar comparativos sobre a distribuição de inscritos, sobre notas entre as regiões do Brasil, e sobre as notas separadas por gênero, raça/cor, renda e tipo de escola.

Já Peixoto *et al.* (2016) [22] busca mostrar como o fenômeno da violência contra a mulher se expressa: espacialmente, utilizando para isto o recorte do município do Rio de Janeiro, e demograficamente, buscando identificar o perfil da mulher que sofre violência.

Esse segundo trabalho relacionado, teria mais referências ao tema e seu clamor para com a agressão e vulnerabilidade da mulher e a violência doméstica que a mesma sofre, em específico às mulheres cariocas.

Para isso o trabalho se dividiu, essencialmente, em três partes. Na primeira parte foi feito um levantamento bibliográfico sobre o histórico de compreensão e combate da violência. A segunda parte consistiu em analisar a distribuição espacial das diferentes formas de violência contra a mulher (psicológica e moral; física; sexual e homicídio). Essa parte foi elaborada a partir de dados coletados no Departamento de Informática do Sistema Único de Saúde (DATASUS) da Secretaria Municipal de Saúde para cada uma dessas violências nos anos de 2010 a 2014 (por conta da disponibilidade dos dados e a proximidade ao Censo de 2010). Já na terceira parte foi relacionada a violência com dados socioeconômicos da população para se traçar o perfil dos grupos que sofrem maior violência. Para isso foram coletados também no DATASUS, para os mesmos anos, a informações de raça/cor, idade e escolaridade relacionadas à violência, e para estabelecer uma comparação foram coletadas, a partir do Banco Multidimensional Estatístico do IBGE, essas mesmas informações e mais a renda para o ano de 2010 (por conta do Censo Demográfico).

Sílva & Aníbal (2020) [21] realizam uma análise da predição da violência infantil por meio de Árvores de Decisão e Regras de Associação. Segundo o Fundo das Nações Unidas para a Infância (*United Nations International Children's Emergency Fund*, UNICEF), atualmente cerca de 300 milhões de crianças sofrem de diversos tipo de violência, podendo ser: psicológica, física, abuso sexual, ou negligência. Considerando a gravidade dessa problemática, adiciona-se que analisá-la torna-se difícil por alguns aspectos: existem diferentes definições de violência utilizadas em diversos contextos, não havendo padrões que facilitem o estudo; determinar qual é o tipo de violência que acontece em cada caso, podendo ser um ou mais; os informes das próprias vítimas e as estatísticas oficiais de violência não sempre possuem uma qualidade apropriada; entre outros. Ainda, cabe ressaltar que em tais levantamentos não é considerado o fato de que em muitas ocasiões as únicas pessoas que sabem que a violência acontece são as crianças e os agressores, sendo a situação invisível e tornando muito mais difícil a tarefa de previsão da agressão. O objetivo deste trabalho é gerar e avaliar modelos baseados em técnicas de aprendizado de máquina que possam estimar em que casos está acontecendo ou irá acontecer algum ato de violência infantil através de modelos que sejam representados por regras simples de entender por qualquer humano.

Outro trabalho relevante é de Moreira (2022) [20]. A evolução tecnológica certamente trouxe e vem trazendo grandes avanços sociais e econômicos para a geração

atual, entretanto, esse desenvolvimento também tem sido utilizado por alguns indivíduos para a prática de novos crimes ou, até mesmo, auxiliar na prática de antigos delitos. É o que acontece com a violência sexual contra crianças e adolescentes, um crime realizado ao longo dos anos sem auxílio da tecnologia que, nas últimas décadas, tem sido impulsionado pela posse e compartilhamento de arquivos digitais possuindo conteúdo pornográfico infanto-juvenil.

Silva (2021) [3] investiga os fatores de risco mais comuns para a violência no âmbito doméstico para responder ao questionamento: porque as agressões continuam acontecendo mesmo após a vítima denunciar o agressor. A hipótese é de que as leis a favor da mulher que sofre esse tipo de violência são folhas brandas. O objetivo geral é evidenciar a violência doméstica, mostrando o processo entre a mulher agredida que denuncia e a atuação da lei a favor dela. Segundo o Ministério da Saúde, a violência doméstica aumentou 36% em 2020.

Por fim, Dinair *et al* [15] buscam identificar interseções entre socialização de gênero e violência contra a mulher por parceiro íntimo. Trata-se de pesquisa qualitativa nos moldes da história de vida tópica. Participaram do estudo 16 mulheres que lavraram B.O em uma Delegacia de Defesa da Mulher de um município do interior paulista. Para as entrevistas, foi utilizado um roteiro semiestruturado que abordou as mulheres a partir de suas lembranças da infância, da dinâmica familiar, da união/casamento e da dinâmica familiar pós-união/casamento. O patriarcado enquanto ordem societária orientou a socialização dessas mulheres, que, por causa das relações de dominação do pai (patriarca) e orientadas pelo modelo de submissão das mães, buscaram liberdade em casamento/uniões precoces com companheiros que reproduziram o comportamento paterno. Tais contextos familiares podem tê-las influenciado na escolha de potenciais agressores. Os resultados demonstraram que a socialização de gênero reforça desigualdades entre os sexos e norteia os comportamentos de homens e mulheres, apontando assim para a importância de se problematizar esse tema em diferentes espaços, especialmente nos serviços de atenção primária à saúde, como forma de prevenir relações de violência contra a mulher por parceiro íntimo.

Percebe-se que a Ciência de Dados se tornou uma ferramenta fundamental na identificação de padrões que levaram ao ato de violência, seja a violência contra criança, mulher, etc. Os trabalhos mencionados assim como este artigo vão de encontro na busca por uma resposta para o cenário de extrema violência que vivemos, principalmente no Brasil. Compreender os fatores que motivam os atos de violência é um passo importante para coibir e estabelecer uma solução para essas ocorrências. Os trabalhos relacionados citados são ótimos exemplos de como a pesquisa poderá auxiliar na busca de uma resposta para a solução do problema. Citamos aqui o autor Aníbal Tomás Silva Osses, que na sua pesquisa buscou encontrar Árvores de Decisões e Regras Associativas para compreender a violência no universo infantil, assim como este artigo, que busca compreender essa violência utilizando regras de Associações e os Dados Frequentes, mas no universo da mulher brasileira.

3. Metodologia

Neste trabalho, utilizamos o processo de extração de conhecimento *KDD* para identificar padrões de ocorrências criminais relacionadas à violência contra a mulher brasileira.

O processo de *KDD*, derivado do inglês *Knowledge Discovery in Databases*, consiste em converter um grande volume de dados em informações úteis [27]. Segundo Fayyad [8], este processo é composto por cinco fases: seleção dos dados, pré-processamento, transformação, mineração de dados e interpretação e avaliação de resultados, conforme a descrição a seguir:

1. **Seleção**: nesta fase inicial, obtém-se os dados que serão utilizados para análise e obtenção de conhecimento útil.
2. **Pré-processamento:** realiza-se a limpeza no conjunto de dados, corrigindo dados inconsistentes e ausência de informações.
3. **Transformação**: esta etapa consiste na normalização, criação, agregação ou redução de atributos, de forma a adequar os dados para a próxima fase do processo *KDD*.
4. **Mineração de dados**: o conjunto de dados pré-processados e transformados é então transformado em informação por meio de aplicação de modelos, técnicas ou algoritmos específicos para extração de padrões de dados, utilizando os Conjuntos de Itens Frequentes e Regras de Associação.
5. **Interpretação e avaliação dos resultados**: por fim, obtém-se o conhecimento útil a partir do conjunto de dados, que poderá auxiliar tomadas de decisão e prover melhor entendimento do objeto de estudo. Ainda nesta última etapa, valida-se o resultado obtido, avaliando-se o desempenho do modelo utilizado.

A Figura 1 apresenta o Diagrama *KDD* aplicado neste trabalho para a análise de dados sobre a vulnerabilidade da mulher brasileira para identificação de padrões.

Na etapa da seleção, obtém-se os *datasets* referentes à violência contra a mulher brasileira por meio do Disque 180, e referente especificamente ao estado de Goiás por meio da Secretaria de Segurança Pública do Estado de Goiás (SSP-GO). Em seguida, no Pré-Processamento, realiza-se a limpeza e normatização dos dados utilizando as ferramentas *Python*[11] e *Gnu/Linux*[10]. Nas terceira e quarta etapas ocorrem a Transformação e Mineração de Dados através das ferramentas *Microsoft Power BI*[18] e *Orange Data Mining* [7]. Por fim, na Interpretação e Evolução apresentam-se os resultados, gerando conhecimento que pode ser utilizado como subsídio para a tomada de decisão dos gestores da área de segurança pública. Maiores detalhes dos procedimentos realizados em cada etapa do processo *KDD* são descritos na seção seguinte.

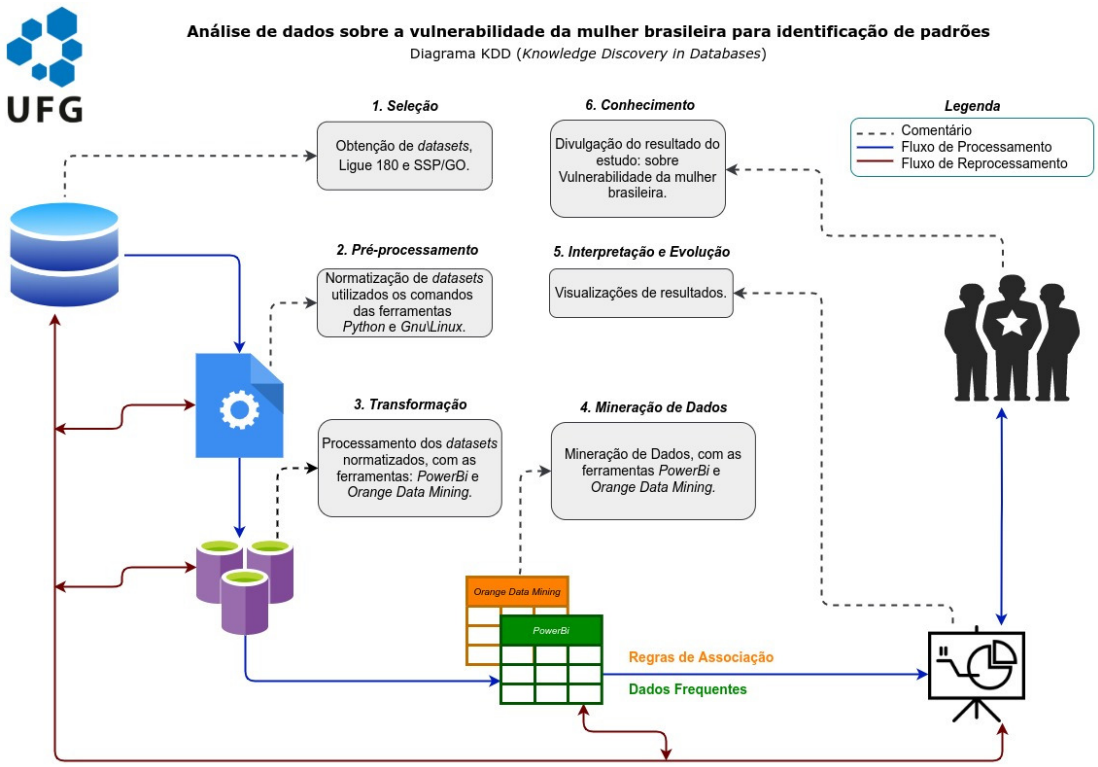


Figura 1. Diagrama *KDD* para a análise de dados sobre a vulnerabilidade da mulher brasileira para identificação de padrões

3.1. Regras de Associação

Nesse trabalho temos como um dos principais objetivos o reconhecimento de padrões junto dos *datasets* do Disque-Denúncia 180 e o *dataset* dos registros de ocorrências criminais contra mulheres do Estado de Goiás, ambos do ano de 2020. Para isso nos apossamos do conhecimento da Mineração de dados, que é conceituado como o conjunto de dados pré-processados e transformados em informação por meio de aplicação de modelos, técnicas ou algoritmos específicos, para extração de padrões de dados, utilizando os Conjuntos de Itens Frequentes e Regras de Associação.

A ferramenta de mineração de dados, *Orange Data Mining*, em sua essência trabalha com o algoritmo *FP-growth*, esse algoritmo utiliza-se da estrutura de árvore para representar o banco de dados, essa representação é chamada de *frequent-pattern tree* ou *FP tree*, como é demonstrado por Abner Suniga (2020) [26].

As regras de associação vieram para auxiliar a descobrir padrões associativos entre conjuntos de itens de dados como citado nos estudos de Agrawal & Srikan (1994) [1]. Nesse contexto de conjunto de dados frequentes temos dois índices que são o ***Suporte e Confiança***, que são medidas para encontrar as associações ou descobrir padrões nos *datasets* submetidos. Tais índices são informados para destacar quais conjuntos de dados são mais significantes. Ainda lembrando que essas associações entre conjuntos se dão sempre por Conjunto(s) Antecedente(s) $\rightarrow$Conjunto(s) Consequente(s), pode ser observado no artigo sobre regras de associação, na documentação original do *Software Orange Data Mining* [2].

Ainda para as regras de associação, se faz necessário ressaltar que fora utilizado a ferramenta *Orange Data Mining* (ferramenta de *Business Inteligence*), onde consegue formular Regras de Associação em um conjunto frequente de dados com base em uma medida de suporte e confiabilidade (geralmente é utilizado nas primeiras execuções do *widget* suporte mínimo de 0,0001% e uma confiança mínima de 1%. Esse Suporte mínimo corresponde a uma proporção mínima de instâncias de dados que devem suportar (conter) o conjunto de itens para que seja gerado. ***Para grandes conjuntos de dados, é normal definir um suporte mínimo inferior (por exemplo, entre 2% e 0,01%)*** [2].

Dessa maneira o ***Índice de Suporte*** é simplesmente o número de transações que incluem todos os itens na parte antecedente e consequente da regra. Assim, para uma determinada regra de associação $\{X\} \rightarrow \{Y\}$, o suporte da regra mede o número total de registros de transação que contém os conjuntos de itens X e Y. O suporte permite-nos aferir quão comum determinado *itemsets* é na base de dados como citado nos estudos de Ulas (1999) [28]. ***Suporte mínimo***: porcentagem de todo o conjunto de dados abrangidos por toda a regra (antecedente e consequente).

Já o ***índice de Confiança***: é a porcentagem de todas as transações satisfazendo X que também satisfazem Y. A confiança também pode ser interpretada como uma estimativa da probabilidade condicional. E assim os índices Suporte e Confiança trazem para as regras de associação: a confiança de uma regra específica para o seu suporte e geralmente é implementado para uma compreensão mais aprofundada da relação entre os itens frequentes [29]. ***O mínimo de confiança***: proporção do número de exemplos que se encaixam o lado direito (consequente) entre aqueles que se encaixam no lado esquerdo (antecedente).

Algumas Regras de Associação foram extraidas no processo de mineração, como por exemplo: Conjunto Antecedente (de mulheres agredidas pela motivação em razão do sexo biológico) →Conjunto Consequente (onde o cenário da violação é a casa onde reside a vítima e o suspeito), nesses padrões associativos foram submetidos ao índice de Suporte 13% e um índice de Confiança 40%. Isso contempla um total de ***21.587 casos denunciados***. Outra regra de associação encontrada foi a do Conjunto Antecedente (mulheres agredida no Estado de São Paulo) →Conjunto Consequente (cenário da violação é casa onde reside a vítima e o suspeito) nesses padrões associativos foram submetidos ao indice de Suporte 9% e um índice de Confiança 20% . Isso contemplando um total de ***13.233 casos denunciados***.

4. Experimentos e Resultados

Nesta seção são descritos os processos utilizados neste trabalho bem como os resultados obtidos. Todos os arquivos referentes aos experimentos: relatório completo, arquivos de instalação, base de dados, *datasets* e Ferramentas de B.I. estão disponibilizados na plataforma GitHub [6].

4.1. Coleta de dados

4.1.1. Denúncias do Disque 180 - Ano referência 2020

A coleta dos dados foi feita através do site governamental: Ministério da Mulher, da Família e dos Direitos Humanos, publicado pelo portal do Governo Federal [9], onde

foram adquiridas duas planilhas -"Primeiro semestre de 2020" e o "Segundo semestre de 2020".

Sendo que na planilha "Primeiro semestre de 2020", tivemos 1.016.398 (***Um Milhão Dezesseis Mil Trezentos Noventa e Oito***) registros e na planilha "Segundo semestre de 2020", tivemos 409.878 (***Quatrocentos Nove Mil Oitocentos Setenta e Oito***) registros. Lembrando que essas planilhas possuem os registros das denúncias registradas através do Disque 180, canal de comunicação oficial do governo federal para denúncias de mulheres em situação de vulnerabilidade e que são acometidas de violência doméstica. Os arquivos totalizam ***1.426.276 (Um Milhão Quatrocentos Vinte Seis Mil Duzentos Setenta e Seis) instâncias***, sendo que a planilha "Primeiro semestre de 2020" com 33 atributos, contendo informações (Ano, Trimestre, Mês, Dia, Tipo_de_violação, Grupo_vulnerável, Espécie_de_violação, País_de_orígem_da_vítima, UF, Canal_de_atendimento, Denúncia_Emergencial, Denunciante, Motivação, Cenário_da_violação, Agravantes, agravantes_policiais, Relação_Vítima_x_Suspeito, Sexo_da_vítima, idade_vitima, Grau_instrução_da_vítima, Raça/cor_da_vítima, Faixa_de_renda_da_vítima, Nacionalidade_da_vítima, deficiencia_vitima, relacao_demandante_vitima, Sexo_do_suspeito, idade_suspeito, Grau_instrução_do_suspeito, Nacionalidade_do_suspeito, Raça/cor_do_suspeito, Faixa_de_renda_do_suspeito, País_de_orígem_do_suspeito, deficiencia_suspeito. Já a planilha "Segundo semestre de 2020" com 28 atributos, contendo informações (Ano, Trimestre, Mês, Dia, País, UF, Grupo_vulnerável, Canal_de_atendimento, Cenário_da_violação, Frequência, Motivação, Denunciante, Sexo_da_vítima, Vítima_LGBT, Raça/Cor_da_vítima, Grau_de_instrução_da_vítima, Faixa_etária_da_vítima, Faixa_de_renda_da_vítima, Profissão_da_vítima, Religião_da_vítima, Deficiência_da_vítima, Deficiência_relacionada_a_doença_rara, Categoria, Categoria_2, Categoria_3, Categoria_4, Categoria_5, Categoria_6).

Como estamos falando de um *big data* do Disque 180 e também como fora apresentado que o mesmo possui em suas planilhas colunas diferentes se faz necessário para Selecão e Processamento dos Dados (Normatização) o uso da linguagem *Python*, para a criação de um *Dataset* único, Normatizado e selecionado para utilização de ferramentas de B.I (*Business Intelligence*) como serão utilizadas nas próximas etapas, ferramentas como *Microsoft Power BI* e *Orange Data Mining*.

4.1.2. Conjunto de dados das Ocorrências criminais de violência contra a mulher no estado de Goiás em 2020

Esta base de dados foi solicitada por meio do Serviço de Informação ao Cidadão (SIC) e foi fornecida pela Gerência do Observatório de Segurança Pública da Secretaria de Segurança Pública do Estado de Goiás (SSP-GO) e é proveniente do Registro de Atendimento Integrado – RAI, composto pela Polícia Militar, Polícia Civil, Corpo de Bombeiros Militar, Superintendência de Polícia Técnico-Científica (SPTC) e Diretoria Geral da Administração Penitenciária (DGAP) [25]. Foram disponibilizados cinco arquivos no formato CSV contendo os registros de ocorrências criminais de violência contra a mulher no estado de Goiás no ano de 2020, sendo cada planilha referente a um tipo de natureza do crime: "AMEACA.csv", "CALUNIA_INJURIA_E_DIFAMACAO.csv",

“ESTUPRO.csv”, “FEMINICIDIO.csv” e “LESAO_CORPORAL.csv”. Os arquivos totalizam 35,25 mil instâncias, com 15 atributos, contendo informações espaço-temporais (estado, Região Integrada de Segurança Pública-RISP, ano, mês dia da semana e faixa de horário) do fato, natureza do crime, e relacionadas ao perfil da vítima (faixa etária, profissão, sexo, cor e grau de instrução), além do número ID, consumação e qualificação. Da mesma forma que a base de dados do Disque 180, também foi utilizada a linguagem *Python* para a criação de um *dataset* único dos dados do estado de Goiás, normatizado e selecionado para análise com o *Microsoft Power BI* e *Orange Data Mining*.

4.2. Pré-processamento e transformação de dados

4.2.1. Denúncias do Disque 180 - Ano referência 2020

Primeiramente, tivemos que unir as duas planilhas: “Primeiro semestre de 2020” e o “Segundo semestre de 2020”, para a criação de um *dataset* único. Isso foi necessário para que os dois semestres do ano de 2020, podessem ser trabalhados de forma única, ou seja, anualmente. Assim foi usado comandos *Shell* em um sistema *Linux* de manipulação de arquivos texto (Comandos: *CAT* , *SORT* e *TR*). Conseguimos unificar as planilhas em um unico Dataset, chamado ”new_dataset_completo_2020.csv”que contem 14 atributos, contendo informações (***Mês, Dia, Grupo_vulnerável, Espécie_de_violação, UF, Denunciante, Motivação, Cenário_da_violação, Sexo_da_vítima, Faixa_etária_da_vítima, Grau_instrução_da_vítima, Raça/cor_da_vítima, Faixa_de_renda_da_vítima, deficiencia_vitima***).

Em seguida, passamos para a parte de Pré-processamento e Transformação de Dados, nesse momento se fez necessário para manipulação de uma quantidade excessiva de dados, ou seja, uma base em *Big Data*. Utilizamos a linguagem *Python*, porém pela facilidade de recursos foi utilizado o serviço *Web* do *Google Colaboratory* [14] que é um ambiente de *notebooks Jupyter* que não requer configuração e é executado na nuvem do Google.

Com a utilização do ***Google Collaboratory*** (*Python*), conseguimos pré-processar e transformar, normatizando os dados e criando um arquivo em *Python* (Normatizar Completo.ipynb), no qual gerou um total de 827 linhas de comando e ao final conseguiu-se um Dataset normatizado, chamado “new_dataset_completo_2020_dia_22_11_2021_Normatizado_Completo.csv”que contém 13 atributos, com dados totalmente selecionados, pré-processados e normatizados, (***Mês, Dia, Grupo_vulnerável, Espécie_de_violação, UF, Denunciante, Motivação, Cenário_da_violação, Sexo_da_vítima, Faixa_etária_da_vítima, Grau_instrução_da_vítima, Raça/cor_da_vítima, Faixa_de_renda_da_vítima***). Apesar de existir apenas uma coluna diferente de um *dataset* para outro, é fato que no *dataset* “new_dataset_completo_2020_dia_22_11_2021_Normatizado_Completo.csv”, foram aplicadas várias técnicas de normatização e que podem ser todas acompanhadas pelo arquivo *Python* (Normatizar Completo.ipynb) disponibilizado no repositório, o qual gerou uma quantidade de ***66.206 (Sessenta Seis Mil Duzentos e Seis***) instâncias. Além que optamos por não utilizar as colunas (atributos) Mês e Dia para a parte de Mineração de Dados, pois ao final das etapas do *KDD*, com a geração de conhecimento formada, a comparação entre os dois *datasets* (Ligue 180 e Ocorrências de Goiás) não seria possível, uma vez que no que se refere a comparação de ambos os *datasets*, tais atributos trariam inconsistência

na análise dos dados.

4.2.2. Conjunto de dados das Ocorrências criminais de violência contra a mulher no estado de Goiás em 2020

Primeiramente, foi identificado que as planilhas "AMEACA.csv", "ESTUPRO.csv", "FEMINICIDIO.csv" e "LESAO_CORPORAL.csv" possuíam a coluna "UF", cujo valor era igual a "GO" para todas as ocorrências, porém a planilha "CALUNIA__INJURIA_E_DIFAMACAO.csv" não possuía tal coluna. Assim, para criação de um *dataset* único e sabendo que todas as planilhas se referem às ocorrências no estado de Goiás, a coluna "UF" foi então excluída nas demais planilhas. Ainda, apenas a planilha "LESAO_CORPORAL.csv" possuía uma coluna com maior detalhamento da natureza do crime, sendo portanto também eliminada para a junção das planilhas. Utilizando o ***Google Collaboratory*** (*Python*) (juntar_go.ipynb), foi então criado o *dataset* único (completo.csv), totalizando 35250 instâncias referentes às ocorrências criminais de violência contra a mulher no estado de Goiás em 2020. Em seguida, realizou-se um tratamento nos dados relacionados ao atributo "faixa etária". Uma vez que o percentual de valores não informados correspondia a 0,91% do total, estes foram considerados na faixa etária com maior percentual, isto é, de 35 a 64 anos, correspondente a 42,83% dos valores (normatizar_go.ipynb).

4.3. Mineração de Dados e Resultados

4.3.1. Denúncias do Disque 180 - Ano referência 2020

Após o *KDD* (*Knowledge Discovery in Databases*) ter completado seu ciclo e conseguirmos chegar ao *dataset* "new_dataset_completo_2020_dia_22_11_2021_Normatizado_Completo.csv", para geração de conhecimento (RESULTADOS), utilizamos duas formas de Análises para os resultados apresentados ao decorrer do artigo: os Itens Frequentes e as Regras de Associação. Isso auxiliará para melhores tomadas de decisão.

Para a demonstração dos conjunto de itens frequentes por ***atributo (colunas)*** das tabelas do nosso *dataset* foi utilizada a ferramenta de *Business Intelligence - Microsoft Power BI* , sendo detectado dentro da ferramenta, através da utilização de painéis, conforme a demonstrado na documentação do site da *Microsoft* [19].

O painel criado para análise deste dataset está disponível para acesso público [5]. Na Figura 2, temos o mapa de formas dos dados por Unidade Federativa. Verificamos que as ocorrências registradas de violência contra à mulher concentram-se na região Sudeste, mais especificamente nos estados do Rio de Janeiro e São Paulo.

Utilizando o *dataset* "new_dataset_completo_2020_dia_22_11_2021_Normatizado_Completo.csv" onde ***foram contabilizados 66206 instâncias (registros) e 9 variáveis (atributos/colunas)***, uma observação se faz necessária quanto a utilização deste *dataset*: a qualidade e quantidade de dados normatizados foram capazes de ***gerarem índices de Suporte e Confiança superiores a 10%***.

Em nossa análise utilizando *Orange Data Mining*, conseguimos Regras de Associação – Com uma parametrização através dos índices de ***Suporte mínimo de 5%***

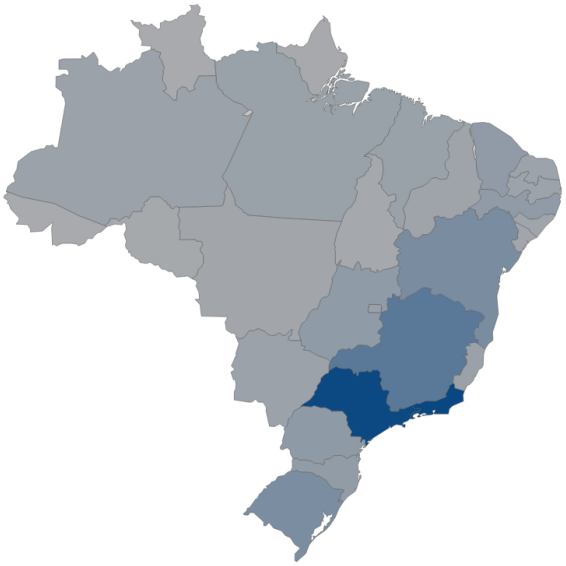


Figura 2. Mapa de formas dos registros do Ligue 180 por Unidade Federativa

e um Mínimo de Confiança 1%. Isso são indicadores de associações fortes, como foi anteriormente citado.

Conjuntos de Itens Frequentes

A) Conjuntos de Itens mais frequentes na variável (Coluna) "Cenário da Violação"

Figura 3. Conjuntos de Itens mais frequentes na variável (Coluna) "Cenário da Violação"

No gráfico em questão fica evidenciado que na variável Cenário da Violação, temos 2 atributos, com fatores de Frequência maiores de 10%. Isso logo nos gera o conhecimento, ou seja, uma informação valiosa para ser disponibilizada aos tomadores de decisão, de que o Cenários onde ocorre a violação contra a mulher de maior frequência seria na ***Residência Comum à Mulher e seu Agressor com 45,34% - Totalizando 30.017 casos denunciados***. Já no segundo atributo com maiores fatores de frequência, temos um cenário de violação onde ***Reside a Vítima com 39,67% - Totalizando 26.263 casos denunciados***.

B) Conjuntos de Itens mais frequentes na variável (Coluna) "Motivação"
No gráfico em questão fica evidenciado que na variável Motivação, temos 4 atributos, com fatores de Frequência maiores de 10%. Isso logo nos gera o

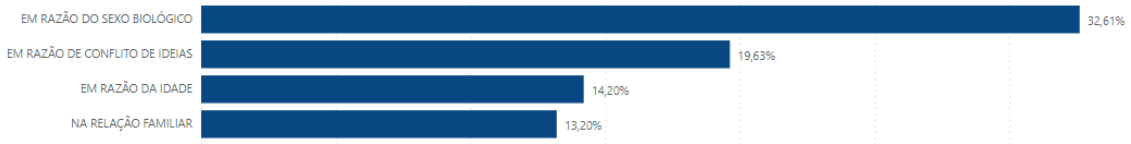


Figura 4. Conjuntos de Itens mais frequentes na variável (Coluna) "Motivação"

conhecimento, ou seja, uma informação valiosa para ser disponibilizada aos tomadores de decisão, de que o fator Motivação que gera violência contra a mulher é oriunda com maior frequência ***Em Razão do Sexo Biológico com 32,61% - Totalizando 21.587 casos dedunciados***, em seguida a Motivação ***Em Razão de Conflito de Ideias com 19,63% - Totalizando 12.993 casos denunciados***, na ordem a Motivação ***Em Razão da Idade com 14,20% - Totalizando 9.401 casos denunciados*** e por ultimo temos a Motivação ***Na Relação Familiar com 13,20% - Totalizando 8.737 casos denunciados***.

C) Conjunto de Itens mais frequentes na variável (Colunas) "UF"

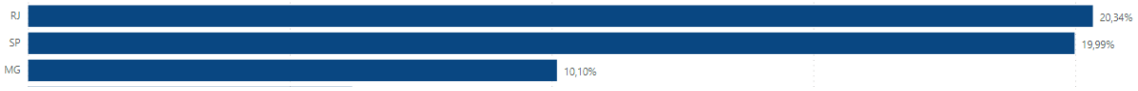


Figura 5. Conjunto de Itens mais frequentes na variável (Colunas) "UF"

No gráfico em questão fica evidenciado que na variável UF (Unidade Federativa) temos 3 atributos, com fatores de Frequência maiores de 10%. Isso logo nos gera o conhecimento, ou seja, uma informação valiosa para ser disponibilizada aos tomadores de decisão, de que um dos Estados da Federação, em que violência contra a mulher é gerada com maior frequência está no ***Estado do Rio de Janeiro com 20,34% - Totalizando 13.464 casos denunciados***, em seguida vem o ***Estado de São Paulo com 19,99% - Totalizando 13.233 casos denunciados*** e por último o ***Estado de Minas Gerais com 10,01% - Totalizando 6.688 casos denunciados***.

D) Conjuntos de Itens mais frequentes na variável (Colunas) "Espécie_de_violação"

Figura 6. Conjuntos de Itens mais frequentes na variável (Colunas) "Espécie_de_violação"

No gráfico em questão fica evidenciado que ***na variável Espécie de violação temos 3 atributos, com fatores de Frequência maiores de 10%***. Isso logo nos gera o conhecimento, ou seja, uma informação valiosa para ser disponibilizada aos tomadores de decisão, de que a espécie de violação em que a mulher é submetida ocorre com maior frequência com ***Ameaça ou Coação com 13,47% - Totalizando 8.919 casos denunciados***, em seguida através do ***Assédio Moral com 13,3% - Totalizando 8.806 casos denunciados*** e por último pelo ***Constrangimento com 11,73% - Totalizando 7.764 casos denunciados***.

Regras de Associação

As Regras de Associação trabalham com dois conjuntos de itens Frequentes, um Antecedente (representado pela condicional SE) e um Consequente (representado pela condicional ENTÃO). Na análise resultante temos as Regras de Associação Máxima e Mínima, ambas com índices de Suporte e Confiança ***(destacados de vermelho no gráfico)***:

A) Regras de Associação - Antecedente **"MOTIVAÇÃO"** com 10 Regras de associação e 38 padrões

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.132	0.405	0.326	1.391	0.894	-0.016	Motivação=EM RAZÃO DO SEXO BIOLÓGICO	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.129	0.395	0.326	1.217	0.996	-0.001	Motivação=EM RAZÃO DO SEXO BIOLÓGICO	→	Cenário_da_violação=CASA DA VÍTIMA
0.093	0.473	0.196	2.021	1.192	0.015	Motivação=EM RAZÃO DE CONFLITO DE IDEIAS	→	Cenário_da_violação=CASA DA VÍTIMA
0.089	0.675	0.132	3.436	1.489	0.029	Motivação=NA RELAÇÃO FAMILIAR	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.077	0.539	0.142	3.193	1.188	0.012	Motivação=EM RAZÃO DA IDADE	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.070	0.214	0.326	0.624	1.050	0.003	Motivação=EM RAZÃO DO SEXO BIOLÓGICO	→	UF=RJ
0.065	0.332	0.196	2.310	0.733	-0.024	Motivação=EM RAZÃO DE CONFLITO DE IDEIAS	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.062	0.190	0.326	0.613	0.949	-0.003	Motivação=EM RAZÃO DO SEXO BIOLÓGICO	→	UF=SP
0.054	0.381	0.142	2.794	0.961	-0.002	Motivação=EM RAZÃO DA IDADE	→	Cenário_da_violação=CASA DA VÍTIMA
0.053	0.402	0.132	1.021	2.986	0.035	Motivação=NA RELAÇÃO FAMILIAR	→	Espécie_de_violação=AMEAÇA ou COAÇÃO

Figura 7. Regras de Associação - Antecedente "MOTIVAÇÃO"

Na regra máxima de associação temos a seguinte situação: ***SE*** uma mulher agredida pela motivação em razão do sexo biologico ***ENTÃO*** o cenário da violação é casa onde reside a vítima e o suspeito com supp-13% e conf-40% . Isso contempla um total de ***21.587 casos denunciados***.

Na regra mínima de associação temos a seguinte situação: ***SE*** uma mulher agredida pela motivação da relação familiar, ***ENTÃO*** o cenário da violação é do tipo ameaça ou coação com índices supp-5% e conf-40% . Isso contempla um total de ***8.737 casos denunciados***.

B) Regras de Associação - Antecedente **"CENÁRIO DA VIOLAÇÃO"** com 16 Regras de associação e 38 padrões

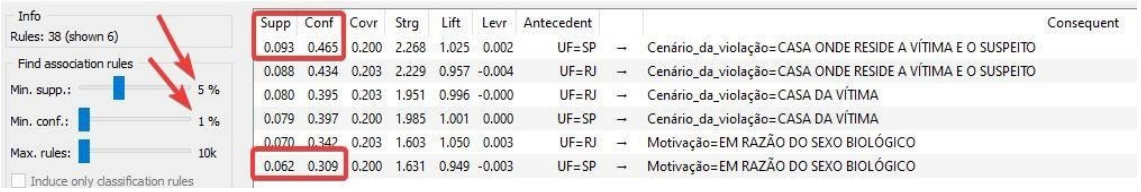
Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.132	0.292	0.453	0.719	0.894	-0.016	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	Motivação=EM RAZÃO DO SEXO BIOLÓGICO
0.129	0.325	0.397	0.822	0.996	-0.001	Cenário_da_violação=CASA DA VÍTIMA	→	Motivação=EM RAZÃO DO SEXO BIOLÓGICO
0.093	0.205	0.453	0.441	1.025	0.002	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	UF=SP
0.093	0.234	0.397	0.495	1.192	0.015	Cenário_da_violação=CASA DA VÍTIMA	→	Motivação=EM RAZÃO DE CONFLITO DE IDEIAS
0.089	0.196	0.453	0.291	1.489	0.029	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	Motivação=NA RELAÇÃO FAMILIAR
0.088	0.195	0.453	0.449	0.957	-0.004	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	UF=RJ
0.080	0.203	0.397	0.513	0.996	-0.000	Cenário_da_violação=CASA DA VÍTIMA	→	UF=RJ
0.079	0.200	0.397	0.504	1.001	0.000	Cenário_da_violação=CASA DA VÍTIMA	→	UF=SP
0.077	0.169	0.453	0.313	1.188	0.012	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	Motivação=EM RAZÃO DA IDADE
0.065	0.144	0.453	0.433	0.733	-0.024	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	Motivação=EM RAZÃO DE CONFLITO DE IDEIAS
0.064	0.141	0.453	0.297	1.043	0.003	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	Espécie_de_violação=AMEAÇA ou COAÇÃO
0.063	0.138	0.453	0.293	1.039	0.002	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	Espécie_de_violação=ASSÉDIO MORAL
0.055	0.138	0.397	0.340	1.026	0.001	Cenário_da_violação=CASA DA VÍTIMA	→	Espécie_de_violação=AMEAÇA ou COAÇÃO
0.054	0.136	0.397	0.358	0.961	-0.002	Cenário_da_violação=CASA DA VÍTIMA	→	Motivação=EM RAZÃO DA IDADE
0.053	0.132	0.397	0.335	0.996	-0.000	Cenário_da_violação=CASA DA VÍTIMA	→	Espécie_de_violação=ASSÉDIO MORAL
0.051	0.113	0.453	0.259	0.959	-0.002	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	→	Espécie_de_violação=CONSTRANGIMENTO

Figura 8. Regras de Associação - Antecedente "CENÁRIO DA VIOLAÇÃO"

Na regra máxima de associação temos a seguinte situação: ***SE*** uma mulher agredida no cenário da violação onde é a casa onde reside a vítima e o suspeito ***ENTÃO*** a motivação é em razão do sexo biológico com índices supp-13% e conf-29% . Isso contemplando um total de ***30.017 casos denunciados***.

Na regra mínima de associação temos a seguinte situação: ***SE*** uma mulher agredida no cenário da violação onde é a casa onde reside a vítima e o suspeito, ***ENTÃO*** a espécie de violação será por constrangimento com índices supp-5% e conf-11%. Isso contemplando um total de ***30.017 casos denunciados***.

C) Regras de Associação - Antecedente **"UF"** com 6 Regras de associação e 38 padrões.



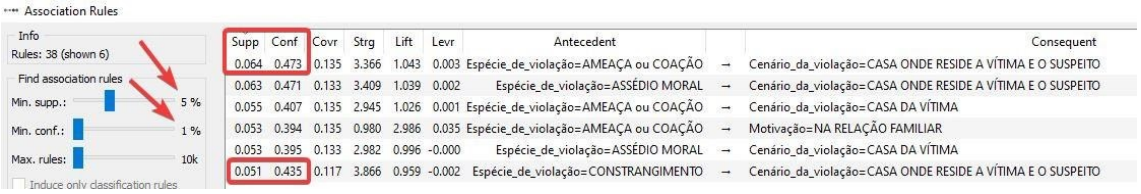
Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.093	0.465	0.200	2.268	1.025	0.002	UF=SP	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.088	0.434	0.203	2.229	0.957	-0.004	UF=RJ	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.080	0.395	0.203	1.951	0.996	-0.000	UF=RJ	→	Cenário_da_violação=CASA DA VÍTIMA
0.079	0.397	0.200	1.985	1.001	0.000	UF=SP	→	Cenário_da_violação=CASA DA VÍTIMA
0.070	0.342	0.203	1.603	1.050	0.003	UF=RJ	→	Motivação=EM RAZÃO DO SEXO BIOLÓGICO
0.062	0.309	0.200	1.631	0.949	-0.003	UF=SP	→	Motivação=EM RAZÃO DO SEXO BIOLÓGICO

Figura 9. Regras de Associação - Antecedente "UF"

Na regra máxima de associação temos a seguinte situação: ***SE*** uma mulher agredida no estado de são paulo ***ENTÃO*** o cenário da violação é casa onde reside a vítima e o suspeito com índices supp-9% e conf-20% . Isso contemplando um total de ***13.233 casos denunciados***.

Na regra mínima de associação temos a seguinte situação: ***SE*** uma mulher agredida no estado de são paulo, ***ENTÃO*** o cenário da violação é em razão do sexo biológico com índices supp-6% e conf-20% . Isso contemplando um total de 13.233 casos denunciados.

D) Regras de Associação - Antecedente **"ESPÉCIE DE VIOLAÇÃO"** Com 6 Regras de associação e 38 padrões.



Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.064	0.473	0.135	3.366	1.043	0.003	Espécie_de_violação=AMEAÇA ou COAÇÃO	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.063	0.471	0.133	3.409	1.039	0.002	Espécie_de_violação=ASSÉDIO MORAL	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO
0.055	0.407	0.135	2.945	1.026	0.001	Espécie_de_violação=AMEAÇA ou COAÇÃO	→	Cenário_da_violação=CASA DA VÍTIMA
0.053	0.394	0.135	0.980	2.986	0.035	Espécie_de_violação=AMEAÇA ou COAÇÃO	→	Motivação=NA RELAÇÃO FAMILIAR
0.053	0.395	0.133	2.982	0.996	-0.000	Espécie_de_violação=ASSÉDIO MORAL	→	Cenário_da_violação=CASA DA VÍTIMA
0.051	0.435	0.117	3.866	0.959	-0.002	Espécie_de_violação=CONSTRANGIMENTO	→	Cenário_da_violação=CASA ONDE RESIDE A VÍTIMA E O SUSPEITO

Figura 10. Regras de Associação - Antecedente "ESPÉCIE DE VIOLAÇÃO"

Na regra máxima de associação temos a seguinte situação: ***SE*** uma mulher agredida por uma especie de violação de ameaça ou coação ***ENTÃO*** o cenário da violação é casa onde reside a vítima e o suspeito com índices supp-6% e conf-47% . Isso contemplando um total de ***8.919 casos denunciados***.

Na regra mínima de associação temos a seguinte situação: ***SE*** uma mulher agredida por uma espécie de violação com constrangimento, ***ENTÃO*** o cenário da violação é na casa onde reside a vítima e o suspeito com índices supp-5% e conf-43% . Isso contemplando um total de ***7.764 casos denunciados***.

4.3.2. Conjunto de dados das Ocorrências criminais de violência contra a mulher no estado de Goiás em 2020

Em relação aos dados do estado de Goiás, por meio da visualização dos dados pelo *Microsoft Power BI* [4], verificou-se que três atributos possuíam mais de 70% de seus valores como "não informado", conforme apresentado nas Figuras 11 a 13.

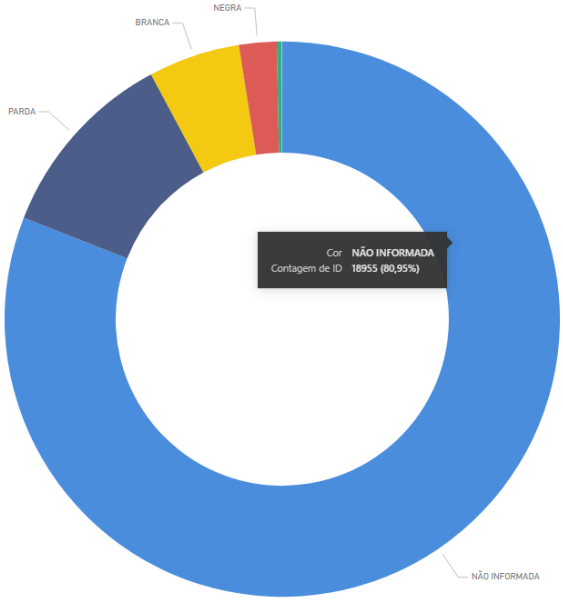


Figura 11. Gráfico dos valores do atributo Cor das vítimas do *dataset* da SSP/GO

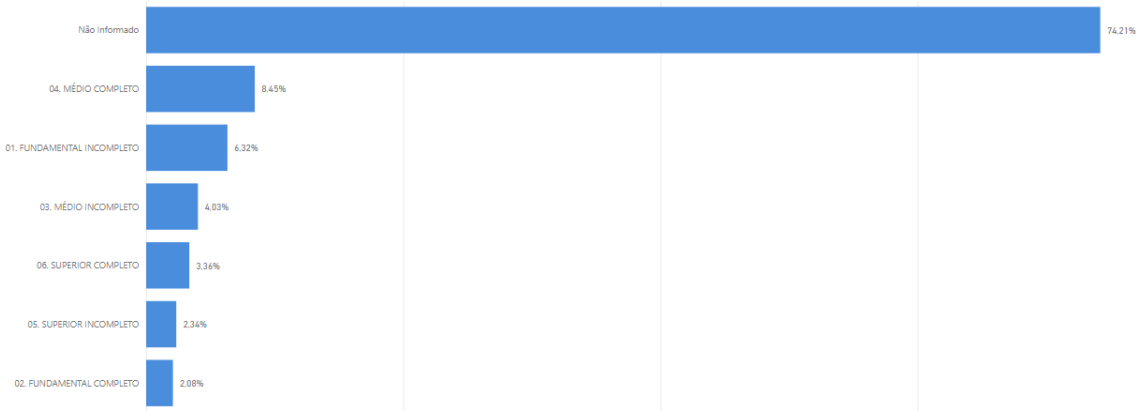


Figura 12. Gráfico dos valores do atributo Grau de Instrução das vítimas do dataset da SSP/GO

Assim, os atributos "Cor", "Grau de Instrução" e "Profissão" não foram considerados na obtenção de regras de associação, tendo em vista a grande ausência de dados. Entretanto, pode-se observar que, dentre os valores informados, o perfil das vítimas no estado de Goiás é em sua maioria mulheres pardas, com ensino médio completo, cuidadoras do lar.

Também não foram considerados na obtenção de regras de associação os atributos que possuíam um único valor para todas as instâncias. Por exemplo, na coluna "Ano"

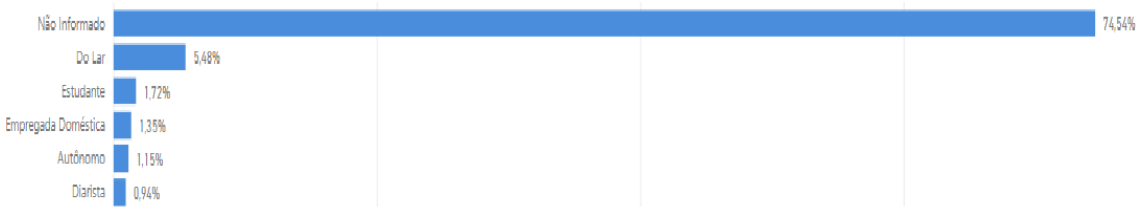


Figura 13. Gráfico dos valores do atributo Profissão das vítimas do dataset da SSP/GO

todas as ocorrências possuíam o valor igual a 2020. Analogamente, possuíam um único valor as colunas: "Consumação" (valor: Consumado), "Qualificação" (valor: Vítima) e "Sexo" (valor: Feminino).

Regras de Associação

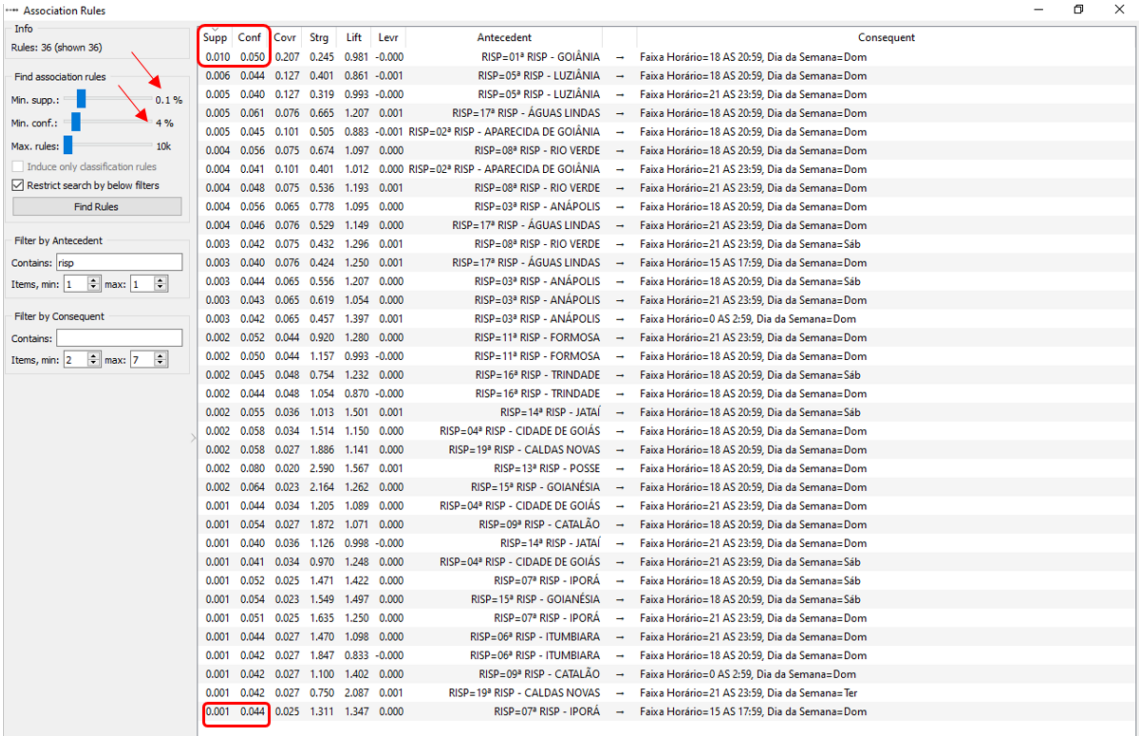
Tendo em vista que o *dataset* das ocorrências no estado de Goiás é menor do que o obtido pelo Disque 180, foi necessário considerar valores menores para os índices mínimos de suporte e confiança a fim de obtermos as Regras de Associação do conjunto de dados. Desta forma, considerou-se em nossa análise utilizando *Orange Data Mining*, os índices de suporte mínimo de 0,1% e um mínimo de confiança de 4%.

A) Regras de Associação - Antecedente **"RISP"** com 36 regras de associação

No estudo das ocorrências criminais contra a mulher no estado de Goiás, buscamos verificar a existência de padrões de associação espaço-temporais. Assim, obtivemos as regras de associação cujo atributo antecedente era a Região Integrada de Segurança Pública (RISP) e os atributos consequentes eram a Faixa de Horário e o Dia da Semana da ocorrência do crime.

Na regra máxima de associação temos a seguinte situação: ***SE*** o local da ocorrência corresponde à 01ª RISP (Goiânia) ***ENTÃO*** a Faixa de Horário do crime é de 18 às 20:59 e o dia da Semana é domingo com índices supp-1% e conf-5%.

Na regra mínima de associação temos a seguinte situação: ***SE*** o local da ocorrência corresponde à 07ª RISP (Iporá) ***ENTÃO*** a Faixa de Horário do crime é de 15 às 17:59 e o dia da Semana é domingo com índices supp-0,1% e conf-4,4%.



Association Rules

Info
Rules: 36 (shown 36)

Find association rules
Min. supp.: 0.1 %
Min. conf.: 4 %
Max. rules: 10k
Induce only classification rules
Restrict search by below filters
Find Rules

Filter by Antecedent
Contains: risp
Items, min: 1 max: 1

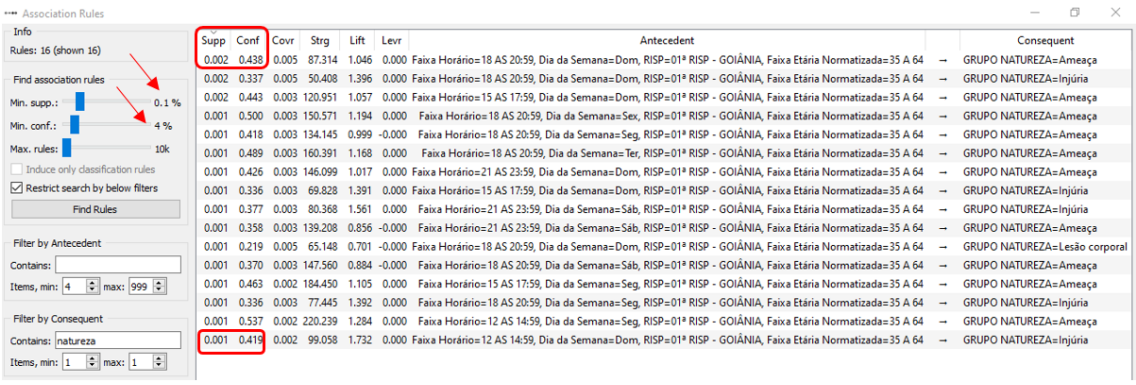
Filter by Consequent
Contains:
Items, min: 2 max: 7

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.010	0.050	0.207	0.245	0.981	-0.000	RISP=01ª RISP - GOIÂNIA	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.006	0.044	0.127	0.401	0.861	-0.001	RISP=05ª RISP - LUZIÂNIA	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.005	0.040	0.127	0.319	0.993	-0.000	RISP=05ª RISP - LUZIÂNIA	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.005	0.061	0.076	0.665	1.207	0.001	RISP=17ª RISP - ÁGUAS LINDAS	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.005	0.045	0.101	0.505	0.883	-0.001	RISP=02ª RISP - APARECIDA DE GOIÂNIA	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.004	0.056	0.075	0.674	1.097	0.000	RISP=08ª RISP - RIO VERDE	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.004	0.041	0.101	0.401	1.012	0.000	RISP=02ª RISP - APARECIDA DE GOIÂNIA	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.004	0.048	0.075	0.536	1.193	0.001	RISP=08ª RISP - RIO VERDE	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.004	0.056	0.065	0.778	1.095	0.000	RISP=03ª RISP - ANÁPOLIS	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.004	0.046	0.076	0.529	1.149	0.000	RISP=17ª RISP - ÁGUAS LINDAS	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.003	0.042	0.075	0.432	1.296	0.001	RISP=08ª RISP - RIO VERDE	→	Faixa Horário=21 AS 23:59, Dia da Semana=Sáb
0.003	0.040	0.076	0.424	1.250	0.001	RISP=17ª RISP - ÁGUAS LINDAS	→	Faixa Horário=15 AS 17:59, Dia da Semana=Dom
0.003	0.044	0.065	0.556	1.207	0.000	RISP=03ª RISP - ANÁPOLIS	→	Faixa Horário=18 AS 20:59, Dia da Semana=Sáb
0.003	0.043	0.065	0.619	1.054	0.000	RISP=03ª RISP - ANÁPOLIS	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.003	0.042	0.065	0.457	1.397	0.001	RISP=03ª RISP - ANÁPOLIS	→	Faixa Horário=0 AS 2:59, Dia da Semana=Dom
0.002	0.052	0.044	0.920	1.280	0.000	RISP=11ª RISP - FORMOSA	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.002	0.050	0.044	1.157	0.993	-0.000	RISP=11ª RISP - FORMOSA	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.002	0.045	0.048	0.754	1.232	0.000	RISP=16ª RISP - TRINDADE	→	Faixa Horário=18 AS 20:59, Dia da Semana=Sáb
0.002	0.044	0.048	1.054	0.870	-0.000	RISP=16ª RISP - TRINDADE	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.002	0.055	0.036	1.013	1.501	0.001	RISP=14ª RISP - JATAÍ	→	Faixa Horário=18 AS 20:59, Dia da Semana=Sáb
0.002	0.058	0.034	1.514	1.150	0.000	RISP=04ª RISP - CIDADE DE GOIÁS	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.002	0.058	0.027	1.886	1.141	0.000	RISP=19ª RISP - CALDAS NOVAS	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.002	0.080	0.020	2.590	1.567	0.001	RISP=13ª RISP - POSSE	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.002	0.064	0.023	2.164	1.262	0.000	RISP=15ª RISP - GOIANÉSIA	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.001	0.044	0.034	1.205	1.089	0.000	RISP=04ª RISP - CIDADE DE GOIÁS	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.001	0.054	0.027	1.872	1.071	0.000	RISP=09ª RISP - CATALÃO	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.001	0.040	0.036	1.126	0.998	-0.000	RISP=14ª RISP - JATAÍ	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.001	0.041	0.034	0.970	1.248	0.000	RISP=04ª RISP - CIDADE DE GOIÁS	→	Faixa Horário=21 AS 23:59, Dia da Semana=Sáb
0.001	0.052	0.025	1.471	1.422	0.000	RISP=07ª RISP - IPORÁ	→	Faixa Horário=18 AS 20:59, Dia da Semana=Sáb
0.001	0.054	0.023	1.549	1.497	0.000	RISP=15ª RISP - GOIANÉSIA	→	Faixa Horário=18 AS 20:59, Dia da Semana=Sáb
0.001	0.051	0.025	1.635	1.250	0.000	RISP=07ª RISP - IPORÁ	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.001	0.044	0.027	1.470	1.098	0.000	RISP=06ª RISP - ITUMBIARA	→	Faixa Horário=21 AS 23:59, Dia da Semana=Dom
0.001	0.042	0.027	1.847	0.833	-0.000	RISP=06ª RISP - ITUMBIARA	→	Faixa Horário=18 AS 20:59, Dia da Semana=Dom
0.001	0.042	0.027	1.100	1.402	0.000	RISP=09ª RISP - CATALÃO	→	Faixa Horário=0 AS 2:59, Dia da Semana=Dom
0.001	0.042	0.027	0.750	2.087	0.001	RISP=19ª RISP - CALDAS NOVAS	→	Faixa Horário=21 AS 23:59, Dia da Semana=Ter
0.001	0.044	0.025	1.311	1.347	0.000	RISP=07ª RISP - IPORÁ	→	Faixa Horário=15 AS 17:59, Dia da Semana=Dom

Figura 14. Regras de Associação - Antecedente "RISP"

B) Regras de Associação - Consequente **"GRUPO NATUREZA"** com 16 regras de associação

Em outra análise no conjunto de dados de ocorrências criminais contra a mulher no estado de Goiás, buscamos verificar se há relação entre as características espaço-temporais da ocorrência e a idade da vítima com o tipo de natureza do crime. Assim, obtivemos as regras de associação com 4 atributos antecedentes e o atributo "GRUPO NATUREZA" como consequente.



Association Rules

Info
Rules: 16 (shown 16)

Find association rules
Min. supp.: 0.1 %
Min. conf.: 4 %
Max. rules: 10k
Induce only classification rules
Restrict search by below filters
Find Rules

Filter by Antecedent
Contains:
Items, min: 4 max: 999

Filter by Consequent
Contains: natureza
Items, min: 1 max: 1

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.002	0.438	0.005	87.314	1.046	0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Dom, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.002	0.337	0.005	50.408	1.396	0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Dom, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Injúria
0.002	0.443	0.003	120.951	1.057	0.000	Faixa Horário=15 AS 17:59, Dia da Semana=Dom, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.500	0.003	150.571	1.194	0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Sex, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.418	0.003	134.145	0.999	-0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Seg, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.489	0.003	160.391	1.168	0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Ter, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.426	0.003	146.099	1.017	0.000	Faixa Horário=21 AS 23:59, Dia da Semana=Dom, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.336	0.003	69.828	1.391	0.000	Faixa Horário=15 AS 17:59, Dia da Semana=Dom, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Injúria
0.001	0.377	0.003	80.368	1.561	0.000	Faixa Horário=21 AS 23:59, Dia da Semana=Sáb, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Injúria
0.001	0.358	0.003	139.208	0.856	-0.000	Faixa Horário=21 AS 23:59, Dia da Semana=Sáb, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.219	0.005	65.148	0.701	-0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Dom, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Lesão corporal
0.001	0.370	0.003	147.560	0.884	-0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Sáb, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.463	0.002	184.450	1.105	0.000	Faixa Horário=15 AS 17:59, Dia da Semana=Seg, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.336	0.003	77.445	1.392	0.000	Faixa Horário=18 AS 20:59, Dia da Semana=Seg, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Injúria
0.001	0.537	0.002	220.239	1.284	0.000	Faixa Horário=12 AS 14:59, Dia da Semana=Seg, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Ameaça
0.001	0.419	0.002	99.058	1.732	0.000	Faixa Horário=12 AS 14:59, Dia da Semana=Dom, RISP=01ª RISP - GOIÂNIA, Faixa Etária Normatizada=35 A 64	→	GRUPO NATUREZA=Injúria

Figura 15. Regras de Associação - Consequente "GRUPO NATUREZA"

Na regra máxima de associação temos a seguinte situação: ***SE*** a Faixa de Horário da ocorrência do crime é de 18 às 20:59, o dia da Semana é Domingo, o local é a 01ª RISP (Goiânia) e a Faixa Etária da vítima é entre 35 a 64 anos ***ENTÃO*** a natureza do crime é ameaça com índices supp-0,2% e conf-43,8%.

Na regra mínima de associação temos a seguinte situação: ***SE*** a Faixa de Horário da ocorrência do crime é de 12 às 14:59, o dia da Semana é Domingo, o local é a 01ª RISP (Goiânia) e a Faixa Etária da vítima é entre 35 a 64 anos ***ENTÃO*** a natureza do crime é injúria com índices supp-0,1% e conf-41,9%.

5. Conclusões e Trabalhos Futuros

O presente trabalho teve como objetivo uma análise sobre a vulnerabilidade da mulher brasileira utilizando o método *KDD* em duas bases de dados: uma referente aos registros de denúncias recebidas no Ligue 180; e a outra fonte é referente aos Boletins de Ocorrências Policiais do estado de Goiás. Os *datasets* têm o foco na violência contra a mulher, e os dados obtidos são referentes aos registros do ano de 2020. Para o processamento dos *datasets*, foram utilizadas as ferramentas *Orange Data Mining* e *Microsoft Power BI*. A ferramenta *Orange Data Mining* foi utilizada na mineração de dados e na identificação das regras de Associação. A ferramenta *Microsoft Power BI* foi utilizada na obtenção dos dados Frequentes. Com auxílio das ferramentas foram obtidos os padrões sobre à violência contra as mulheres, que infelizmente continua fazendo novas vítimas diariamente e contribuindo para os índices em uma escalada vertiginosa. Uma dessas regras de associação tem haver com cenário de violação que na maioria das situações haverá a probabilidade de ocorrem agressões às mulheres que residem junto com os seus agressores, motivado pela simples fato de serem do sexo feminino. Outra regra de associação, é que essa mulher poderá com frequência enorme ser do Estado do Rio de Janeiro, e para fechar esse primeiro ciclo de associações, essa mulher ainda terá como maior fator de agressão as ameaças e coações à sua integridade física e mental. Na obtenção das regras de associação dos dados do estado de Goiás, foi possível observar que as mulheres goianas com idade entre 35 (Trinta e cinco) a 64 (Sessenta e quatro) anos, no intervalo de horário das 18h às 20h:59, no dia de Domingo, na área de atuação da RISP Goiânia, estarão mais vulneráveis ao crime de ameaça física.

A violência no Brasil é uma dura realidade para uma grande parcela da população, os fatos apresentados neste artigo vão de encontro à essa triste realidade, as mulheres estão sujeitas aos mais diversos tipos de violências existentes por essas terras. Infelizmente, a violência contra elas se tornou um problema crônico no Brasil, sem a perspectiva de solução a médio ou longo prazo. O caminho para a solução definitiva do problema passará por investimentos governamentais e da iniciativa privada em setores chaves, principalmente o setor de educação. Melhorar o nível de escolaridade dos brasileiros não resolverá o problema em definitivo, a educação possibilitará à vítima o conhecimento dos seus direitos, e a consequente busca por esses direitos com a possível condenação do acusado, criará um senso de que a justiça é de fato cumprida, inibindo novos agressores ou agressões recorrentes. Ainda no item investimentos, as cidades brasileiras em sua grande maioria não possuem uma infraestrutura mínima para registrar ou prestar um serviço de acolhidas para essas vítimas, esse é um setor carente de investimentos públicos, ficando esse tipo assistência relegada ao terceiro setor. ONG (Organização não Governamental) como CEVAM (Centro de Valorização da Mulher Consuelo Nasser) existente na cidade de Goiânia/GO, busca suprir essa carência. O CEVAM, em diversos momentos da sua história de mais de 40 (Quarenta) anos teve o seu funcionamento comprometido por falta de investimentos mínimos. Os investimentos são responsabilidade dos entes públicos, da iniciativa privada e de nós cidadãos, membros da sociedade civil. Quando a sociedade

organizada falha em criar uma infraestrutura básica de atendimento e de proteção das vítimas, estamos colocando ou empurrando essas pessoas para serem expostas aos atos de violência continua, sendo que a violência doméstica é a líder absoluta dos registros em ambos *datasets*.

Por fim, como trabalhos futuros sugerimos o estudo sobre violência contra a mulher com a utilização *dataset* disponibilizado pelo SUS (Sistema Único de Saúde), utilizando ferramentas como *Weka* e Linguagem R a fim de se obter uma comparação com a análise realizada neste artigo, ou seja, uma comparação dos padrões criminais denunciados no Ligue 180 e SSP-GO com os padrões criminais que resultaram em dano físico na vítima registrados no SUS.

Referências

[1] R. Agrawal and R. Srikan. Fast algorithms for mining association rules. Proc. 20th Int. Conf. Very Large Data Bases,VLDB, 1994. 1215, 487–499.

[2] AJDA. Association rules in orange. Disponível em: https://bit.ly/3u7AyOO, 2016. Accessado em: 09 de fev. 2022.

[3] Luiz Felipe Vieira da Silva. ViolÊncia contra a mulher: Fatores de risco no ambiente domÉstico. Disponível em: http://periodicos.processus.com.br/index.php/multi/article/view/431/508, 2021. Acessado em: 12 de fev. 2022.

[4] Dirlei Ribeiro de Mendonça Junior, Flavia Megumi Ohara, and Huziel Correa Soares. Análise do dataset ssp/go - ano referência 2020. Disponível em: https://bit.ly/BI_sspgo, 2021.

[5] Dirlei Ribeiro de Mendonça Junior, Flavia Megumi Ohara, and Huziel Correa Soares. Análise do dataset denúncias do disque 180 - ano referência 2020. Disponível em:https://bit.ly/BI_180, 2021.

[6] Dirlei Ribeiro de Mendonça Junior, Flavia Megumi Ohara, and Huziel Correa Soares. Análise de dados sobre a vulnerabilidade da mulher brasileira para identificação de padrões. Disponível em: https://github.com/huzielcorrea/UFG_TCC_DATA_MINING, 2022. Accessado em: 29 de jan. 2022.

[7] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan. Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, 14:2349–2353, 2013. Disponível em: http://jmlr.org/papers/v14/demsar13a.html.

[8] U. M. Fayyad and et al. Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, 1996.

[9] Governo Federal. Relatórios com os dados de 2019 e 2020 sobre violência contra a mulher e violência doméstica e familiar contra a mulher registrados pelo ligue 180. Disponível em: https://bit.ly/3gHv8SC, 2021. Acessado em: 11 de fev. de 2022.

[10] Free Software Foundation. Gnu/linux. Disponível em: https://www.gnu.org/software/software.html. Acesso em: 9 de nov. de 2021.

[11] PYTHON SOFTWARE FOUNDATION. The python language reference. Disponível em: https://www.python.org/. Acesso em: 22 de nov. de 2021.

[12] F. Freire, F. Grandin, and A. Martins. Justiça sem medo: 10 anos depois da morte da juíza patrícia acioli, polícia mata 9 vezes mais em são gonçalo. Disponível em: https://g1.globo.com/rj/rio-de-janeiro/noticia/2021/08/11/justica-sem-medo-10-anos-depois-da-morte-da-juiza-patricia-acioli-policia-mata-9-vezes-mais-em-sao-goncalo.ghtml, 2021. Acesso em: 13 de jan. de 2022.

[13] G1. Roger abdelmassih é condenado a mais de 200 anos de prisão. Disponível em: http://g1.globo.com/sao-paulo/noticia/2010/11/roger-abdelmassih-e-condenado-mais-de-200-anos-de-prisao.html, 2010. Acesso em: 16 de out. de 2021.

[14] Google. Google colaboratory (colab). Disponível em: https://colab.research.google.com/. Acesso em: 22 de nov. de 2021.

[15] Dinair Ferreira Machado, Elen Rose Lodeiro Castanheira, and Margareth Aparecida Santini de Almeida. Interseções entre socialização de gênero e violência contra a mulher por parceiro íntimo. Disponível em: https://www.scielo.br/j/csc/a/yxT3jYZczC9LvqH8WRj79rf/, 2021. Acessado em: 12 de fev. 2022.

[16] G. Maria and P. Bittar. Ligue 180 é o mais importante projeto de enfrentamento à violência contra a mulher. Disponível em: https://www.camara.leg.br/noticias/543357-ligue-180-e-o-mais-importante-projeto-de-enfrentamento-a-violencia-contra-a-mulher-diz-secretaria/, 2018. Acesso em: 13 de jan. de 2022.

[17] Rafaela Martins. Ginecologista é preso em anápolis por violação sexual; uma vítima é do df. Disponível em: https://www.correiobraziliense.com.br/cidades-df/2021/09/4952466-ginecologista-e-preso-em-goiania-por-violacao-sexual-uma-vitima-e-do-df.html, 2021. Acesso em: 13 de jan. de 2022.

[18] Microsoft. Power bi. Disponível em: https://powerbi.microsoft.com/pt-br/. Acesso em: 22 de nov. de 2021.

[19] Microsoft. Usar o painel análise do power bi desktop. Disponível em: https://bit.ly/3KT2ZGm, 2021. Acessado em: 11 de fev. 2022.

[20] Danilo Coura MOREIRA. Detecção de pornografia infanto-juvenil baseada em pornografia adulta e estimativa de idade facial. Disponível em: http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/17946, 2021. Acessado em: 12 de fev. 2022.

[21] Silva Osses and Aníbal Tomás. Análise da predição da violência infantil por meio de Árvores de decisão e regras de associação. Disponível em: https://repositorio.ufscar.br/handle/ufscar/12915, 2020. Acessado em: 12 de fev. 2022.

[22] D. F. Peixoto, M. S. Silva, and P. Ag. T. do Amaral. Mulheres em situação de violência na cidade do rio de janeiro: identificação e análise dos padrões espaciais. 2016. Disponível em: http://www.abep.org.br/xxencontro/files/_paper/1066-787.pdf, 2016. Anais da ABEP. Acesso em: 13 de jan. de 2022.

[23] Klark Gable Souza Porto. Descoberta de conhecimento através da análise e mineração em dados do enem. 2020. Disponível em: https://bdm.unb.br/handle/10483/28902, 2019. Universidade de Brasília. Acesso em: 13 de jan. de 2022.

[24] Senado. Painel de violência contra mulheres. Disponível em: http://www9.senado.gov.br/painelstrans, 2019. Acesso em: 16 de out. de 2021.

[25] SSPGO. Tutorial painel de ocorrências ssp. Disponível em: https://www.seguranca.go.gov.br/apps/observatorio/manuais/tutorial-painel-de-ocorrencias.pdf, 2019. Acesso em: 15 de dez. de 2021.
[26] Abner Suniga. Minerações de dados frequentes com apriori e fp growth.
[27] P.N. Tan, M. Steimbach, and V. Kumar. *Introduction to Data Mining*. Boston: Addison Wesley, 2006. 769p.
[28] M. A. Ulas. Market basket analysis for data mining. Master's thesis, Bogazici University, 1999.
[29] Wikipedia. Association rule learning. Disponível em: https://bit.ly/3o7DiI7, 2021. Acessado em: 11 de fev. de 2022.