

# Aplicação de Regras de Associação em Dados da Criminalidade da Cidade do Recife

*Application of association rules between criminality data from the city of Recife*

**Bettina Cavalcanti Araújo**<sup>1</sup>  [orcid.org/0000-0002-9821-1812](https://orcid.org/0000-0002-9821-1812)

**Alexandre Magno Andrade Maciel**<sup>1</sup>  [orcid.org/0000-0003-4348-9291](https://orcid.org/0000-0003-4348-9291)

<sup>1</sup> Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

E-mail do autor principal: [bca@ecomp.poli.br](mailto:bca@ecomp.poli.br)

## Resumo

---

Os crescentes números da criminalidade na cidade do Recife fez com que o Governo do Estado aumentasse a frota de policiais nas ruas no primeiro trimestre de 2018 esperando uma queda nos crimes. Porém, combater a violência apenas de forma reativa é uma estratégia insuficiente para uma redução significativa, por conta da complexidade do perfil criminal de cada localidade, dificultando a organização tática dos policiais, sendo necessária a criação de ações preventivas em paralelo. Com a aplicação dos algoritmos Apriori e FP-Growth, foram extraídas Regras de Associação que geraram uma análise preditiva de dados de ocorrências de crimes relatados nas delegacias do Recife, como o fato de um crime ocorrer no período da noite implica que sua natureza seja homicídio. Espera-se assim auxiliar a polícia nas investigações e estratégias para um combate mais eficaz à criminalidade.

**Palavras-Chave:** Criminalidade na cidade do Recife; Delegacias do Recife; Mineração de Dados; Regras de Associação; Apriori; FP-Growth.

## Abstract

---

The rising numbers of crime in the city of Recife have prompted the State Government to increase the fleet of police officers on the streets in the first quarter of 2018, expecting a decrease in crime. However, combating violence only in a reactive way is an insufficient strategy for a significant reduction, because to the complexity of the criminal profile of each locality, making difficult the tactical organization of the police, and it is necessary to create preventive actions in parallel. With the application of the algorithms Apriori and FP-Growth, Association Rules were extracted that generated a predictive analysis of data of occurrences of crimes reported in the police stations of Recife, as the fact that a crime occurs in the night period implies that its nature is homicide. It's hoped to assist the police in investigations and strategies for a more effective fight against crime.

**Key-words:** Criminality in the city of Recife; Police stations in Recife; Data Mining; Association Rules; Apriori; FP-Growth.

### 1 Introdução

Em Pernambuco, temos visto a criminalidade crescer nos últimos anos, num ritmo três vezes maior do que o Rio de Janeiro, cidade que possui a maior fama de ser o epicentro da violência no país, apresentando de 2014 a 2017 um avanço de 37,8%, comparado aos 12,6% do Rio [1]. Em paralelo, Recife atingiu em 2017, a 22ª posição no ranking das cidades mais violentas do mundo, pela ONG Mexicana Segurança, Justiça e Paz, com uma média de 54,43 homicídios a cada 100 mil habitantes [2]. Analisando os números de 2006 a 2016, percebe-se que o Estado está voltando ao patamar de dez anos atrás, sendo alguns desses anos de sucesso na redução de crimes com a implantação do programa Pacto Pela Vida [3].

O Governo de Pernambuco aumentou a frota de policiais em 1.214 novos contratos realizados nos primeiros meses de 2018, visando reprimir a violência com a reativação de delegacias que estavam fechadas por falta de delegados e mais profissionais focados na investigação do tráfico de drogas, grande problema instaurado na cidade do Recife [4]. Embora essas medidas reativas não sejam suficientes para obter uma queda significativa nos números de crimes, pois assim como foi feito em Medellín, é preciso investir em inteligência de dados para maior eficácia no combate à criminalidade. Segundo Weriqui Bezerra, “quando você parar de olhar para o problema, talvez encontrará a solução que tanto procura. Não foque nos sintomas apenas, mas na raiz do problema”.

A cidade colombiana Medellín, é um exemplo de sucesso na redução da violência por meio de investimento em segurança pública, inteligência e equipamentos para polícia. Nos anos 90, Medellín foi considerada a metrópole mais violenta do mundo, com média de 380 homicídios por 100 mil habitantes. Após aplicação de tecnologia e maior capacitação da polícia, a cidade possui atualmente a taxa de 21 homicídios por 100 mil, tornando-se uma cidade-modelo em um país subdesenvolvido [5].

Foi realizado um estudo utilizando uma abordagem estatística para analisar possíveis relações entre variáveis de ocorrências na cidade do Recife [6]. Os resultados descritivos foram

obtidos através de testes de relações conduzidos manualmente pelo usuário, o que traz algumas limitações para alcançar um nível maior de abrangência e detalhe nesses resultados. Com o objetivo de realizar uma análise preditiva dos dados de ocorrências criminais da capital Pernambucana, através da utilização da mineração de dados para encontrar possíveis padrões frequentes de crimes. Com o presente estudo espera-se auxiliar os órgãos competentes à segurança pública a compreender possíveis relações associativas entre as variáveis cadastradas nos boletins de ocorrências nas delegacias da cidade, para contribuir com a promoção de estratégias preventivas de combate ao crime.

### 2 Fundamentação teórica

#### 2.1 Trabalhos relacionados

Em busca de auxiliar a segurança pública a traçar estratégias mais assertivas no combate ao crime, no Brasil já existem alguns trabalhos com objetivos semelhantes tendo a criminalidade como centro do estudo. Na Universidade Federal Fluminense, foi construída uma ferramenta web chamada SiAPP (Sistema de Apoio ao Policiamento Preditivo) utilizando dados referentes a ocorrências da cidade de Niterói, em busca da predição de crimes a partir do aprendizado de máquina. Foram usados 781 registros que geraram 281 pontos de importância na cidade. Por meio da geolocalização e uso de algoritmos, foram extraídas regras de associação visualizadas em um mapa da região, onde cada regra foi evidenciada pelo bairro a qual pertence. Após a visualização de resultados e análise, foram encontradas indicações do comportamento da criminalidade em determinados pontos, como por exemplo, que houve um aumento na ocorrência de furtos entorno de escolas em dias comerciais e aumento de roubos de carteira na região do centro a noite [7].

Outro estudo foi realizado na Universidade Federal de Viçosa, com o objetivo de realizar estudos em relação a segurança pública em cidades de pequeno porte, com a implementação

de um sistema SIG (Sistemas de Informações Geográficas) para auxiliar no cadastro de atividades e mapeamento de crimes, já que cidades pequenas geralmente não possuem recursos tecnológicos avançados. Foi utilizada uma base disponibilizada pela Polícia Militar da cidade de Rio Pomba, com dados de 2009, 2010 e 2011. Com isso, foram utilizados os algoritmos de redes neurais para classificar o quanto a polícia deve atuar nas ações de combate e prevenção para alcançar a diminuição de crimes, e Apriori para extrair padrões criminais relevantes para tornar as ações policiais mais eficazes contra a violência [8].

Ao analisar os resultados obtidos nesses estudos, pode-se compreender o quanto as regras de associação extraídas dos algoritmos aplicados às bases de dados, podem ser relevantes para pesquisa e análise em relação ao combate a crimes. Os resultados gerados dificilmente seriam encontrados sem mineração de dados, por conta do tamanho e diversidade dos registros de ocorrências criminais. A expectativa é que assim como foi feito em Niterói e Viçosa, sejam encontrados padrões frequentes de crimes da cidade do Recife.

## 2.2 Regras de Associação

Por sua grande capacidade de aplicações, é umas das técnicas de mineração de dados mais utilizadas atualmente em campanhas de marketing, logística no comércio, atividades que atuam com criação de estratégias e tomada de decisão [9]. De origem descritiva, estas aplicações têm como princípio descobrir possíveis associações, da presença de um item com outro em uma mesma transação, ou seja, na mesma operação de consulta, no conjunto de dados, encontrando padrões ou tendências frequentes relevantes para análise, em forma de regras de associação.

O problema clássico que mostra o poder das regras de associação surgiu com a observação da cesta de compras dos clientes de uma rede de supermercados dos Estados Unidos, para descobrir relações entre os itens comprados juntos frequentemente. Descobriu-se uma regra curiosa, que indicava {cerveja} -> {fraldas}, informação essa que após análise, concluiu que homens que iam comprar fraldas para os filhos, aproveitavam para levar cerveja. Seguindo o

resultado do estudo, os produtos foram colocados um ao lado do outro, e as vendas aumentaram 30%.

Uma regra de associação é uma implicação da forma  $A \Rightarrow B$ , onde A e B são conjuntos de itens, também chamados de *itemsets*, e  $A \cap B = \emptyset$ . A regra  $A \Rightarrow B$  vale no conjunto de transações T com suporte s, onde s é a porcentagem de transações em T que contém  $A \cup B$ , ou seja, pode ser entendida como a relevância estatística de uma regra (1). A regra  $A \Rightarrow B$  tem confiança c no conjunto de transações T, onde c é a porcentagem de transações em T contendo A que também contém B, sendo considerada a certeza da ocorrência de uma regra (2). A determinação do suporte e da confiança é de extrema importância, pois serve para eliminar regras que sejam pouco significativas por apresentarem valores muito baixos. Outro fator para verificação de regras consideráveis, é o lift, capaz de indicar o quanto B é frequente, quando A aparece (3). A convicção é uma medida responsável por indicar a força de uma implicação, tentando exprimir até que ponto o antecedente de determinada regra e a negação do conseqüente dessa regra são independentes (4). Além disso, dois parâmetros devem ser informados pelo usuário ao algoritmo utilizado: o *minSup* e *minConf*, que são respectivamente, um valor mínimo para o suporte e um valor mínimo para a confiança. Uma regra é dita como frequente, se atende a um suporte mínimo. E caso atenda um suporte e confiança mínima, é dita como uma regra forte.

$$Sup(A \Rightarrow B) = \frac{P(A \cup B)}{|T|} \quad (1)$$

$$Conf(A \Rightarrow B) = \frac{P(A | B)}{|T|} \quad (2)$$

$$Lift(A \Rightarrow B) = \frac{Conf(A \Rightarrow B)}{Sup(B)} \quad (3)$$

$$Conv(A \Rightarrow B) = \frac{Sup(A) \times Sup(\neg B)}{Sup(A \cup \neg B)} \quad (4)$$

### 2.2.1 Apriori

O algoritmo foi proposto em 1994, por Agrawal e Srikant, foi o pioneiro, sendo atualmente o mais famoso e utilizado quando o assunto é regras de

<http://dx.doi.org/10.25286/rep.v3i3.974>

associação por ele ser eficaz em encontrar *itemsets* frequentes em grandes bancos de dados, gerando posteriormente regras fortes de associação. O nome Apriori, é baseado no fato do algoritmo usar conhecimento *a priori* para estudo das características dos *itemsets* frequentes [10]. O funcionamento do Apriori, é dividido em duas partes, geração do conjunto de itens frequentes e geração das regras, ocasionando uma economia de custo computacional.

Primeiramente ocorre uma geração de um conjunto de *k-itemsets* candidatos, logo após ele percorre toda base de dados, verificando se os candidatos são mesmo frequentes, ou seja, aqueles que possuem o suporte com valor maior do que o *minSup* determinado, gerando um conjunto de apenas itens frequentes.

De posse de um conjunto de *k-itemsets* frequentes, com  $k \geq 2$ , é gerado as regras de associação, de forma que caso os itens AB e ABCD sejam frequentes, podemos avaliar a regra  $AB \Rightarrow CD$ , aplicando o cálculo da confiança, sendo  $conf(AB \Rightarrow CD) = sup(ABCD)/sup(AB)$ . Caso o valor da confiança seja maior ou igual ao *minConf* determinado, a regra é considerada válida.

O princípio dos *Itemsets* Frequentes do algoritmo, diz que se um itemset é frequente, todos os seus subconjuntos não vazios também são frequentes, logo todo *itemset* que não é frequente, da mesma forma seus subconjuntos também não são frequentes. Este preceito só é válido de acordo com a propriedade anti-monotônica do suporte, que garante que o suporte de um conjunto frequente nunca exceda o suporte de seus subconjuntos (5).

$$\forall A, B: (A \subseteq B) \Rightarrow s(A) \geq s(B) \quad (5)$$

A validade do princípio permite que ao identificar um conjunto de itens não frequentes, o subgrafo que o contém pode ser podado, evitando que todos os nós sejam visitados desnecessariamente, consequentemente economizando tempo de execução.

### 2.2.2 FP-Growth

Com o uso do Apriori e de outros algoritmos que utilizam abordagem semelhante, foram

encontradas algumas dificuldades, como a execução de muitos acessos ao banco de dados e no tratamento de uma grande quantidade de conjuntos de itens candidatos, ocasionados por um grande número de itens frequentes ou caso o valor do *minSup* seja muito baixo. Para solucionar essas e outras limitações, foi desenvolvido por Han, Pey e Yin, em 2000, o algoritmo FP-Growth (*Frequent Pattern Growth*) [11].

Baseado em uma estrutura de dados de árvore de prefixos para padrões frequentes, usada para extração dos conjuntos de itens constantes na própria estrutura, capaz de armazenar essas informações de forma compactada, permitindo uma mineração de dados bastante eficaz pois não necessita de vários acessos a base de dados, sendo apenas duas vezes, uma para encontrar e ordenar os *itemsets* frequentes e outra para construção da árvore, chamada de FP-Tree (*Frequent Pattern Tree*).

A solução do algoritmo é dada através de três pilares, a compactação do banco de dados é realizada dando lugar a uma estrutura geralmente bastante menor em uma árvore FP-Tree, em seguida é usado o algoritmo para minerar a árvore em busca de evitar uma grande geração de conjuntos de itens candidatos, e por fim, as tarefas de mineração são decompostas em tarefas menores usando o método particional. O processo de construção da FP-Tree acontece primeiramente após a seleção do valor do *minSup*, com a varredura da base de dados e o armazenamento e ordenação decrescente dos conjuntos de itens frequentes encontrados.

## 3 Metodologia

A metodologia empregada neste trabalho está baseada em um dos mais conhecidos processos para realização de um projeto de mineração de dados chamado CRISP-DM (*Cross Industry Standard Process for Data Mining*). Criado em 1996, para ser aplicado em diversas áreas de negócios, sem a dependência de uma ferramenta, o modelo é composto por seis etapas não rigorosas que podem prosseguir ou retornar em diferentes fases, possuindo o caráter cíclico da mineração de dados, como mostra a Figura 1.

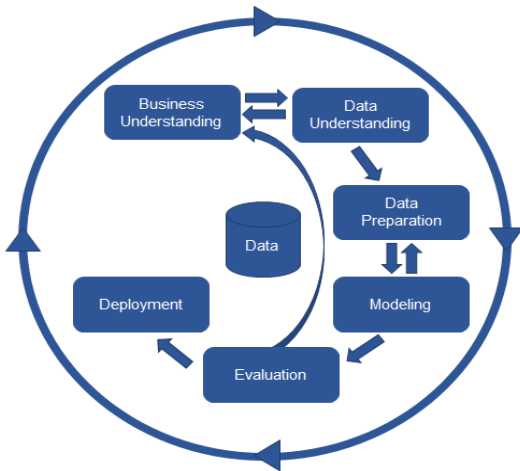


Figura 1 – Ciclo de vida do modelo CRISP-DM

### 3.1 Entendimento do negócio

Para realização deste estudo, foram fornecidas pela Secretaria de Defesa Social de Pernambuco, cinco bases de dados, referentes a ocorrências criminais registradas nas delegacias do Estado ou pela *internet* no período de 1 de Janeiro à 30 de Setembro de 2017. Cada base é caracterizada pela separação por natureza do crime, onde tem seus registros relacionados a apenas essa natureza identificada no cadastro do boletim de ocorrência, sendo elas: CVLI (Crimes Violentos Letais Intencionais), CVP (Crimes Violentos contra o Patrimônio), Furto, Furto de Veículo e Roubo de Veículo. De posse do banco de dados de ocorrências de crimes, é possível analisar, preparar e minerar os dados para encontrar regras de associação entre os registros, ou seja, relacionamentos frequentes entre determinados atributos.

### 3.2 Entendimento dos dados

A base CVLI contém 9 variáveis: ÁREA, DATA, MÊS, PERÍODO DO DIA, IDADE, SEXO, NATUREZA JURÍDICA, LOCAL GENÉRICO, TOTAL. A base CVP contém 5 variáveis: ÁREA, LOCAL, DATA, PERÍODO DO DIA, TOTAL. As base de Roubo e Furto de Veículo contém as mesmas 5 variáveis: ÁREA, DATA, NATUREZA, PERÍODO, TOTAL. E por último a base de Furto contém 5 variáveis: ÁREA, DATA, LOCAL, PERÍODO, TOTAL.

Todas as bases possuem o atributo ÁREA como informação geográfica da localidade onde

ocorreu o crime, sendo os registros desse campo, dados por AIS (Área Integrada de Segurança), divisão territorial feita em Pernambuco para melhor acompanhamento de ações e resultados, gerando 26 AIS para todo o Estado, porém nas bases de dados fornecida, encontram-se apenas registros das AIS referentes a capital Pernambucana, sendo elas:

- AIS 1 - Santo Amaro (compreende bairros como Boa Vista, São José, Ilha Joana Bezerra).
- AIS 2 - Espinheiro (compreende bairros como Cordeiro, Água Fria, Campo Grande).
- AIS 3 - Boa Viagem (compreende bairros como Ipsep, Jordão, COHAB).
- AIS 4 - Várzea (compreende bairros como Jardim São Paulo, Afogados, Mustardinha).
- AIS 5 - Apipucos (compreende bairros como Casa Amarela, Vasco da Gama, Macaxeira).

Para melhor entendimento dos atributos e seus registros presentes nas seis bases de dados, será descrito informações e significados no quadro 1 referente ao dicionário de dados:

Quadro 1 – Dicionário de Dados dos campos presentes nas bases utilizadas (continua)

CAMPO	DESCRIÇÃO
ÁREA	Atributo nominal relacionado a área onde ocorreu o crime (AIS1, AIS2, AIS3, AIS4 e AIS5).
MÊS	Atributo nominal derivado da variável DATA, que informa o mês da ocorrência do crime.
PERÍODO	Atributo nominal que informa o período do dia em que o crime aconteceu (manhã, tarde, noite ou madrugada).
NATUREZA	Atributo nominal que informa a natureza do crime (CVP, furto, homicídio, roubo de veículo, etc.).

Quadro 1 – Dicionário de Dados dos campos presentes nas bases utilizadas (continuação)

CAMPO	DESCRIÇÃO
DIA_SEMANA	Atributo nominal que informa o dia da semana que ocorreu o crime.

SEXO	Atributo nominal que informa o gênero da vítima (feminino, masculino ou desconhecido).
LOCAL	Atributo nominal que informa o local que o crime ocorreu.

### 3.3 Preparação dos dados

Após um estudo das bases de dados fornecidas e dos tipos de entradas necessárias para aplicação dos algoritmos de regras de associação, Apriori e FP-Growth, foi realizado um pré-processamento dos dados para obtenção de melhores resultados.

O primeiro algoritmo utilizado foi o Apriori, e ele admite apenas variáveis nominais, pensando nisso, primeiramente foi criada uma nova base chamada de INTEGRADA, com a junção de variáveis em comum e relevantes entre as cinco bases, considerando campos com possibilidade de transformação das variáveis de quantitativas para categóricas, foram selecionados cinco atributos: ÁREA, MÊS, PERÍODO, NATUREZA, DIA, SEMANA.

Cada base foi analisada individualmente, e também sofreram transformações. A base CVLI, que tinha 9 variáveis, ficou com 5, todas nominais: ÁREA, PERÍODO, SEXO, NATUREZA, LOCAL. As bases CVP e Furto, dos 5 atributos, ficaram 4: ÁREA, MÊS, LOCAL, PERÍODO. E as bases Furto de Veículo e Roubo de Veículo, que tinham 5 variáveis, ficaram com 3: ÁREA, MÊS, PERÍODO.

### 3.1 Modelagem

Com as bases de dados devidamente tratadas, foi escolhido para minerar os dados o *software Weka* na versão pois possui alguns algoritmos já implementados agregados a ferramenta, e uma das técnicas presentes são regras de associação, com os algoritmos Apriori e FP-Growth, os quais podem ser aplicados nas bases fornecidas, resultando em padrões frequentes nos registros. Para o algoritmo Apriori foi usado o banco de dados transformado, com todas as variáveis categóricas. Para aplicação do algoritmo FP-Growth na ferramenta *Weka*, foi realizada uma transformação nos atributos, tornando todos binários.

### 3.4 Avaliação

Durante a aplicação dos algoritmos na ferramenta *Weka*, foram experimentados vários valores de *minSup* e *minConf* para extrair as melhores regras. Foi percebido que valores muito altos, conseguiram extrair poucas ou nenhuma regra por conta da grande variedade de registros nas bases, e insistir nesses valores poderia ocasionar a perda de alguns conjuntos de itens raros e relevantes. Foi preciso realizar diversos testes até encontrar os valores ideais de *minSup* e *minConf* como parâmetros para serem aplicados em cada base de dados.

### 3.5 Desenvolvimento

Para esta pesquisa não foi realizada a etapa de desenvolvimento de um sistema, mas sim a entrega da análise dos resultados gerados pela aplicação de regras de associação por meio deste artigo.

## 4 Experimentos e Resultados

Inicialmente foi aplicado o algoritmo Apriori, nas seis bases individualmente, gerando diferentes regras relevantes para análise. Para a base Integrada, que possui 52.873 registros foi escolhido um *minSup* de 0,05 e um *minConf* de 0,5 que geraram os *itemsets* candidatos na tabela 1.

Tabela 1 – Conjunto de itens candidatos da base Integrada

k	Quantidade
1	28
2	49
3	1

Dos itens candidatos encontrados, podemos destacar que as áreas de maior ocorrência são as AIS 1 e AIS 3, a natureza dos crimes mais cometidos é a CVP, sendo a maior parte dos crimes ocorridos à noite, os meses de mais registros foram Agosto e Janeiro, e o dia da

semana de maior incidência foi a sexta-feira. Logo após a geração dos candidatos, foram geradas as melhores regras, onde algumas delas e os valores de sua Confiança estão descritas respectivamente na tabela 2.

**Tabela 2** - Regras de associação geradas para a base Integrada

Regra	Confiança
{PERÍODO=Noite} ⇒ {NATUREZA=CVP}	0,66
{ÁREA=AI3, PERÍODO=Noite} ⇒ {NATUREZA=CVP}	0,65
{ÁREA=AI4} ⇒ {NATUREZA=CVP}	0,60
{MÊS=Janeiro} ⇒ {NATUREZA=CVP}	0,59
{DIA_SEMANA=Segunda} ⇒ {NATUREZA=CVP}	0,58

Para a base CVLI, que possui 601 registros foi escolhido um *minSup* de 0,05 e um *minConf* de 0,5 que geraram os *itemsets* candidatos na Tabela 3.

**Tabela 3** – Conjunto de itens candidatos da base CVLI

k	Quantidade
1	8
2	11
3	6
4	4

Dos itens candidatos encontrados, podemos analisar que a área de maior incidência de casos tem seus dados não informados, sendo caracterizada pela sigla NI(Não Informada), a natureza dos crimes mais cometidos foi Homicídio, sendo a maior parte dos crimes ocorridos no período da noite, a grande maioria das vítimas é do sexo masculino, o local de maior ocorrência de crimes acontece em logradouros público, e o mês com mais registros de crimes CVLI foi Setembro. Logo após a geração dos candidatos, foram geradas as melhores regras, onde algumas delas e os valores de sua Confiança estão descritas respectivamente na tabela 4.

**Tabela 4** – Regras de associação geradas para a base CVLI (continua)

Regra	Confiança
{MÊS=Setembro} ⇒ {NATUREZA=Homicídio}	0,98
{PERÍODO=Noite} ⇒ {NATUREZA=Homicídio}	0,98
{SEXO=Masculino, LOCAL= Logradouro público} ⇒ {NATUREZA=Homicídio}	0,97
{PERÍODO=Noite, LOCAL=Logradouro público, NATUREZA=Homicídio} ⇒ {SEXO=Masculino}	0,97
{PERÍODO=Manhã, LOCAL=Logradouro público, NATUREZA=Homicídio} ⇒ {SEXO=Masculino}	0,95

Para a base CVP, que possui 18.026 registros foi escolhido um *minSup* de 0,01 e um *minConf* de 0,3 que geraram os *itemsets* candidatos na tabela 5.

**Tabela 5** - Conjunto de itens candidatos da base CVP

k	Quantidade
1	36
2	161
3	25

Dos itens candidatos encontrados, podemos destacar que a área com mais registros de Crimes Violentos contra o Patrimônio foi a AIS1, ocorridos principalmente em Vias públicas, sendo a maioria dos delitos cometidos no período da noite, e o mês de maior incidência desse tipo de crime foi em Janeiro. Logo após a geração dos candidatos, foram geradas as melhores regras, onde algumas delas e os valores de sua Confiança estão descritas respectivamente na tabela 6.

**Tabela 6** – Regras de associação geradas para a base CVP

Regra	Confiança
{ÁREA=AI4, MÊS=Janeiro} ⇒ {PERÍODO=Noite}	0,44
{LOCAL=Praça pública} ⇒ {ÁREA=AI1}	0,42
{LOCAL=Ônibus} ⇒ {PERÍODO=Noite}	0,40

{ÁREA=AIS3, LOCAL=Calçada} ⇒ {PERÍODO=Noite}	0,38
{ÁREA=AIS5, PERÍODO=Manhã} ⇒ {LOCAL=Via pública}	0,35

Para a base Furto, que possui 14.154 registros, foi escolhido um *minSup* de 0,01 e um *minConf* de 0,3, que geraram os itemsets candidatos na tabela 7.

**Tabela 7** – Conjunto de itens candidatos da base Furto

k	Quantidade
1	40
2	156
3	9

Dos itens candidatos encontrados, podemos destacar que a área com maior incidência de furtos foi a AIS 1, sendo a maioria deles ocorridos em Vias públicas, e principalmente acontecidos no período da tarde, com Agosto como mês de maior ocorrências. Logo após a geração dos candidatos, foram geradas as melhores regras, onde algumas delas e os valores de sua Confiança estão descritas respectivamente na tabela 8.

**Tabela 8** - Regras de associação geradas para a base Furto

Regra	Confiança
{LOCAL=Parada de ônibus} ⇒ {ÁREA=AIS1}	0,57
{LOCAL=Shopping Center} ⇒ {ÁREA=AIS3}	0,54
{LOCAL=Ônibus, PERÍODO=Tarde} ⇒ {ÁREA=AIS1}	0,44
{LOCAL=Calçada} ⇒ {PERÍODO=Tarde}	0,37
{LOCAL=Residência térrea} ⇒ {PERÍODO=Madrugada}	0,35

Para a base Furto de Veículo, que possui 1.098 registros, foi escolhido um *minSup* de 0,01 e um *minConf* de 0,3, que geraram os itemsets candidatos tabela 9.

**Tabela 9** - Conjunto de itens candidatos da base Furto de Veículo

k	Quantidade
1	18
2	100
3	20

Dos itens candidatos encontrados, podemos destacar que a área com mais registros de furto de veículos é a AIS4, sendo a maioria dos crimes ocorridos no período da madrugada, e o mês de maior incidência foi Maio. Logo após a geração dos candidatos, foram geradas as melhores regras, onde algumas delas e os valores de sua Confiança estão descritas respectivamente na tabela 10.

**Tabela 10** – Regras de associação geradas para a base Furto de Veículo

Regra	Confiança
{ÁREA=AIS2, MÊS=Janeiro} ⇒ {PERÍODO=Madrugada}	0,52
{PERÍODO=Tarde, MÊS=Janeiro} ⇒ {ÁREA=AIS3}	0,52
{ÁREA=AIS5, MÊS=Março} ⇒ {PERÍODO=Madrugada}	0,50
{ÁREA=AIS2, MÊS=Julho} ⇒ {PERÍODO=Tarde}	0,44
{PERÍODO=Madrugada, MÊS=Fevereiro} ⇒ {ÁREA=AIS4}	0,43

E por fim, para a base Roubo de Veículo, que possui 1.884 registros, foi escolhido um *minSup* de 0,01 e um *minConf* de 0,3, que geraram os itemsets candidatos tabela 11.

**Tabela 11** – Conjunto de itens candidatos da base Roubo de Veículo

k	Quantidade
1	18
2	82
3	35



Dos itens candidatos encontrados, podemos destacar que a área com a maioria das ocorrências foi a AIS4, onde a maior parte dos registros no período da noite, e Março foi o mês com mais ocorrências de roubo de veículos. Logo após a geração dos candidatos, foram geradas as melhores regras, onde algumas delas e os valores de sua Confiança estão descritas na tabela 12.

**Tabela 12** – Regras de associação geradas para a base Roubo de Veículo

Regra	Confiança
{ÁREA=AIS5, MÊS=Janeiro} ⇒ {PERÍODO=Noite}	0,69
{ÁREA=AIS2, MÊS=Setembro} ⇒ {PERÍODO=Noite}	0,51
{ÁREA=AIS2, MÊS=Março} ⇒ {PERÍODO=Noite}	0,48
{ÁREA=AIS3, MÊS=Maio} ⇒ {PERÍODO=Noite}	0,45
{ÁREA=AIS4, MÊS=Março} ⇒ {PERÍODO=Noite}	0,42

Em seguida foi realizada uma transformação nos dados na própria ferramenta *Weka*, para converter os atributos de nominais para binários, para aplicar o algoritmo FP-Growth. Foram utilizados os mesmos parâmetros de *minSup* e *minConf* para cada base usados no algoritmo Apriori. Após a aplicação do FP-Growth, foi observado que as regras extraídas e o valor de sua confiança, foram as mesmas do Apriori, corroborando a relevância das regras geradas.

## 5. Análise e discussão

Em comparativo com o estudo estatístico [6], percebe-se que há uma confirmação entre os resultados encontrados entre as pesquisas, porém foram encontradas novas associações entre as variáveis existentes no banco de dados que oferecem mais detalhes em relação ao comportamento criminal. Informações como o dia

da semana com mais registros, sendo a sexta-feira, a AIS 1 liderando as ocorrências referentes a Crimes Violentos contra o Patrimônio, além da AIS 4 com maior ocorrência de furtos e roubos de veículos, foram confirmadas pela geração de itens candidatos na aplicação do algoritmo Apriori.

Uma observação pode ser realizada considerando a natureza de crimes para exemplificar e comprovar a efetividade dos resultados dos estudos feitos com o mesmo banco de dados, onde a maior quantidade de registros é de Crimes Violentos ao Patrimônio, de forma que no estudo anterior foi identificado que 40,45% dos crimes dessa natureza acontecem à noite, e no presente estudo encontramos uma regra que mostra que {PERÍODO=Noite}⇒{NATUREZA=CVP} com 0,66 de Confiança, fortalecendo a veracidade e importância dos fatos para análise.

Por possuir um nível alto de granularidade dos dados das ocorrências de crimes, os resultados encontrados no estudo de Viçosa sugerem poucas informações sobre os crimes, como na regra que indica que {DATA=Julho, SETOR=S8, TURNO=Noite}⇒{GRUPO=B}, onde o atributo SETOR refere-se às subdivisões da cidade e o atributo GRUPO refere-se ao tipo de crime, onde B equivale a crimes contra pessoas. No estudo realizado em Niterói, como possui um nível mais baixo de granularidade dos dados das ocorrências de crimes, os resultados encontrados foram mais objetivos pela presença da localização geográfica do crime, como por exemplo, na regra que indica que {BAIRRO=Icaraí, PERÍODO=Manhã, ENTORNO=Banco}⇒{OCORRÊNCIA=Furto}, enquanto possuímos como localização apenas as AIS.

## 6. Considerações Finais

De posse dos resultados em forma de regras de associação, temos a representação de padrões de relacionamento frequentes entre itens das bases de dados estudadas, que permitem um aprofundamento em pesquisas e análises em relação a essas regras. Os resultados mostraram que as predições geradas tiveram um valor de Confiança médio maior que 57%, confirmando a relevância das informações, sendo possível através delas, que a Polícia atue de forma mais

<http://dx.doi.org/10.25286/rep.v3i3.974>

efetiva no combate e prevenção de crimes, pois eles saberão por exemplo quais dias ou períodos precisarão intensificar a frota de policiais nas ruas, assim como determinada AIS que apresenta maior ocorrência de crimes precisa de maior prioridade nas investigações.

Um fator importante para estudos posteriores, seria o fornecimento de novos registros referentes à todo Estado de Pernambuco, e não apenas da capital, permitindo um estudo geral da Segurança Pública do Estado. Para que regras mais robustas e com maiores valores de confiança sejam geradas, é necessário a inclusão de mais variáveis às bases de dados fornecidas pela Secretaria de Defesa Social de Pernambuco, que possam mostrar mais detalhes sobre as ocorrências registradas, possibilitando uma análise urbanística, geográfica e comportamental dos crimes ocorridos em Recife.

Com este estudo damos um passo para pesquisas maiores, para que com o uso da mineração de dados, possa haver uma contribuição no auxílio à tomada de decisões para um melhor planejamento das rondas policiais de acordo com a AIS e a natureza dos crimes, foco em investigações para combate a determinados crimes frequentes, e com a criação de novas estratégias de medidas preventivas. Com o uso desta pesquisa, espera-se cooperar com as autoridades e órgãos responsáveis pela Segurança Pública, para alcançarmos uma redução significativa na violência da cidade do Recife.

## Agradecimentos

À Secretaria de Defesa Social de Pernambuco, pelo fornecimento das bases de dados utilizadas neste trabalho.

## Referências

[1] FALCÃO, Marina. Violência em Pernambuco é maior e homicídios crescem mais que no Rio. **Valor Econômico**. Disponível em:

<<http://www.valor.com.br/brasil/5138178/violencia-em-pernambuco-e-maior-e-homicidios-crescem-mais-que-no-rio>> Acesso: 21 abr. 2018.

[2] JC ONLINE. Recife é a 22ª cidade mais violenta do mundo, segundo ONG Mexicana. **Jornal do Comercio**, Recife, 7 mar. 2018. Disponível em:

<<http://jconline.ne10.uol.com.br/canal/cidades/policia/noticia/2018/03/07/recife-e-a-22-cidade-mais-violenta-do-mundo-segundo-ong-mexicana-330506.php>> Acesso: 22 abr. 2018.

[3] MONTESANI, Beatriz. Por que a violência sobe de forma preocupante em Pernambuco, segundo especialista. **Nexo Jornal**, São Paulo, 18 abr. 2017. Disponível em:

<<https://www.nexojornal.com.br/entrevista/2017/04/18/Por-que-a-viol%C3%Aancia-sobe-de-forma-preocupante-em-Pernambuco-segundo-este-especialista>> Acesso: 10 mai. 2018.

[4] CAVALCANTE, Diogo. Com novos policiais, Governo de PE espera reforçar delegacias. **Folha de Pernambuco**, Recife, 5 fev. 2018. Disponível em:

<<https://www.folhape.com.br/noticias/noticias/coltidiano/2018/02/05/NWS,57935,70,449,NOTICIAS,2190-COM-NOVOS-POLICIAIS-GOVERNO-ESPERA-REFORCAR-DELEGACIAS.aspx>> Acesso em: 21 abr. 2018.

[5] SANT'ANNA, Lourival. Como Medellín virou a cidade-modelo que está vencendo o crime.

**Revista Exame**, 5 out. 2017. Disponível em:

<<https://exame.abril.com.br/revista-exame/menos-violenta-e-mais-prospera/>> Acesso em: 10 mai. 2018.

[6] MELO, Carolina; RODRIGUES, Rodrigo Lins; CAVALCANTI, Bettina. **Análise de relações entre variáveis de ocorrências de crimes da cidade do Recife**. Universidade de Pernambuco, 2018.

[7] LOURENÇO, Vitor; MANN, Paulo; PAES, Aline; OLIVEIRA, Daniel de. SiAPP: Um sistema para análise de ocorrências de crimes baseado em aprendizado lógico-relacional. In: BRAZILIAN SYMPOSIUM ON INFORMATION SYSTEMS, 12., 2016, Florianópolis. **Anais...** Florianópolis: SBSI, 2016. p.168-175.

[8] LAMAS, João Paulo Campolina. **Predição de crimes e otimização de ações de segurança pública para cidades de pequeno porte utilizando geotecnologias**. Tese de Doutorado. Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Viçosa. Viçosa, 2013. Disponível em: <<http://www.locus.ufv.br/bitstream/handle/123456789/840/texto%20completo.pdf?sequence=1>>

[9] BARANAUSKAS, José Augusto. **Regras de Associação**. Departamento de Física e Matemática, Universidade de São Paulo. Disponível em: <<http://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-Regras-Associacao.pdf>>

[10] AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. Fast algorithms for mining association rules. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, Santiago. **Proceedings...** Santiago, 1994. p. 487-499. Disponível em: <<http://www.vldb.org/conf/1994/P487.PDF>>

[11] HAN, Jiawei.; PEI, Jian; YIN, Yiwen. Mining Frequent Patterns Without Candidate Generation In: INTERNATIONAL CONFERENCE 2000 ON MANAGEMENT OF DATA, 2002, Dallas. **Proceedings...** Dallas: ACM SIGMOD, 2002. p. 1-12.