

UNIVERSIDADE FEDERAL DO PARANÁ

KETHELLYN J. SILVEIRA DOS SANTOS

**ANÁLISE CRIMINAL NA *GLOBAL TERRORISM DATABASE***

CURITIBA  
2017

KETHELLYN J. SILVEIRA DOS SANTOS

**ANÁLISE CRIMINAL NA *GLOBAL TERRORISM DATABASE***

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção de grau de Bacharel no Curso de Gestão da Informação, Departamento de Ciência e Gestão da Informação do Setor de Ciências Sociais Aplicadas, da Universidade Federal do Paraná.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Denise Fukumi Tsunoda.

CURITIBA  
2017

## **TERMO DE APROVAÇÃO**

KETHELLYN J. SILVEIRA DOS SANTOS

### **ANÁLISE CRIMINAL NA *GLOBAL TERRORISM DATABASE***

Trabalho apresentado como requisito parcial à obtenção do grau de bacharel em Gestão da Informação no curso de graduação em Gestão da informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, pela seguinte banca examinadora:

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Denise Fukumi Tsunoda

Orientadora - Setor de Ciências Sociais Aplicadas da Universidade  
Federal, UFPR

---

Prof. Dr. Cícero Aparecido Bezerra

Setor de Ciências Sociais Aplicadas da Universidade Federal, UFPR

---

Prof. Me. André José Ribeiro Guimarães

Setor de Ciências Sociais Aplicadas da Universidade Federal, UFPR

Curitiba, 26 de junho de 2017

*Aos meus pais Jair e Adienez pelo amor, apoio e dedicação*

## **AGRADECIMENTOS**

Agradeço a Deus por permitir que eu chegasse até aqui.

Aos meus pais Jair e Adienez que com muito carinho e dedicação me acompanharam durante toda esta jornada e proporcionam todo o apoio, incentivo e motivação para que este momento chegasse.

Ao meu namorado Everton, por ter compreendido os momentos em que estive ausente e pelas palavras de apoio e coragem nos momentos de *stress*.

A minha orientadora Prof<sup>a</sup> Dr<sup>a</sup> Denise Fukumi Tsunoda pelo conhecimento e experiência compartilhados e por toda a atenção e incentivo durante as orientações.

Ao Pr<sup>o</sup> Dr<sup>o</sup> Cícero Aparecido Bezerra pelo auxílio na análise estatística.

Aos meus amigos, Daniel, Pamela, Carolina, Talita e Mariana que se mantiveram próximos a mim durante a graduação, proporcionando muitas risadas e momentos de descontração.

Vocês foram essenciais para a realização desta pesquisa.

Muito Obrigada!

*“A melhor maneira de lutar contra o terrorismo não é com armas.  
É com canetas, livros, professores e escolas”*

Malala Yousafzai

## RESUMO

Este estudo realiza uma análise criminal com base nos registros de incidentes criminais disponíveis na *global terrorism database* (GTD) por meio da realização de análise estatística e tarefas de mineração de dados. Relacionado principalmente aos incidentes ocorridos na América do Sul e no Brasil. Pesquisa aplicada quanto à finalidade, quantitativa quanto à abordagem, exploratória quanto aos objetivos. Considera o País, ano, cidade, grupo responsável, número de vítimas fatais, número de feridos, alvo, tipo de ataque, armas utilizadas e o sucesso para realização da análise estatística utilizando o SPSS® *Statistics Base* e aplicação de tarefas de classificação, associação e agrupamento com os algoritmos J48 (C4.5), Apriori e K-means. Conclui que o número de óbitos e feridos gerados no incidente está relacionado aos atributos, tipo de ataque e armas utilizadas. Sugere a continuidade de pesquisa, tendo em vista que pode ser utilizada como auxílio a prevenção de futuros ataques, faz-se necessário a atenção aos demais atributos que constituem a GTD.

Palavras-Chave: Terrorismo. Análise criminal. Análise estatística. Mineração de dados. *Global Terrorism Database*.

## **ABSTRACT**

This study aims to conduct a criminal analysis based on the records of criminal incidents available in the global terrorism data base (GTD) by performing statistical analysis and data mining tasks. It is mainly concerned with the incidents that occurred in South America and in Brazil. Applied research as to the purpose, quantitative as to the approach, exploratory regarding the objectives. It considers the country, year, city, responsible group, number of fatalities, number of injured, target, type of attack, weapons used and success for statistical analysis using SPSS® Statistics Base and application of classification, association And grouping with the algorithms J48 (C4.5), Apriori and K-means. It concludes that the number of deaths and injuries generated in the incident is related to the type of attack and arming attributes used. It suggests the continuity of research, since it can be used as an aid to prevent future attacks, it is necessary to pay attention to the other attributes that make up the GTD.

Keywords: Terrorism. Criminal Analysis. Statistical Analysis. Data Mining. Global Terrorism Database.

## LISTA DE FIGURAS

Figura 1 - Processo de análise estatística.....	28
Figura 2 - Processos de KDD.....	29
Figura 3 - Modelo CRISP-DM.....	31
Figura 4 - Árvore de decisão .....	35
Figura 5 - Resultados de uma busca na GTD .....	41
Figura 6 - Parâmetros default J48 (C4.5) .....	67
Figura 7 Árvore de decisão j48, experimento 1 (c4.5).....	68
Figura 8 - Árvore de decisão no tipo texto j48 (c4.5), experimento 1 .....	69
Figura 9 - Árvore de decisão no tipo texto j48 (c4.5), experimento 3 .....	70
Figura 10 - Árvore de decisão no tipo texto j48 (c4.5), experimento realizado com a amostra referente aos incidentes ocorridos no Brasil.....	73
Figura 11 - Parâmetros default Apriori .....	75
Figura 12 - Parâmetros default K-means .....	79
Figura 13 - Resultado Apriori e K-means apresentados no mapa.....	81

## LISTA DE TABELAS

Tabela 1 - Resultados de pesquisa realizada nas bases de dados CAPES, Scielo, Web of Science e BDTD com os termos análise criminal, terrorismo e Global Terrorismo Data Base .....	17
Tabela 2 - Filtro aplicado a pesquisa com o termo "terrorism" (CAPES).....	18
Tabela 3 - Filtro aplicado a pesquisa com o termo "terrorism" (Web of Science).....	18
Tabela 4 Publicações encontradas nas bases de dados através da busca com dois termos (Análise criminal e bases de dados, Análise criminal e mineração de dados, Terrorismo e base de dados, Terrorismo e mineração de dados, Global terrorism data base and data mining).....	18
Tabela 5 - Filtro aplicado a pesquisa com os termos Criminal analysis e database..	19
Tabela 6 - Filtro aplicado a pesquisa com os termos Terrorism e Database.....	20
Tabela 7 - Registros de ataques categorizados por região .....	43
Tabela 9 - Distribuição de ataques por país (América Central e Caribe) .....	48
Tabela 10 - Distribuição de ataques por país (América do Norte).....	50
Tabela 11 - Distribuição de ataques por país (América do Sul) .....	51
Tabela 12 - Distribuição de ataques por país (Ásia central).....	54
Tabela 13 - Distribuição de ataques por país (Ásia Meridional) .....	55
Tabela 15 - Distribuição de ataques por país (Austrália e Oceania) .....	58
Tabela 16 - Distribuição de ataques por país (Europa Ocidental).....	59
Tabela 17 - Distribuição de ataques por país (Europa Oriental).....	61
Tabela 18 - Distribuição de ataques por país (Oriente Médio e Norte da África) .....	63
Tabela 19 - Distribuição de ataques por país (Sudeste Asiático).....	65
Tabela 20 - Matriz de confusão gerada no experimento 1 utilizando o J48 (C4.5)....	68
Tabela 21 - Matriz de confusão gerada no experimento 2 com a aplicação J48 (C4.5) .....	69
Tabela 22 - Matriz de confusão gerada no experimento 3 utilizando o J48 (C4.5)....	71
Tabela 23 - Matriz de confusão gerada no experimento 4 com a aplicação J48 (C4.5) .....	71
Tabela 24 - Matriz de confusão gerada no experimento com a amostra que representa os incidentes ocorridos no Brasil utilizando o J48 (C4.5) .....	74

## LISTA DE QUADROS

Quadro 1 - Atributos da GTD considerados neste estudo.....	44
Quadro 2 - Discretização dos atributos numéricos.....	45
Quadro 3 - Experimentos realizados com o J48 (C 4.5).....	72
Quadro 5 - Regras e conclusões através do Apriori (América do Sul).....	77
Quadro 6 - Regras e conclusões através do Apriori (Brasil).....	78
Quadro 7 - Resultado K-means (América do Sul).....	80
Quadro 8 - Resultado K-means utilizando a amostra referente aos incidentes ocorridos no Brasil.....	81

## LISTA DE GRÁFICOS

Gráfico 1 - Ataques registrados na GTD ao longo dos anos (África Subsariana).....	46
Gráfico 2 - Ataques registrados na GTD ao longo dos anos (América Central e Caribe) .....	48
Gráfico 3 - Ataques registrados na GTD ao longo dos anos (América do Norte).....	50
Gráfico 4 - Ataques registrados na GTD ao longo dos anos (América do Sul).....	51
Gráfico 5 - Distribuição de mortos e feridos ao longo dos anos (América do Sul).....	52
Gráfico 6 - Distribuição de mortos e feridos ao longo dos anos (Brasil).....	53
Gráfico 7 - Ataques registrados na GTD ao longo dos anos (Ásia Central) .....	54
Gráfico 8 - Ataques registrados na GTD ao longo dos anos (Ásia Meridional) .....	55
Gráfico 9 - Ataques registrados na GTD ao longo dos anos (Ásia Oriental) .....	56
Gráfico 10 - Ataques registrados na GTD ao longo dos anos (Austrália e Oceania).58	
Gráfico 11 - Ataques registrados na GTD ao longo dos anos (Europa Ocidental) ....	59
Gráfico 12 - Ataques registrados na GTD ao longo dos anos (Europa Oriental).....	61
Gráfico 13 - Ataques registrados na GTD ao longo dos anos (Oriente Médio e Norte da África).....	63
Gráfico 14 - Ataques registrados na GTD ao longo dos anos (Sudeste Asiático) .....	65

## LISTA DE SIGLAS

- BDTD - Biblioteca Digital Brasileira de Teses e Dissertações
- DECIGI - Departamento de Ciência e Gestão da Informação
- GTD - *Global Terrorism Database*
- ONU - Organização das Nações Unidas
- CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- START - Consórcio Nacional para o Estudo do Terrorismo e Respostas ao Terrorismo
- CETIS - Centro de Estudos de Terrorismo e Inteligência
- ISVG - Instituto de Estudos de Grupos Violentos

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	15
1.1	PROBLEMA	16
1.2	OBJETIVOS GERAL E ESPECÍFICOS	16
1.3	JUSTIFICATIVA	17
1.4	ESTRUTURA DO DOCUMENTO	20
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	22
2.1	Terrorismo	22
2.1.1	Terrorismo no Brasil	24
2.2	Análise Criminal	25
2.3	Análise estatística	27
2.4	KDD (KNOWLEDGE DISCOVERY IN DATABASE)	28
2.5	MINERAÇÃO DE DADOS	30
<b>3</b>	<b>ENCAMINHAMENTOS METODOLÓGICOS</b>	38
3.1	CARACTERIZAÇÃO DA PESQUISA	38
3.1.1	Tipo de pesquisa	38
3.2	MATERIAIS E MÉTODOS	39
3.2.1	GTD	39
3.3	FERRAMENTAS	41
3.3.1	SPSS	41
3.3.2	Excel	42
3.3.3	WEKA	42
3.3.4	Power BI	43
3.4	PREPARAÇÃO DOS DADOS	43
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	46
4.1	ANÁLISE ESTATÍSTICA	46
4.2	TAREFAS DE MINERAÇÃO DE DADOS	66
4.2.1	J48	66
4.2.2	Apriori	75
4.2.3	<i>K-means</i>	78
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	83

5.1 Alcance dos objetivos .....	83
5.2 Contribuições da pesquisa .....	84
5.3 Trabalhos futuros .....	85
<b>REFERÊNCIAS</b> .....	86
<b>APÊNDICE A – Atributos GTD</b> .....	90
<b>APÊNDICE B – Distribuição de ataques por país (áfrica subsariana) .....</b>	96
<b>APÊNDICE C – Regras geradas a partir do algoritmo <i>apriori</i> (América do Sul).....</b>	98
<b>APÊNDICE D – Regras geradas a partir do algoritmo <i>Apriori</i> (Brasil)....</b>	100

## 1 INTRODUÇÃO

Os ataques terroristas fazem parte da história da humanidade desde a antiguidade, onde o “terror” era praticado como forma de mostrar à população as formas de punição aos contraventores do poder vigente. Entretanto foi somente no final do século XVIII, durante a revolução francesa que conseqüentemente foi um período de grande violência que o terrorismo passou a ser entendido como uma expressão política e caracterizado como terrorismo de estado.

Os estudos pertinentes a este tema tornaram-se mais frequentes após os ataques ao *World Trade Center* em 11 de setembro, onde o terrorismo passa a ter mais atenção por parte da comunidade internacional. Tais estudos apontam que dificilmente os atentados deixaram de acontecer, pois no pensamento dos grupos terroristas, eles estão lutando por uma causa “justa”. Segundo eles, há inimigos primordiais a serem combatidos, “os regimes autoritários brutais, corruptos e opressivos do mundo árabe”.

Atualmente 156.772 ataques terroristas ocorridos estão registrados na *Global Terrorism Database* (GTD), uma base de dados gratuita desenvolvida pela iniciativa de um grupo de estudos da universidade de *Maryland*, por meio de uma grande base de dados coletada pelos serviços de inteligência global de *Pinkerton*.

A América do Sul representa 11% do total da base com 18.628 incidentes registrados, sendo que Colômbia e Peru apresentam o maior número de ataques registrados.

O Brasil, país que historicamente não contrapõe os ideais de grupos terroristas, aparece com 267 ataques registrados na GTD até o dia 5 novembro de 2015. Cabe ressaltar que até março de 2016 o Brasil não dispunha de uma legislação específica para caracterizar crimes como ataques terroristas fazendo com que estes atos fossem investigados como outros crimes ou infrações, tais como falsificação de documentos e discriminação racial. Neste sentido, torna-se relevante a realização de uma análise criminal, a fim de verificar as possíveis relações entre os ataques ocorridos no Brasil e até mesmo sua relação com ataques ocorridos em outros países. Neste estudo tal análise é realizada através da aplicação de técnicas estatística aliadas a tarefas de mineração de dados,

utilizando-se de algoritmos previamente definidos e adequadas ao tipo de dado disponível na GTD.

## 1.1 PROBLEMA

Os atentados ao *World Trade Center* ocorridos em 11 de setembro 2001, organizados pela Al-Qaeda e os atentados a Paris em 13 e 14 de novembro de 2015 são apenas dois exemplos dos milhares de ataques terroristas ocorridos.

É importante ressaltar também que até março de 2016 o Brasil ainda não dispunha de uma legislação tratando especificamente à prática de ataques terroristas, fazendo com que estes atos fossem investigados como outros crimes ou infrações, tais como falsificação de documentos e discriminação racial acarretando em punições inadequadas aos criminosos. Neste sentido a lei 13.260/2016 tipifica os crimes envolvendo atos terroristas e estabelece as devidas punições aos praticantes.

Com isso a situação problema do presente estudo é: existem relações entre os incidentes terroristas disponíveis na *Global Terrorism Database* (GTD)?

## 1.2 OBJETIVOS GERAL E ESPECÍFICOS

Como objetivo geral define-se realizar uma análise criminal na *Global Terrorism Database* (GTD), por meio de tarefas de mineração de dados e análise estatística, dos ataques terroristas registrados na GTD, principalmente aos incidentes ocorridos na América do Sul.

Derivados do objetivo geral acima, os objetivos específicos são:

- escolher o algoritmo adequado para realizar uma efetiva mineração nos dados presentes na base;
- identificar padrões e relações entre os atributos da base de dados;
- comparar os resultados obtidos por meio deste estudo, com outras análises já realizadas na GTD.

### 1.3 JUSTIFICATIVA

O fator determinante para este tema é a falta de estudos acadêmicos realizados na GTD ainda que a base disponha atualmente de 150.000 registros de ataques.

Na tabela 1 nota-se que o termo “*Global Terrorism Database*” retorna um baixo número de publicações, quando utilizado como palavra-chave de pesquisa em bases de dados considerando que apenas em uma das quatro bases de dados pesquisadas foi encontrado um número considerável de resultados (58 publicações).

Tabela 1 - Resultados de pesquisa realizada nas bases de dados CAPES, Scielo, Web of Science e BDTD com os termos análise criminal, terrorismo e Global Terrorism Database

Base	Análise criminal	<i>Criminal analysis</i>	Terrorismo	<i>Terrorism</i>	<i>Global terrorism database</i>
Capes	248	118	997	65.747	6
Scielo	102	390	209	195	0
<i>Web of science</i>	0	7.914	8	15.302	58
BDTD	1.082	844	30	17	0
Total	1.432	9.266	1.244	81.261	64

FONTE: Elaborado pela autora (2017)

Os dados apresentados também mostram que o termo “*terrorism*” quando mencionado de forma isolada e sem nenhum outro filtro adicional (por exemplo por tópico ou assunto), retorna um número elevado de publicações, principalmente no portal de periódicos da capes e da *Web of Science*. Entretanto muitas destas publicações estão relacionadas a temas que não são o foco do presente estudo, com isso foram utilizados filtros conforme a tabela 2 abaixo:

Tabela 2 - Filtro aplicado a pesquisa com o termo "terrorism" (CAPES)

Tópico	Publicações
Terrorismo	28.095
<i>Terrorist organizations</i>	833
<i>Antiterrorism measures</i>	1.841
Total	30.769

FONTE: Elaborado pela autora (2017)

Com a utilização da filtragem por tópico disponível no portal de periódicos da Capes é possível notar que houve uma significativa diminuição no número de publicações. Os tópicos "*Terrorist Organizations*", "*Antiterrorism Measures*" e o próprio termo "*terrorism*", foram destacados, pois através destes poderia ser possível encontrar estudos que utilizaram a mineração de dados para realizar análises em bases de dados semelhantes à GTD.

Os dados obtidos na *Web of Science* foram filtrados de acordo com categorias específicas (tabela 3) disponíveis na própria base, escolhidas também com o objetivo de encontrar estudos que relacionem bases semelhantes ou até mesmo a própria GTD à mineração de dados.

Tabela 3 - Filtro aplicado a pesquisa com o termo "terrorism" (Web of Science)

CATEGORIAS	PUBLICAÇÕES
<i>Computer science artificial intelligence</i>	264
<i>Computer science interdisciplinary applications</i>	230
<i>Computer science software engineering</i>	86
<i>Computer science hardware architecture</i>	47
Total	627

FONTE: Elaborado pela autora (2017)

Apesar dos filtros utilizados nota-se ainda um elevado número de publicações, considerando que seria inviável analisá-las uma a uma, foi realizada uma nova busca, nas mesmas bases de dados agregando os termos

“mineração de dados”, “*data mining*”, “base de dados” e “*database*” a pesquisa anterior, conforme apresentado na tabela 4.

A busca com os termos *Criminal analysis* e *data base* e *Terrorism And Database* realizadas na *web of Science* foram novamente filtradas de acordo com categorias específicas (tabelas 4 ,5 e 6).

Tabela 4 - Publicações encontradas nas bases de dados através da busca com dois termos (Análise criminal e bases de dados, Análise criminal e mineração de dados, Terrorismo e base de dados, Terrorismo e mineração de dados, Global terrorism data base and data mining

BASE	Análise criminal e bases de dados	<i>Criminal analysis and data base</i>	Análise criminal e Mineração de dados	<i>Criminal analysis and Data mining</i>	Terrorismo e base de dados	<i>Terrorism And Database</i>	Terrorismo e Mineração de dados	<i>Terrorism and Data mining</i>	<i>Global terrorism data base and data mining</i>
Capes	8	2	0	1	0	90	0	171	0
Scielo	1	2	0	0	0	2	0	0	0
<i>Web of science</i>	0	315	0	87	0	286	0	89	1
BDTD	137	37	2	3	3	2	0	0	0
Total	46	356	2	91	3	380	0	260	1

FONTE: Elaborado pela autora (2017)

Tabela 5 - Filtro aplicado a pesquisa com os termos *Criminal analysis* e *database*

Categorias	Publicações
<i>Computer science theory methods</i>	33
<i>Computer science information systems</i>	27
<i>Computer science software engineering</i>	6
<i>Computer science interdisciplinary applications</i>	7
Total	73

FONTE: Elaborado pela autora (2017)

Tabela 6 - Filtro aplicado a pesquisa com os termos Terrorism e Database

CATEGORIAS	PUBLICAÇÕES
<i>Computer science artificial intelligence</i>	32
<i>Computer science theory methods</i>	27
<i>Computer science information systems</i>	6
<i>Computer science software engineering</i>	7
Total	72

FONTE: Elaborado pela autora (2017)

A partir dos dados aqui apresentados, nota-se um grande número de publicações relacionadas aos temas “*terrorism*” e “*Criminal analysis*” de forma isolada. Entretanto nota-se que foram poucos os estudos que integram ambos os termos a mineração de dados, principalmente se tratando de publicações em português, onde foram encontrados apenas dois registros.

Por fim, pode-se notar que o tema proposto abre mais um campo de estudo na área de gestão de informação, considerando que a base teórica obtida nas disciplinas ministradas no curso, mais especificamente a disciplina de mineração de dados, capacita os alunos para realizar análises em bases de dados como a que está sendo proposta sendo proposta. Com isso, pretende-se, que este estudo sirva como base para futuras pesquisas dentro do DECIGI.

#### 1.4 ESTRUTURA DO DOCUMENTO

O presente documento está organizado em cinco seções principais. A primeira delas apresenta os elementos essenciais para a realização de um projeto de pesquisa sendo eles: a problematização, os objetivos geral e específicos, e a justificativa.

A segunda seção denominada fundamentação teórica consiste no embasamento teórico necessário para a realização da análise criminal proposta, realizado através de estudos propostos por autores consagrados nos temas: terrorismo, terrorismo no Brasil, análise criminal, análise estatística e mineração de dados.

A terceira seção aborda os procedimentos metodológicos que serão utilizados no projeto. Neste sentido a pesquisa é caracterizada quanto a à sua finalidade, abordagem, objetivos e procedimentos. São apresentados os *softwares* utilizados na realização da análise estatística e tarefas de mineração de dados

A quarta seção apresenta os resultados obtidos durante a realização da análise estatística bem como em relação as tarefas de mineração de dados. Preocupa-se ainda em realizar a análise dos respectivos dados obtidos.

A quinta seção compreende as considerações finais desta pesquisa relatando o cumprimento dos objetivos propostos e a possibilidade da realização de trabalhos futuros através do presente estudo. Por fim, é apresentada a lista de referências utilizadas na elaboração deste documento.

## 2 FUNDAMENTAÇÃO TEÓRICA

Esta seção concentra-se em contextualizar a pesquisa a ser realizada, através de estudos publicados por especialistas no tema proposto. Neste sentido aborda as temáticas: terrorismo seguido pelo terrorismo no Brasil, análise criminal, análise estatística e mineração de dados.

### 2.1 Terrorismo

A palavra terrorismo deriva de terror, que por sua vez deriva do verbo em latim *terrere*, cujo significado é: “fazer alguém tremer por meio de grande medo” (SCHMID, 2011, tradução nossa).

De acordo com resolução n.49/60 de 17 de fevereiro de 1995 da assembleia geral da ONU terrorismo

São atos criminosos pretendidos ou calculados para provocar um estado de terror no público em geral, num grupo de pessoas ou em indivíduos para fins políticos são injustificáveis em qualquer circunstância, independentemente das considerações de ordem política, filosófica, ideológica, racial, étnica, religiosa ou de qualquer outra natureza que possam ser invocadas para justificá-lo. (ONU. Assembléia Geral, 1995)

Santos (2013) lembra que a existência de atos similares às ações terroristas já se faziam presente desde a antiguidade com a presença de violência física e psicológica, em que Sun Tzu já ensinava como táticas de guerra o uso de atos de terror aplicando a máxima do “Mate um, amedronte dez mil”. Porém, os atos de terror se tornaram mais conhecidos com o advento da Revolução Francesa no século XVIII, como forma de aterrorizar a população, por parte de quem detinha o poder. A autora ainda cita Mattos (2010), explicando que os atos terroristas foram alterados ao longo dos últimos séculos, sobretudo no século XX, em que passaram a ser cometidos por agentes distintos e com motivações variadas, indo do cunho tanto religioso, como ideológico, psicológico, político ou a junção de alguns desses, podendo tais atividades serem financiadas ou não por um Estado soberano ou grupos terroristas.

Alcântara (2015) afirma que a definição do termo terrorismo está relacionada com a história, a cultura e as políticas das nações e organizações

internacionais, o que torna o trabalho de alcançar um consenso quase impossível, considera ainda que tal definição está relacionada a questões políticas, crenças, valores e ideologia daquele que a define, logo, como consequência, o termo terrorismo é considerado um conceito contestado.

Dissenha (2013) entende que à dificuldade em definir o terrorismo por não se tratar de um fenômeno homogêneo, já que há diversas realidades diferentes às quais se agrega o conceito: há o terrorismo do radicalismo religioso; há o terrorismo do *freedom fighter*, em oposição a regimes políticos; há o terrorismo das associações criminosas (“máfias”); há o terrorismo dos grupos para militares que se opõem ao governo. Mas não é só: há o terrorismo que sequestra e o que explode, há o terrorismo civil e há o terrorismo de Estado.

Vasconcelos (2015) cita a definição criada pela GTD que entende como ação terrorista toda ameaça ou uso real da força praticada por pessoas ou grupos sem autorização do estado e que atuam para atingir um objetivo político, econômico, religioso ou social através do medo, coerção ou intimidação. No Brasil, algumas ações de traficantes podem receber essa classificação. O autor faz uma breve análise dos incidentes registrados na GTD onde os Entre os ataques mais comuns dos terroristas, considerando os dados desde 1970, destacando os ataques por explosão de bombas (48%), assaltos, assassinatos e sequestros. A análise realizada pelo autor ainda afirma que o ano de 2014 registrou o maior número de mortos em ações terroristas pelo mundo desde 1970.

Santos (2013) complementa afirmando que com as transformações dos últimos anos o conceito e os tipos de terrorismo também mudaram, razão pela qual tratar o terrorismo à luz do direito e das relações internacionais passou a ser algo cada vez mais difícil, isso por que em função dos variados motivos que levam aos ataques terroristas, os meios de financiamento e as partes envolvidas, criar um padrão conceitual e um consenso universal sobre o que pode ou não ser considerado terrorismo tornou-se um trabalho árduo para Organismos Internacionais, doutrinadores e demais estudiosos da área.

### 2.1.1 Terrorismo no Brasil

Lasmar (2015) entende que apesar da posição do Brasil como acolhedor de diversas culturas e dos fortes negativos do governo, o país se encontra vulnerável ao terrorismo internacional assim como qualquer outro. O mesmo autor ainda afirma que grande parte dos argumentos sobre a inexistência do problema do terrorismo internacional no Brasil se baseia no histórico recente da inexistência de ataques terroristas em território brasileiro.

Entretanto, grande parte do governo brasileiro adota uma postura de “negacionismo” que não apenas insistentemente nega publicamente a existência de qualquer atividade terrorista internacional no Brasil, mas também reage fortemente contra qualquer manifestação que relate contrário. Além disso, geralmente os crimes relacionados ao terrorismo não são tipificados e acabam sendo investigados sob a guisa de outros crimes ou infrações tanto penais quanto administrativas tais como falsificação de documentos, divulgação de propaganda racista, entrada ilegal no país dentre outros (LASMAR, 2015).

A fim de tipificar tais atos terroristas ocorridos no Brasil, entrou em vigor, em março deste ano a lei nº 13.260 que regulamentou o disposto no inciso XLIII do art. 5º, da Constituição Federal de 1988. Netto (2016) lembra que essa tipificação do crime de terrorismo, era uma obrigação que o Brasil havia assumido no concerto internacional das nações. E que estava colocando o país em uma situação bastante desconfortável, dada a demora na sua implementação.

Com isso, no art. 2º, desta nova legislação, foi traçada a tipificação dos crimes de terrorismo, nos seguintes termos:

Art. 2º O terrorismo consiste na prática por um ou mais indivíduos dos atos previstos neste artigo, por razões de xenofobia, discriminação ou preconceito de raça, cor, etnia e religião, quando cometidos com a finalidade de provocar terror social ou generalizado, expondo a perigo pessoa, patrimônio, a paz pública ou a incolumidade pública.

§ 1º São atos de terrorismo:

I - Usar ou ameaçar usar, transportar, guardar, portar ou trazer consigo explosivos, gases tóxicos, venenos, conteúdos biológicos, químicos, nucleares ou outros meios capazes de causar danos ou promover destruição em massa;

II – (VETADO);

III - (VETADO);

IV - Sabotar o funcionamento ou apoderar-se, com violência, grave ameaça a pessoa ou servindo-se de mecanismos cibernéticos, do controle total ou parcial, ainda que de modo temporário, de meio de comunicação ou de transporte, de portos, aeroportos, estações ferroviárias ou rodoviárias, hospitais, casas de saúde, escolas, estádios esportivos, instalações públicas ou locais onde funcionem serviços públicos essenciais, instalações de geração ou transmissão de energia, instalações militares, instalações de exploração, refino e processamento de petróleo e gás e instituições bancárias e sua rede de atendimento;

V - Atentar contra a vida ou a integridade física de pessoa: Pena - reclusão, de doze a trinta anos, além das sanções correspondentes à ameaça ou à violência. (BRASIL, 2016)

Netto (2016) questiona o veto do artigo III, que também classifica como ato terrorista, “*interferir, sabotar ou danificar sistemas de informática ou bancos de dados*”, afirmando que, apesar das brechas de tal redação, uma das ameaças terroristas mais temidas é exatamente a cibernética, promovida por meio (ou com o emprego) da informática, valendo-se das redes mundiais de computadores.

Alguns governantes se opuseram a nova lei afirmando que, os delitos previstos no projeto aprovado já estavam caracterizados no Código Penal brasileiro, como lesões corporais, danos a patrimônios ou tentativas de homicídio. Além disso, parte das organizações não governamentais (ONGs) que atua na área de direitos humanos reclama da aprovação do projeto e disse que ele criminaliza os movimentos sociais. Um grupo de 42 ONGs entende que a lei é inconstitucional porque fere a liberdade de manifestação, expressão e associação e vai contra o princípio da legalidade estrita, que diz que todas as leis penais devem ser de fácil compreensão a sociedade em geral (BENITES, 2016).

## 2.2 Análise Criminal

A Análise Criminal, conforme aponta Gottlieb (1998) citado por Feitosa e Carvalho (2013) é definida como:

Um conjunto de processos sistemáticos direcionados para o provimento de informação oportuna e pertinente sobre os padrões do crime e suas correlações de tendências, de modo a apoiar as áreas operacional e administrativa no planejamento e distribuição de recursos

para prevenção e supressão das atividades criminais (GOTTLIEB,1994 apud FEITOSA; CARVALHO,2013, p.9).

Já de acordo com Rodrigues (2012, p. 18 apud BOBA, 2009), análise criminal pode ser entendida como: [...] um estudo sistemático de problemas de crime e desordem, bem como outros relacionados à segurança pública, incluindo fatores sócio demográficos, espaciais e temporais, para apoiar os órgãos de segurança pública na prevenção e redução de crimes e desordem, na prisão de criminosos e na avaliação de resultados das medidas e ações de segurança pública.

Peixoto (2008, p.4) afirma que análise criminal é o "uso de uma coleção de métodos para planejar ações e políticas de segurança pública, obter dados, organizá-los, analisá-los, interpretá-los e deles tirar conclusões".

Magalhães (2008,*on-line*) entende que a análise criminal pode ser dividida em três grandes vertentes sendo elas:

Análise Criminal estratégica, que é voltada para o estudo dos fenômenos e suas influências em longo prazo. Esta possui como focos prioritários à formulação de políticas públicas; Produção de conhecimento para redução da criminalidade; Planejamento e desenvolvimento de soluções; Interação com outras secretarias na construção de ações de Segurança Pública; Direcionamento de investimentos; Formulação do plano orçamentário; Controle e acompanhamento de ações e projetos; Formulação de Indicadores de desempenho.

Análise Criminal Tática, que é voltada para o estudo dos fenômenos e suas influências em médio prazo. Esta vertente realiza a produção de conhecimento voltado para os operadores de segurança pública que no jargão policial são conhecidos como: "homens da linha de frente", tanto das polícias ostensivas, quanto das polícias investigativas. Possui como focos prioritários a produção de conhecimento para orientar as atividades de policiamento ostensivo nas atividades preventivas e repressivas. Na vertente das policiais investigativas, o mesmo conhecimento produzido, facilitará, subsidiará e indicará possíveis vertentes investigativas na busca da solução dos casos investigados no que concerne a busca da autoria e materialidade dos delitos.

Análise Criminal Administrativa, que é voltada para o público alvo funciona como um editor chefe que seleciona quais assuntos e profundidades desses assuntos serão divulgados para cada cliente. Os focos prioritários dessa vertente são fornecimento de informações sumarizadas para seus diversos públicos (Cidadãos, Gestores Públicos, Instituições Pública, Organismos Internacionais, Organizações Não-Governamentais, etc); Elaboração de Estatísticas (descritiva); Elaboração de informações gerais sobre tendências criminais; Comparações com períodos similares passados; Comparações com outras cidades similares (*benchmarking*). (MAGALHÃES, 2008, *ON-LINE*).

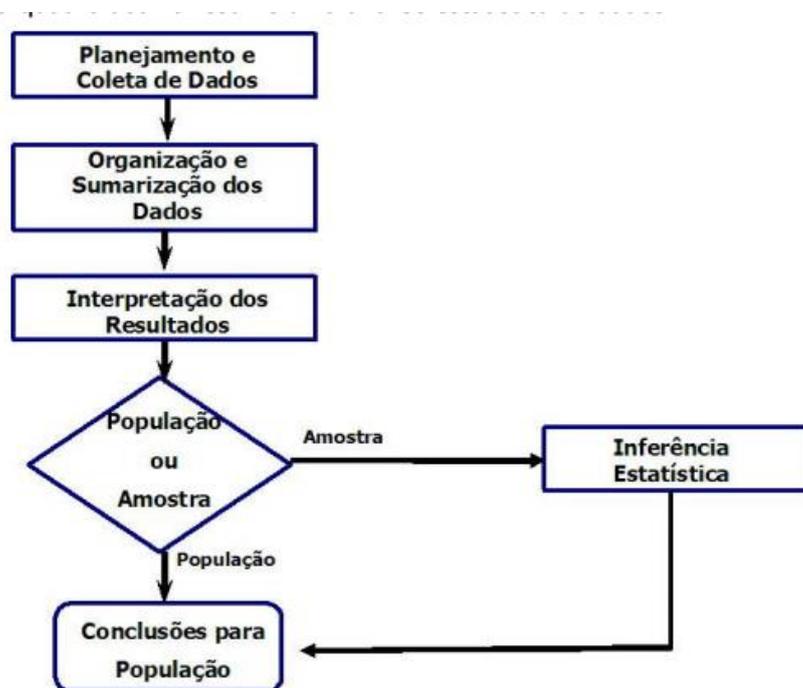
Considerando as definições apresentadas podemos concluir conforme Filipe (2007, p. 8, apud BOBA 2005) que as diversas definições, sobre análise criminal, utilizadas e aceitas nos meios acadêmicos e policiais, diferem nas particularidades umas das outras, mas que compartilham vários componentes comuns: todos os autores concordam que a análise criminal dá suporte às missões policiais, utiliza métodos sistemáticos e informações, e gera conhecimento para diversos usuários.

### 2.3 Análise estatística

A estatística, é considerada um ramo da matemática, que tem como principais objetivos obter, organizar e analisar dados, determinar as correlações entre eles, proporcionando conclusões e previsões, fornecendo métodos e técnicas para se lidar com situações sujeitas a incertezas. Neste sentido pode ser considerada como um conjunto de técnicas e métodos de pesquisa que entre outros tópicos envolve a planificação de experiências, a coleta e organização de dados, a inferência, o processamento, a análise e a disseminação de informação.

Ferreira (2009) entende que o papel da estatística pode ser semelhante ao da mineração de dados. Os dados devem ser cuidadosamente coletados (observados), devidamente conhecidos e utilizados para analisar e interpretar a sua variabilidade de forma a possibilitar uma correta resposta à hipótese em estudo. O mesmo autor afirma que a análise estatística, representada na figura 1, nada mais é do que o processo de organização, processamento, sumarização e retirada de conclusões sobre um determinado conjunto de dados (amostra).

Figura 1 Processo de análise estatística



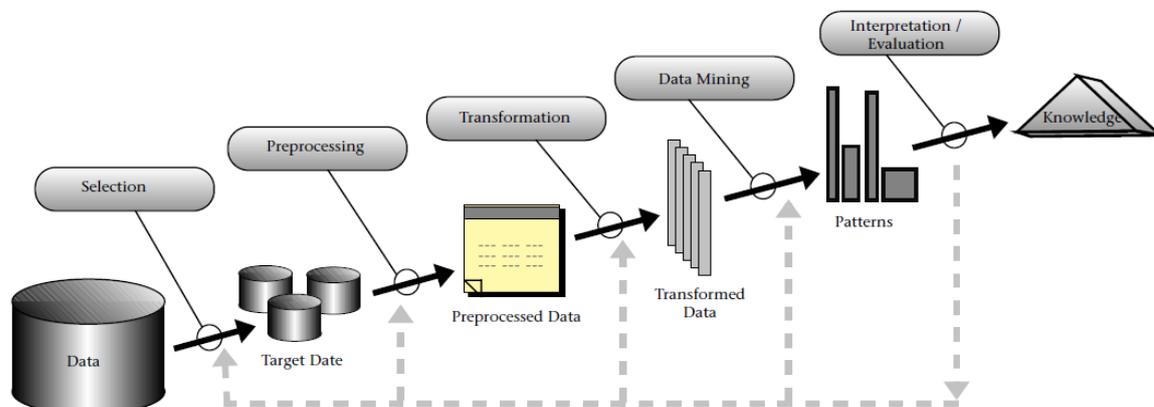
FONTE: FERREIRA (2009)

As hipóteses (questões de interesse) daqueles que realizam o estudo indicam o tipo de dado que precisa ser obtido e conseqüentemente a inferência a ser realizada. A inferência por sua vez é a metodologia utilizada para se fazer a passagem dos resultados obtidos na amostra para conclusões populacionais é chamada “inferência estatística” Ferreira (2009).

#### 2.4 KDD (KNOWLEDGE DISCOVERY IN DATABASE)

O KDD (*Knowledge Discovery In Database* – ou Descoberta de Conhecimento em Bases de Dados, em português) apresentado na figura 2, consiste no processo de descoberta de padrões pela análise de grandes conjuntos de dados, tendo como principal etapa o processo de mineração, consistindo na execução prática de análise e de algoritmos específicos que, sob limitações de eficiência computacionais aceitáveis, produz uma relação particular de padrões a partir de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Figura 2 - Processos de KDD



FONTE: Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 41)

Os autores entendem que o KDD é realizado em nove etapas:

- 1- Desenvolver uma compreensão do domínio da aplicação e do conhecimento prévio relevante e identificar o objetivo do processo de KDD do ponto de vista do cliente;
- 2- Criar um conjunto de dados de destino, ou seja, selecionar um conjunto de dados, ou simplesmente um subconjunto de variáveis ou amostras de dados, em que a “descoberta” será realizada;
- 3- Remover ruídos, coletar a informação necessária para modelagem, decidir estratégias para manusear campos perdidos;
- 4- Reduzir e projetar dados considerando as características úteis para representar os dados, dependendo do objetivo da tarefa e o uso de redução dimensionável e métodos de transformação para reduzir o número efetivo de variáveis que deve ser levado em consideração ou encontrar representações invariantes para os dados;
- 5- Escolher a tarefa de mineração de dados (classificação, regressão, regras de associação e agrupamento, dentre outras) considerando o objetivo definido na etapa 1;
- 6- Definir o algoritmo de mineração de dados, selecionando assim o(s) método(s) para encontrar padrões nos dados;

- 7- Pesquisar de padrões de interesse em uma forma representacional particular ou um conjunto de tais representações, incluindo regras de classificação ou árvores, regressão e agrupamento;
- 8- Interpretar o modelo descoberto, possivelmente retornando a algum passo anterior para maior interação;
- 9- Usar o conhecimento diretamente, incorporando-o em outro sistema de novas ações, ou simplesmente documentá-lo e apresentá-lo às partes interessadas. Este processo também inclui a verificação e resolução dos potenciais conflitos por meio deste novo conhecimento extraído.

Fayyad, Piatetsky-Shapiro e Smyth (1996, p.41) ainda afirmam que o processo de KDD é interativo e iterativo, envolvendo vários passos com muitas decisões tomadas pelo usuário.

## 2.5 MINERAÇÃO DE DADOS

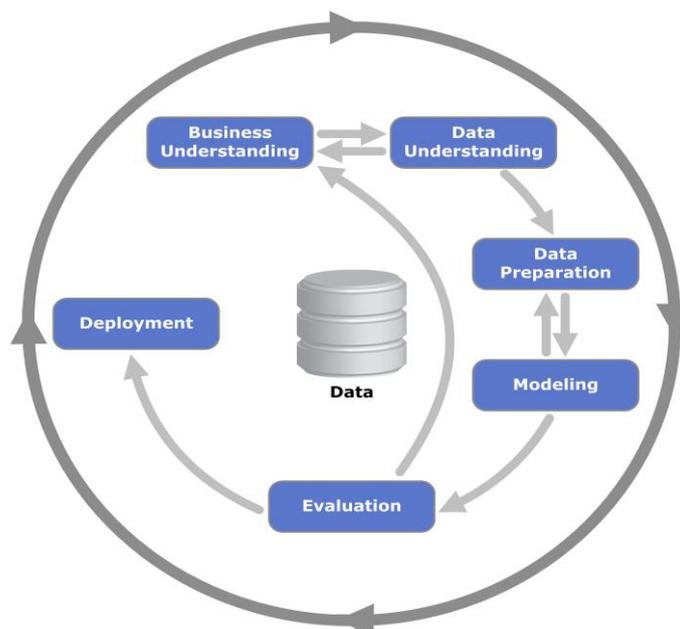
O termo mineração de dados (do inglês *data mining*) é entendido por Fayyad et al. (apud CAMILO e SILVA, 2009) como um passo no processo de descoberta de conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, produzem um conjunto de padrões de certos dados." Para Sholom (1999 apud CORTES, PORCARO e LIFSCHITZ, 2002).

Mineração de dados é a busca de informações valiosas em grandes bancos de dados. É um esforço de cooperação entre homens e computadores. Os homens projetam bancos de dados, descrevem problemas e definem seus objetivos. Os computadores verificam dados e procuram padrões que casem com as metas estabelecidas pelos homens. (SHOLOM, 1999 apud CORTES; PORCARO; LIFSCHITZ, 2002, p.1).

Apesar de ambas as definições sobre a mineração de dados suporem que tal processo se dá de uma forma totalmente automática, isso não é verdade, existem diversas ferramentas que nos auxiliam na execução dos algoritmos de mineração, no entanto os resultados ainda requerem de uma análise humana (CAMILO; SILVA, 2009).

Ainda de acordo com Camilo e Silva (2009) diversos processos definem e padronizam as fases e atividades da Mineração de Dados. Apesar das particularidades, todos em geral contêm a mesma estrutura, como exemplo o CRISP-DM representado na figura 3.

Figura 3 - Modelo CRISP-DM



FONTE: SFERRA; CORRÊA (2003)

O modelo CRISP-DM, segundo Sferra e Corrêa (2003), é composto por seis fases descritas a seguir:

1. entendimento do negócio (*Business Understanding*): visa o entendimento dos objetivos e requisitos do projeto, do ponto de vista do negócio. Baseado no conhecimento adquirido, o problema de mineração de dados é definido e um plano preliminar é projetado para alcançar os objetivos;
2. entendimento dos dados (*Data Understanding*): inicia com uma coleção de dados e prossegue com atividades que visam buscar familiaridade, identificar problemas de qualidade, descobrir os primeiros discernimentos nos dados ou detectar subconjuntos interessantes para formar hipóteses da informação escondida;
3. preparação dos dados (*Data Preparation*): cobre todas as atividades de construção do *dataset* final. As tarefas de preparação de dados

são, provavelmente, desempenhadas várias vezes e sem qualquer ordem prescrita. Essas tarefas incluem a seleção de tabelas, registros e atributos, bem como a transformação e limpeza dos dados para as ferramentas de modelagem;

4. modelagem (*Modelling*): várias técnicas de modelagem são selecionadas e aplicadas, e seus parâmetros são ajustados para valores ótimos. Geralmente, existem várias técnicas para o mesmo tipo de problema de mineração de dados. Algumas delas têm requisitos específicos na formação de dados. Portanto, retornar à fase de preparação de dados é frequentemente necessário;
5. avaliação (*Evaluation*): o modelo (ou modelos) construído na fase anterior é avaliado e os passos executados na sua construção são revistos, para se certificar que o modelo representa os objetivos do negócio. Seu principal objetivo é determinar se existe alguma questão de negócio importante que não foi suficientemente considerada. Nesta fase, uma decisão sobre o uso dos resultados de mineração de dados deverá ser obtida;
6. utilização ou aplicação (*Deployment*): após a construção e avaliação do modelo (ou modelos), ele pode ser utilizado de duas formas: em uma, o analista pode recomendar ações a serem tomadas baseando-se, simplesmente, na visão do modelo e de seus resultados; na outra, o modelo pode ser aplicado a diferentes conjuntos de dados. (SFERRA; CORRÊA, 2003).

Witten et al.(2005), Olson et al (2008) e Bramer (2007) citados por Camilo e Silva (2009) apresentam algumas das áreas nas quais a mineração de dados é aplicada de forma satisfatória sendo elas: retenção de clientes, na identificação de perfis para determinados produtos, venda cruzada; em bancos na identificação de padrões para auxiliar no gerenciamento de relacionamento com o cliente; por empresas de cartões de crédito na identificação de segmentos de mercado e padrões de rotatividade; cobrança na detecção de fraudes; eleitoral na identificação do perfil de possíveis votantes; medicina na indicação de diagnósticos mais precisos e segurança na detecção de atividades terroristas e criminais. Neste sentido a mineração de dados pode ser utilizada em diversos

domínios, logo, suas tarefas vem se diversificando cada vez mais. (CASTANHEIRA, 2008).

Carvalho (2005) destaca três principais tarefas em mineração de dados sendo elas: classificação, regras de associação e *clustering*.

Cortes, Porcaro e Lifschitz (2002) entendem que a classificação consiste em examinar uma característica nos dados e atribuir uma classe previamente definida. Os dados podem ser associados a classes ou a conceitos através de um processo de discriminação ou de caracterização. Na discriminação os resultados são obtidos através da atribuição de um valor a um atributo no registro, em função de um ou mais atributos do mesmo. Por exemplo, em um supermercado podemos classificar os produtos por tipo como alimentício, vestuário, higiene e limpeza etc. Já a caracterização é a classificação de um atributo de estudo por uma característica de um ou mais atributos. Por exemplo, podemos caracterizar um empregado pelo seu salário anual, identificando faixas da agregação mensal de seus salários em baixa, média e alta.

Goldschmidt e Bezerra (2016) exemplificam a tarefa de classificação citando uma financeira que possui o histórico de seus clientes e o comportamento destes em relação ao pagamento de empréstimos contraídos previamente, tais clientes são classificados em adimplentes e inadimplentes, nesta situação, uma tarefa de classificação consiste em descobrir uma função que mapeie corretamente os clientes, a partir de seus dados, tal função pode ser utilizada para prever o comportamento de novos clientes que desejem contrair empréstimos junto à financeira.

A tarefa de associação por sua vez, tem o intuito de identificar associações entre registros de dados que, de alguma maneira, estão ou devem estar relacionados. (VASCONCELOS E CARVALHO, 2004)

Para Baptista e Carvalho (2003) apud Moreira (2004) regras de associações procuram identificar afinidades entre registros de um subconjunto de dados. Essas afinidades/associações são expressas na forma de regras.

Os mesmos autores exemplificam, as tarefas de associação a partir de uma base de dados, na qual registram-se os itens adquiridos por clientes.

Uma estratégia de mineração, com o uso de regras de associação, poderia gerar a seguinte regra: {cinto, bolsa}  $\rightarrow$  {sapato}, a qual indica que o cliente que compra cinto e bolsa, com um determinado grau de certeza, compra também sapato. Este grau de certeza de uma regra é definido por dois índices: o fator de suporte e o fator de confiança. (VASCONCELOS; CARVALHO,2014, p.8).

Neste sentido o problema de mineração por regras de associação está em gerar todas as regras que contenham o suporte e confiança iguais ou maiores do que os valores mínimos determinados pelo usuário, referenciados como suporte mínimo e confiança mínima, respectivamente; onde o suporte de uma regra  $X \Rightarrow Y$ , em que X e Y são conjuntos de itens, é dado pela frequência de X e Y dividido pelo total de transações e a confiança é calculada através da frequência de X e Y dividida pelo número de transações de X. (VASCONCELOS; CARVALHO,2014).

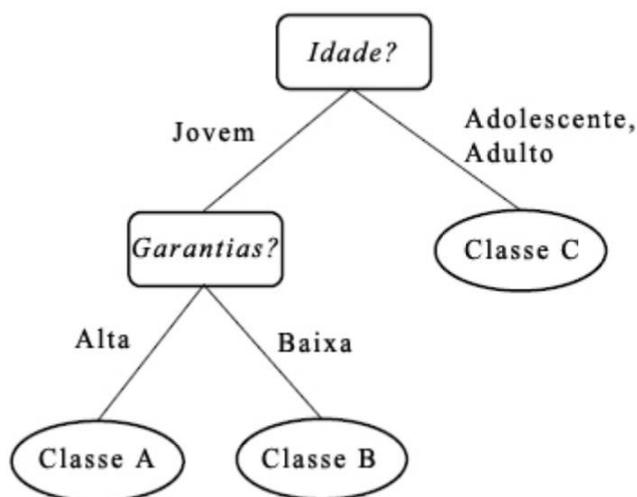
Goldschmidt e Bezerra (2016) exemplificam citando que por meio de análise das transações de compra, os analistas de dados puderam perceber que os compradores eram homens que, ao comprarem fraldas para seus filhos, compravam também cerveja para consumo enquanto cuidavam das crianças e assistiam aos jogos na televisão durante o final de semana. De acordo com os autores, essa empresa utilizou o novo conhecimento para aproximar as gôndolas de fraldas e cervejas na rede de mercados, incrementando assim a venda conjunta dos dois produtos.

A tarefa de *clustering* ou agrupamento visa identificar e aproximar os registros similares. Um *cluster* é uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. Esta tarefa difere da classificação, pois não necessita que os registros sejam previamente categorizados (aprendizado não-supervisionado). Além disso, ela não tem a pretensão de classificar, estimar ou prever o valor de uma variável. Ela apenas identifica os grupos de dados similares (CAMILO; SILVA,2009).

Além destas, ainda merece destaque neste estudo a tarefa de regressão. Está de acordo com Goldschmidt e Bezerra (2016) é similar à tarefa de classificação, com a diferença de que o atributo alvo assume valores numéricos. Uysal e Gúvenir (1999 apud DOSUALDO, 2003) afirmam que as pesquisas relacionadas à mineração de dados têm se voltado também para problemas de

regressão (quando o atributo meta é do tipo contínuo), já que muitos problemas são desse tipo, como problemas relacionados com habilidades esportivas e controle dinâmico de robôs. Existem vários métodos que buscam resolver as tarefas citadas acima, em classificação, Halmenschlager (2003) cita o método de árvore de decisão (figura 4) como o método mais utilizado. O método funciona como um fluxograma em forma de árvore, onde cada nó indica um teste feito sobre um valor. As ligações entre os nós representam os valores possíveis do teste do nó superior, e as folhas indicam a classe (categoria) a qual o registro pertence (CAMILO; SILVA,2009). Após a árvore de decisão montada, para classificar um novo registro, basta seguir o fluxo na árvore começando no nó raiz até chegar a uma folha.

Figura 4 - Árvore de decisão



FONTE: Camilo e Silva (2009)

Adriaans (1997 apud HALMENSCHLAGER, 2003) cita como principal vantagem deste método sua estrutura simples e de grande legibilidade, expressando assim seus resultados de uma forma muito clara, sendo facilmente entendível pelo usuário.

Existem vários algoritmos para geração de árvores de decisão, o pioneiro foi o ID3 desenvolvido por Quilan (1986) que é um algoritmo recursivo e baseado em busca gulosa, procurando, sobre um conjunto de atributos, aqueles que

melhor dividem os exemplos, gerando sub árvores. Posteriormente o ID3 foi aperfeiçoado para o algoritmo C4.5. (BASGALUPP, 2010).

Uma alternativa ao método de árvore de decisão são os algoritmos de cobertura que buscam uma maneira de cobrir todos os objetos de todas as classes e ao mesmo tempo, excluem todos os objetos fora da classe, resultando em um conjunto de regras. (CASTRO; FERRARI, 2016). Neste caso o pioneiro é o algoritmo PRISM proposto por Cendrowska em 1987, onde para cada regra gerada é escolhida a melhor condição, de acordo com a heurística desejada, criando um subconjunto de treinamento que englobe esta condição.

Tratando-se dos problemas de regras de associação Agrawal e Srikant propõe o algoritmo Apriori, desenvolvido com o objetivo de descobrir associações de produtos em grandes bases de dados transacionais obtidas em supermercados. O algoritmo gera conjuntos de itens candidatos de  $k$  elementos a partir de conjuntos de itens com  $k - 1$  elementos, onde os itens candidatos não frequentes são eliminados, e toda a base de dados é rastreada e os conjuntos de itens frequentes, obtidos a partir dos conjuntos de itens candidatos. (CASTRO; FERRARI, 2016).

As tarefas de agrupamento podem ser resolvidas a partir de algoritmos com o *k-means* onde dado um conjunto de dados  $X$ , o algoritmo seleciona de forma aleatória  $k$  registros, cada um representando um agrupamento. Para cada registro restante, é calculada a similaridade entre o registro analisado e o centro de cada agrupamento. O objeto é inserido no agrupamento com a menor distância, ou seja, maior similaridade. O centro do *cluster* é recalculado a cada novo elemento inserido (CAMILO; SILVA, 2009).

Ainda merecem os métodos que utilizam redes neurais, que são inspiradas na arquitetura e no funcionamento do cérebro humano. Entre suas principais características sob uma perspectiva de processamento de informação destacam-se seu processamento paralelo e distribuído e sua capacidade de generalização. (CASTRO; FERRARI, 2016). Neste caso o pioneiro é o *perceptron* criado Rosenblatt em 1958. O algoritmo consiste em uma rede neural simples constituída de uma camada de entrada e uma camada de saída. A cada entrada existe um peso relacionado, sendo que o valor de saída será a soma

dos produtos de cada entrada pelo seu respectivo peso.(CARDON; MÜLLER, 1994).

Por fim, vista a relação entre a análise estatística e tarefas de mineração de dados a próxima seção preocupa-se em apresentar os procedimentos metodológicos utilizados neste estudo, visando o alcance dos objetivos propostos.

### 3 ENCAMINHAMENTOS METODOLÓGICOS

Esta seção apresenta os encaminhamentos metodológicos utilizados neste trabalho, abordando assim, a caracterização da pesquisa, procedimentos e suas fases.

#### 3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa é caracterizada quanto a acordo com a sua finalidade, abordagem, objetivos e procedimentos.

##### 3.1.1 Tipo de pesquisa

Gil (2008) entende que uma pesquisa científica pode ser classificada de acordo com à sua finalidade, abordagem, objetivos e procedimentos.

Quanto a finalidade da pesquisa Gil (2008) afirma que a mesma pode ser classificada em pura ou aplicada. A pesquisa pura busca o avanço da ciência e procura desenvolver os conhecimentos científicos sem a preocupação direta com suas aplicações e consequências práticas. A pesquisa aplicada tem como característica fundamental o interesse na aplicação, utilização e consequências práticas dos conhecimentos gerados. Com isso este estudo pode ser classificado como uma pesquisa aplicada.

Quanto a abordagem a pesquisa pode ser caracterizada como qualitativa ou quantitativa. A pesquisa qualitativa trabalha com o universo de significados, motivos, aspirações, crenças, valores e atitudes, o que corresponde a um espaço mais profundo das relações, dos processos e nos fenômenos que não podem ser reduzidos à operacionalização de variáveis (MINAYO, 2002). Na pesquisa quantitativa os dados podem ser quantificados. Como as amostras geralmente são grandes e consideradas representativas da população, os resultados são tomados como se constituíssem um retrato real de toda a população alvo da pesquisa. Com isso considera que a realidade só pode ser compreendida com base na análise de dados brutos, recolhidos com o auxílio de instrumentos padronizados e neutros. Neste sentido, esta pesquisa pode ser classificada como quantitativa.

Quanto aos objetivos Gil (2002) afirma que pesquisa científica como exploratória, explicativa e descritiva. As pesquisas exploratórias têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipótese têm como objetivo principal o aprimoramento de ideias ou a descoberta de intuições. As pesquisas descritivas têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis. As pesquisas explicativas procuram identificar os fatores que determinam ou que contribuem para a ocorrência dos fenômenos. Neste sentido, esta pesquisa pode ser entendida como exploratória.

Por fim quanto aos procedimentos Gil (2008), classifica a pesquisa como: pesquisa bibliográfica; documental; participante; experimental; estudo de campo, e estudo de caso. Neste sentido, esta pesquisa é classificada como um estudo de caso, visto que esta abordagem “investiga um fenômeno atual dentro do seu contexto de realidade, quando as fronteiras entre o fenômeno e o contexto não são claramente definidas” (GIL, 2008).

## 3.2 MATERIAIS E MÉTODOS

Esta seção descreve a base de dados utilizada, bem como os *software* que foram necessários para realização da análise criminal proposta.

### 3.2.1 GTD

A *Global Terrorism Database* (GTD) é um banco de dados de código aberto que inclui informações sobre eventos terroristas em todo o mundo de 1970 a 2015. É uma iniciativa da universidade de *Maryland*.

O conjunto original de incidentes que compõem a GTD ocorreu entre 1970 e 1997 e foram coletados pelo *Pinkerton Global Intelligence Service* (PGIS), uma agência de segurança privada. O Consórcio Nacional para o Estudo do Terrorismo e Respostas ao Terrorismo (START), que atualmente é responsável pela manutenção da GTD realizou a digitalização desses registros que até então estavam manuscritos. É importante ressaltar que os incidentes de terrorismo ocorridos em 1993 foram perdidos antes da compilação da GTD, estima-se que

apenas 15% dos ataques foram recuperados, o que acarretaria em análises inconsistentes, por isso, os dados relacionados a este ano não estão disponíveis na base.

Para continuar a coleta de dados além de 1997, foi realizada uma parceria com o Centro de Estudos de Terrorismo e Inteligência (CETIS) que realizou a coleta dos dados referentes aos ataques terroristas ocorridos entre 1º janeiro de 1998 à 3 de março de 2008. Entre 4 de março de 2008 a 31 de outubro de 2011 o Instituto de Estudos de Grupos Violentos (ISVG) passou a ser o responsável pela coleta de dados. Após esta data o próprio START é que recolhe os dados disponibilizados.

A GTD considera um ataque terrorista o uso ameaçado ou real de força ilegal e violência por um ator não-estatal para atingir um objetivo político, econômico, religioso ou social através do medo, coerção ou intimidação. Neste sentido consideram-se três critérios para que um ataque seja incluso na GTD sendo eles:

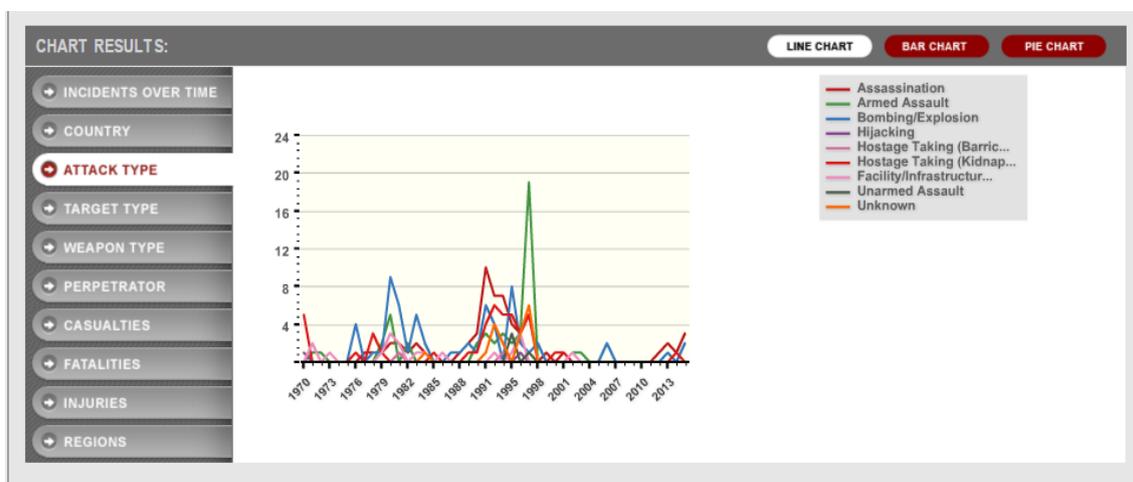
- a) o incidente deve ser intencional;
- b) o incidente deve implicar algum nível de violência ou ameaça imediata de violência (incluindo a violência contra a propriedade e contra as pessoas);
- c) os autores dos incidentes devem ser atores subnacionais (o banco de dados não inclui atos de terrorismo de estado).

Além destes, existem outros três critérios adicionais dos quais no mínimo dois devem ser atendidos para a inclusão de um registro na GTD.

- a) o ato deve ser destinado a atingir um objetivo político, econômico, religioso ou social (a procura exclusiva de lucro não satisfaz este critério);
- b) deve haver evidência da intenção de coagir, intimidar ou transmitir alguma mensagem a um público maior do que as vítimas imediatas;
- c) a ação deve estar fora do contexto de atividades de guerra.

Para cada registro, estão disponíveis a data, local do incidente, armas utilizadas, natureza do alvo, o número de vítimas e quando identificado o grupo ou indivíduo responsável pelos ataques registrados. Estes registros são representados em forma de gráficos conforme a figura 5.

Figura 5 - Resultados de uma busca na GTD



FONTE: GTD

As informações estatísticas contidas na GTD baseiam-se em relatórios de uma variedade de fontes de mídia abertas, logo, é importante ressaltar que novos registros não são adicionados a base sem que seja verificada a autenticidade da informação coletada. Visando ainda a credibilidade dos registros disponíveis os usuários não são autorizados a alterar qualquer dado obtido nas pesquisas realizadas na GTD.

Para análise estatística realizada neste estudo foram utilizados os 156.772 registros presentes na base. Para as tarefas as tarefas de mineração de dados foram utilizadas duas amostras, a primeira contendo os 18.628 registros referentes aos ataques registrados na América do Norte e a segunda contendo os 267 registros referentes aos incidentes registrados no Brasil.

### 3.3 FERRAMENTAS

Esta seção apresenta as ferramentas utilizadas para realizar as tarefas de análise estatística e mineração de dados.

#### 3.3.1 SPSS®

O *SPSS® Statistics Base* desenvolvido pela IBM é um software de análise estatística que fornece os principais recursos necessários para executar um processo de análise do início ao fim. Nesta pesquisa foi utilizado na realização na etapa de análise estatística.

### 3.3.2 Excel

O *Microsoft Office Excel* ® produzido pela *Microsoft* é um editor de planilhas utilizado para realização de cálculos e elaboração de gráficos, foi utilizado na preparação dos dados a serem utilizados no *Weka*. Importante ressaltar que os resultados encontrados nas pesquisas realizadas na GTD, estão disponíveis para *download* no formato .xls.

### 3.3.3 WEKA

O WEKA desenvolvido pela universidade de Waikato é um *software* de código aberto que consiste em uma coleção de algoritmos de aprendizagem de máquina para tarefas de mineração de dados. É utilizado principalmente em grandes bases de dados, onde torna-se inviável realizar uma análise manual. Nesta pesquisa, será utilizado na realização das tarefas de mineração de dados.

Torna-se necessário esclarecer alguns conceitos básicos que são fornecidos como resultado das tarefas de mineração de dados no Weka estes conceitos são descritos por Martinez, Casal e Janeiro (2009), sendo eles:

**Kappa Statistic:** Índice que compara o valor encontrado nas observações com aquele que se pode esperar do acaso. É o valor calculado dos resultados encontrados nas observações e relatado como um decimal (0 a 1). Quanto menor o valor de Kappa, menor a confiança de observação, o valor 1 implica a correlação perfeita.

**Mean Absolute Error:** média da diferença entre os valores atuais e os preditos em todos os casos, é a média do erro da predição.

**True Positives (TP):** são os valores classificados verdadeiramente positivos.

**False Positives (FP):** são os falsos positivos, são os dados classificados erroneamente como positivos pelo classificador.

**Precision (Precisão):** É o valor da predição positiva (número de casos positivos por total de casos cobertos), muito influenciada pela especificidade e pouco pela sensibilidade. Sensibilidade é o número de casos positivos que são verdadeiramente positivos e especificidade é o número de casos negativos que são verdadeiramente negativos.

**Recall (Cobertura):** É o valor da cobertura de casos muito influenciada pela sensibilidade e pouco pela especificidade. É calculada por número de casos cobertos pelo número total de casos aplicáveis.

**F-measure:** Usada para medir o desempenho, pois combina valores de cobertura e precisão de uma regra numa única fórmula  $[2 * Prec * Rec / (Prec + Rec)]$ .

**Root Relative Squared Error:** reduz o quadrado do erro relativo na mesma dimensão da quantidade sendo predita incluindo raiz quadrada. Assim como a raiz quadrada do erro significativo (root mean-squared error), este exagera nos casos em que o erro da predição foi significativamente maior do que o erro significativo.

Relative Absolute Error: É o erro total absoluto. Em todas as mensurações de erro, valores mais baixos significam maior precisão do modelo, com o valor próximo de zero temos o modelo estatisticamente perfeito.

Root Mean-Squared Error: usado para medir o sucesso de uma predição numérica. Este valor é calculado pela média da raiz quadrada da diferença entre o valor calculado e o valor correto. O root mean-squared error é simplesmente a raiz quadrada do mean-squared-error (dá o valor do erro entre os valores atuais e os valores preditos). (MARTINEZ, CASAL e JANEIRO,2009, p .4,5)

Tais conceitos serão utilizados para análise das tarefas de mineração de dados apresentadas na seção de resultados.

### 3.3.4 Power BI ®

É um conjunto de ferramentas para análise de dados de negócios. Como solução de *Business Intelligence* para o meio corporativo. Integrado ao *Excel* foi utilizado para elaboração dos gráficos de forma dinâmica e intuitiva.

## 3.4 PREPARAÇÃO DOS DADOS

Os dados utilizados neste estudo foram coletados diretamente da GTD em 13 de março de 2017 no formato.exe. Os 156.772 registros disponíveis atualmente foram categorizados em 12 regiões distintas (tabela 7) sugeridas pela própria GTD afim de otimizar as análises bem como observar possíveis semelhanças e diferenças entre elas.

Tabela 7 - Registros de ataques categorizados por região

(Continua)

Região	Registros
África Subsaariana	13.434
América Central e Caribe	10.337
América do Norte	3.628
<b>América do Sul</b>	<b>18.628</b>
Ásia Central	538
Ásia Meridional	10.360
Ásia Oriental	786
Austrália e Oceania	246
Europa Ocidental	16.020

Europa Oriental	4.892	(Conclusão)
Oriente médio e Norte da África	40.422	
Sudeste Asiático	10.631	
Total	156.772	

FONTE: Elaborado pela autora (2017)

Como a GTD dispõe de um grande número de atributos (Apêndice 1), para este estudo considerou-se apenas os atributos entendidos como principais pela equipe da CTD devido a relação existente entre os mesmos (quadro 1).

Quadro 1 - Atributos da GTD considerados neste estudo

(Continua)		
Atributo	Descrição	Tipo
Ano	Ano em que ocorreu o ataque	Numérico
País	Nome do país em que ocorreu o ataque	Nominal
Cidade	Nome da cidade em que ocorreu o ataque	Nominal
Região	África Subsaariana, América Central e Caribe, América do Norte, América do Sul, Ásia Central, Ásia Meridional, Ásia Oriental, Austrália e Oceania, Europa Ocidental, Europa Oriental, Oriente médio e Norte da África, Sudeste Asiático	Nominal
Grupo Responsável	Nome do grupo que reivindicou a responsabilidade do ataque	Nominal
Feridos	Número de feridos ocorridos no ataque	Numérico
Vítimas fatais	Número de óbitos ocorridos no ataque	Numérico
Tipo de ataque	Assalto_Armado, Assalto_Desarmado, Assassinato, Bombardeio_Explosao, Instalacao_Infraestrutura, Sequestro, Tomada_Refem_(Sequestro), Tomada_refem_Incidente_de_Barricada	Nominal
Sucesso	0 = o alvo pretendido não foi atingido 1 = o alvo pretendido foi atingido	Categórico
Alvo	Aborto_Relacionado, Aeroportos_Aeronaves, Alimentos_Abastecimento_Agua,	Nominal

(Conclusão)

	Cidadãos_Privados_Propriedades, Instituicoes_educacionais, Empresas, Figuras_Religiosas_Instituicoes, Governo_Geral, Governo_Diplomatico, Maritimo, Militar, ONG, Partidos_Politicos_Violentos, Policia, Telecommunicacoes, Terrorismo_Milicia_nao_estatal, Transporte, Turistas, Utilidades	
Armas utilizadas	Armas_Falsas, Armas_Fogo, Biologica, Corpo_Corpo, Explosivos_Bombas_Dinamite, Incendiaria, Quimica, Sabotagem_Equipamento, eiculos_(exceto_carros e caminhos_bomba)	Nominal

FONTE: Elaborado pela autora (2017)

Para a realização das tarefas de mineração de dados realizada com os incidentes relacionados a América do Sul e ao Brasil foi realizada a limpeza dos dados eliminando os caracteres especiais (&, ', "", @ e /) e espaços entre os atributos do tipo nominal (país, cidade, alvo, tipo de ataque, grupo responsável e tipo de arma). Os registros em branco foram substituídos por -1 no caso de atributos numéricos (número de vítimas e número de feridos) e por desconhecido no caso dos atributos nominais. Após o arquivo foi salvo no formato .CSV e em seguida convertido para ARFF sendo este o formato suportado pelo WEKA.

Além disso, como os algoritmos relacionados as tarefas de associação não podem ser executados através de atributos numéricos, os atributos relacionados ao número de vítimas fatais e feridos foram discretizados conforme os limites apresentados no quadro 2.

Quadro 2 - Discretização dos atributos numéricos

A	Não houve vítimas / feridos
B	$1 \leq 50$
C	$>50$
D	Número de vítimas desconhecido

FONTE: Elaborado pela autora (2017)

A próxima seção apresenta os resultados obtidos através da análise estatística e tarefas de mineração de dados.

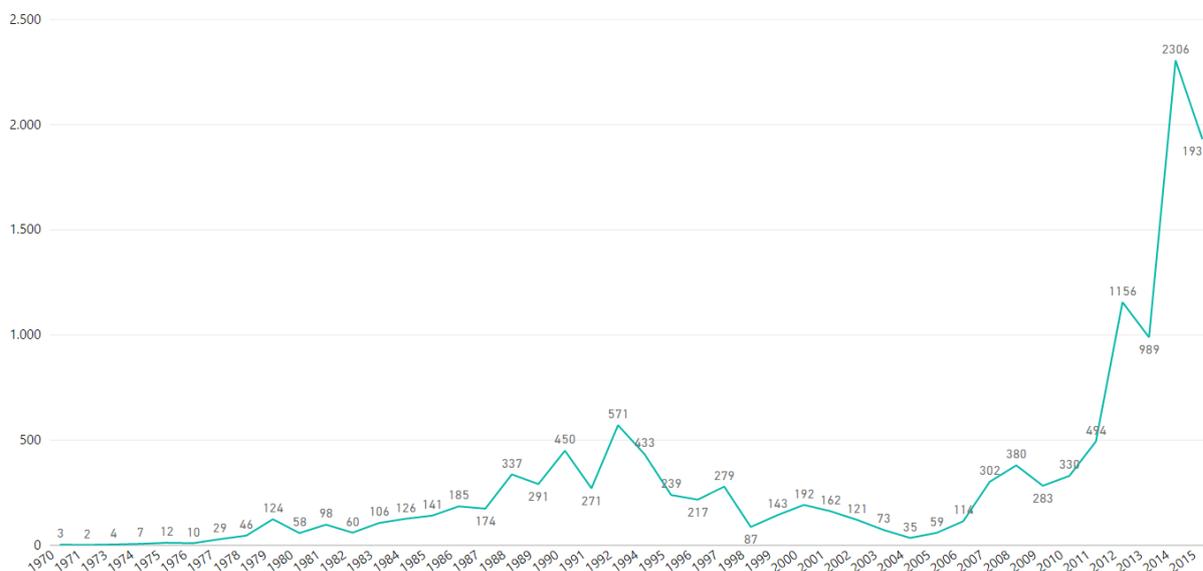
## 4 RESULTADOS E DISCUSSÕES

Apresentam-se os resultados obtidos através da análise estatística bem como tarefas de mineração de dados.

### 4.1 ANÁLISE ESTATÍSTICA

A África Subsariana apresenta 13.434 ataques registrados, ocupando assim o 5º lugar dentre as 12 regiões analisadas. O gráfico 1 mostra que na década de 90 houve oscilações em relação ao número de incidentes registrados, nota-se também que a partir de 2011 há um crescimento elevado de registros destaca-se o ano de 2014 com 2.306 registros.

Gráfico 1 - Ataques registrados na GTD ao longo dos anos (África Subsariana)



FONTE: Elaborado pela autora (2017)

A distribuição de ataques por país pode ser observada na tabela 8, onde a Somália e a Nigéria apresentam o maior número de ataques registrados, totalizando 42% do total de ataques ocorridos nesta região; 33 países apresentam entre 2 e 120 ataques, ou seja, menos de 1% do total registrado, estes estão apresentados no apêndice 2.

Tabela 8 - Distribuição de ataques por país (África Subsariana)

País	Frequência	Porcentagem
Somália	2.890	21,5%
Nigéria	2.888	21,5%
África do sul	1.957	14,6%
Sudão	688	5,1%
Quênia	517	3,8%
Burundi	504	3,8%
Angola	491	3,7%
República democrática do congo	459	3,4%
Uganda	366	2,7%
Mali	319	2,4%
Moçambique	272	2,0%
República centro-africana	202	1,5%
Camarões	180	1,3%
Ruanda	156	1,2%
Namíbia	151	1,1%
TOTAL	12.040	90%

FONTE: Elaborado pela autora (2017)

Em relação aos responsáveis pelos ataques em 4.851 (36%) registros nenhum grupo se responsabilizou pelo ataque, desconsiderando estes, o *Al-Shabaab* é o responsável pelo maior número de ataques sendo identificado em 2.127 (15,8%) registros, seguido pelo *Boko Haram* responsável por 1.839 (13,7%) ataques.

A média de vítimas fatais foi de 5,25 por ataque sendo que o maior número registrado foi 1.180 na cidade de Gikoro em Ruanda no ano 1994, o grupo responsável pelo ataque foi o *Hutus*. É importante ressaltar que em 4.724 (35%) ataques não houve vítimas fatais. A média de feridos foi de 3,89 por ataque, o maior número registrado foi de 4.000 no ano de 1998 na cidade de Nairobi no Quênia, o grupo responsável pelo ataque foi o *Al-Qaida*; em 6.915 (51,4%) ataques não houve feridos.

Com 10.337 registros a América Central e Caribe ocupa o 7º lugar dentre as regiões analisadas. O gráfico 2 mostra que houve um aumento significativo de ataques entre os anos de 1979 e 1981 e novamente entre os anos de 1989 e 1991.

Gráfico 2 - Ataques registrados na GTD ao longo dos anos (América Central e Caribe)



FONTE: Elaborado pela autora (2017)

A distribuição destes registros por país pode ser visualizada na tabela 9 onde é possível observar que 51,5 % ocorreram em um único país (El Salvador). Nota-se também que em 14 dos 22 países analisados, foram registrados entre 89 e 1 ataque representando menos de 1% do total de ataques identificados.

Tabela 9 - Distribuição de ataques por país (América Central e Caribe)

(Continua)

País	Frequência	Porcentagem
El salvador	5.320	51,5%
Guatemala	2.050	19,8%
Nicaragua	1.970	19,1%
Honduras	321	3,1%
Haiti	212	2,1%
Panama	126	1,2%
República dominicana	89	0,9%
Costa rica	67	0,6%
Guadalupe	56	0,5%
Jamaica	34	0,3%
Cuba	30	0,3%
Trindade e Tobago	22	0,2%
Martinica	12	0,1%
Belize	8	0,1%
Bahamas	5	0,0%

(Conclusão)

Granada	5	0,0%
Barbados	3	0,0%
Dominica	3	0,0%
São cristóvão e nevis	2	0,0%
Antígua e barbuda	1	0,0%
Santa lúcia	1	0,0%
<b>Total</b>	<b>10.337</b>	<b>100,0%</b>

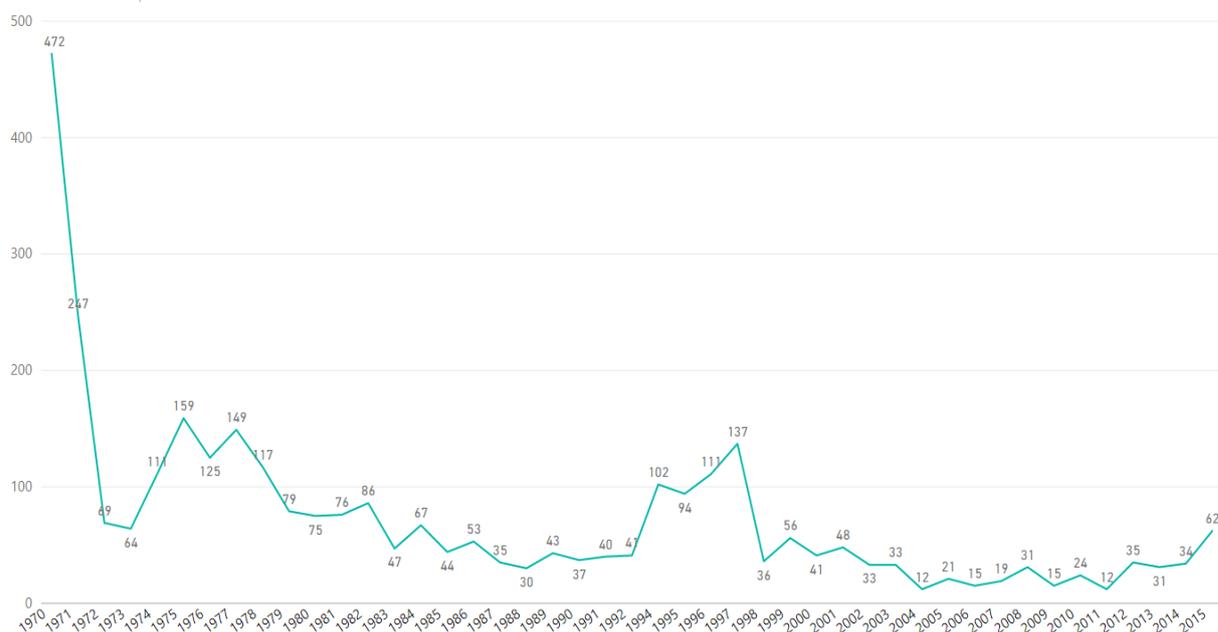
FONTE: Elaborado pela autora (2017)

Em 3.673 (35,5%) ataques não foi possível identificar o grupo responsável pelo ataque, o *Farabundo Marti National Liberation Front* (FMLN) foi responsável por 3.351 (32%) ataques.

A média de vítimas fatais foi de 3,58 por ataque, sendo que o maior número foi de 300 mortos em El Salvador na cidade de Suchitoto no ano de 1980, o grupo responsável pelo ataque é desconhecido. Em 4.268 ataques (41,3%) não houve mortes. A média de feridos foi de 4,43 por ataque, o maior número foi de 130 feridos também em El Salvador na cidade de El Paraiso o grupo responsável pelo ataque foi o *Farabundo Marti National Liberation Front* (FMLN), em 6.637 (61,3%) casos não houve feridos registrados.

A América do Norte ocupa o 9º lugar dentre as regiões analisadas com 3.268 ataques, a distribuição de ataques por país pode ser observada no gráfico 3, nota-se que houve uma grande concentração de ataques com 472 registros no ano de 1970.

Gráfico 3 - Ataques registrados na GTD ao longo dos anos (América do Norte)



FONTE: Elaborado pela autora (2017)

A distribuição dos ataques por país apresentada na Tabela 10 mostra que 2.693 (82,6%) ataques ocorreram nos Estados Unidos.

Tabela 10 - Distribuição de ataques por país (América do Norte)

País	Frequência	Porcentagem
Estados unidos	2.693	82,40%
México	499	15,30%
Canadá	76	2,30%
Total	3.268	100,00%

FONTE: Elaborado pela autora (2017)

Em relação aos grupos responsáveis pelos ataques em 792 casos (24,2%) não foi possível identificar o responsável pelo ataque, em outros 227(6,9%), casos os responsáveis foram pessoas não afiliadas a nenhum grupo terrorista, desta forma, o grupo responsável pelo maior número de ataques foi o *Left-Wing Militants*, identificado em 169 (5,2%) casos.

A média de vítimas fatais foi de 1,47 por ataque sendo que o maior número foi de 1.381 em 2001 nos Estados Unidos, na cidade de *Nova York*, o grupo responsável pelo ataque foi o grupo *Al-Qaida*. Em 103 (3,15%) casos não foi possível identificar o número de vítimas do ataque e em (82%) não houve mortes registradas. A média de feridos foi de 1,25 por ataque, sendo que o maior

número foi de 751 em 1984 também nos Estados Unidos, na cidade de *The Dalles*. Em 125 (3,85) casos não foi identificado o número de feridos em 2.682(82%) registros não houve feridos.

Tratando-se da América do Sul que ocupa a 3ª posição no *ranking* das regiões apresentadas com 18.628 ataques registrados, nota-se que a maior parte dos ataques ocorreu entre 1983 e 1992 (gráfico 4).

Gráfico 4 - Ataques registrados na GTD ao longo dos anos (América do Sul)



FONTE: Elaborado pela autora (2017)

A distribuição dos ataques por país pode ser observada na tabela 11 onde destaca-se a Colômbia com 8.077 ataques (43,4%).

Tabela 11 - Distribuição de ataques por país (América do Sul)

País	Frequência	(Continua)
		Porcentagem
Colômbia	8.077	43,4%
Peru	6.085	32,7%
Chile	2.334	12,5%
Argentina	804	4,3%
Bolívia	314	1,7%
Venezuela	269	1,4%
Brasil	267	1,4%
Equador	215	1,2%
Paraguai	89	0,5%
Uruguai	75	0,4%
Suriname	66	0,4%

		(Conclusão)
Guiana	25	0,1%
Guiana francesa	7	0%
Ilhas Malvinas	1	0%
Total	18.628	100%

FONTE: Elaborado pela autora (2017)

Distribuindo a quantidade de mortos e feridos ao longo dos anos observa-se que houve uma concentração de vítimas fatais entre os anos de 1980 e 1985, sendo que no ano de 1984 foram registradas 2.995 mortes. Em relação aos feridos a maior concentração foi no ano de 1994 com 1.497 registros (gráfico 5).

Gráfico 5 - Distribuição de mortos e feridos ao longo dos anos (América do Sul)



FONTE: Elaborado pela autora (2017)

O *Shining\_Path\_(SL)* foi responsabilizado por 4.547 (24,4%) dos ataques ocorridos na América do Sul. Em 5.159 (27,7%) nenhum grupo responsabilizou-se pelo ataque.

A média de mortes foi de 1,69 por ataque sendo que o maior número registrado foi de 275 no ano de 1998 na Colômbia, a cidade onde ocorreu o incidente é desconhecida o grupo responsável pelo ataque foi o *Revolutionary\_Armed\_Forces\_of\_Colombia\_(FARC)*. Em 10.736 ataques não houve vítimas fatais. Em relação aos feridos a média foi de 0,98 por ataque, o maior número registrado foi de 250 no ano de 1989 na cidade de Bogotá na Colômbia, o grupo responsável pelo ataque foi o *The\_Extraditables*.

O Brasil possui 267 ataques registrados, sendo um número baixo se comparado a Colômbia, Peru e Chile que registram em média 6.000 ataques representando assim 1,4% dos incidentes ocorridos na América do Sul. Assim como no restante da região o número de mortos e feridos sofreram várias oscilações significativas ao longo dos anos conforme apresentado no gráfico 6. Observa-se que o número de vítimas, bem como o número de feridos, vem diminuindo ao longo dos anos da mesma forma que no restante da América de Sul.

Gráfico 6 - Distribuição de mortos e feridos ao longo dos anos (Brasil)



FONTE: Elaborado pela autora (2017)

Com 538 registros a Ásia central ocupa o penúltimo lugar dentre as regiões dispostas pela GTD, diferente das demais regiões não foram registrados ataques antes da década de 90 conforme apresentado no gráfico 7, o maior número de ataques ocorreu 1992.

Gráfico 7 - Ataques registrados na GTD ao longo dos anos (Ásia Central)



FONTE: Elaborado pela autora (2017)

A tabela 12 mostra a distribuição de ataques por país onde destaca-se a Geórgia onde ocorreram 212 (39,5%) dos ataques registrados nesta região.

Tabela 12 - Distribuição de ataques por país (Ásia central)

País	Frequência	Porcentagem
Armênia	22	4,1%
Azerbaijão	45	8,4%
Geórgia	212	39,4%
Cazaquistão	24	4,5%
Cazaquistão	28	5,2%
Tajiquistão	184	34,2%
Turquemenistão	2	0,4%
Uzbequistão	21	3,9%
Total	538	100,0%

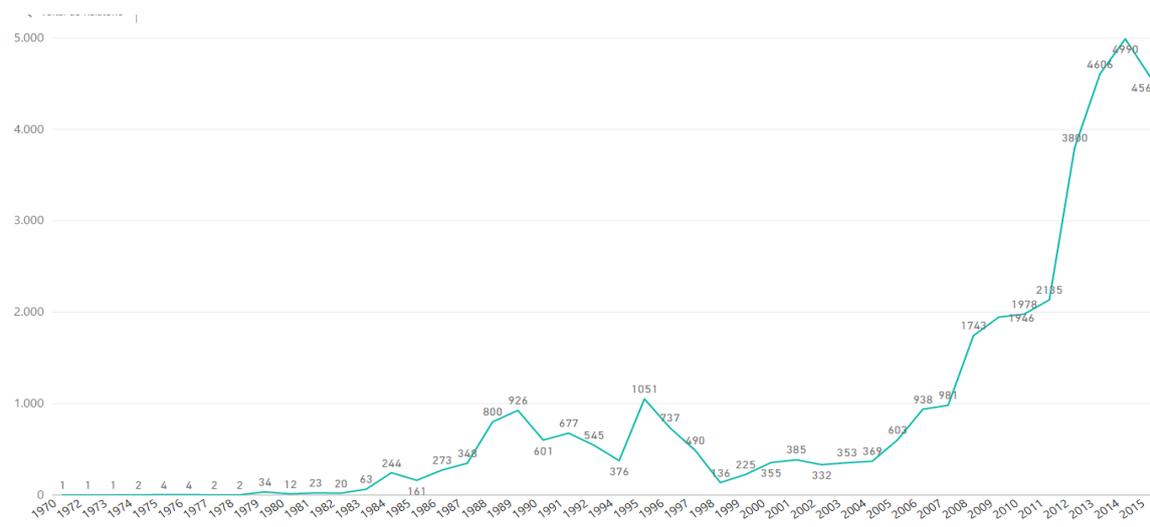
FONTE: Elaborado pela autora (2017)

Em 427 (79,4%) casos não foi possível identificar o grupo responsável pelo ataque, o *South Ossetian Separatists* e o *Supporters of Ex-President Gamsakhurdia*, foram responsabilizados por 10 ataques cada.

A média de vítimas fatais foi de 1,82 por ataque sendo que o maior número registrado foi de 50 mortos no ano de 1992 no Azerbaijão na cidade de Nagorno-Karabakh o grupo responsável pelo ataque foi o *Armenian Guerrillas*. Em relação aos feridos a média foi de 3,72 sendo que o maior número foi de 800 também no ano de 1992 no Tajiquistão na cidade de Dushanbe, o grupo responsável pelo ataque é desconhecido.

A Ásia Meridional que ocupa a 2º posição dentre as regiões analisadas com 37.841 ataques registrados. Através do gráfico 8 observa-se que houve um aumento significativo de ataques entre os anos de 2012 e 2015.

Gráfico 8 - Ataques registrados na GTD ao longo dos anos (Ásia Meridional)



FONTE: Elaborado pela autora (2017)

A distribuição de ataques por país pode ser observada na tabela 13 onde nota-se que a maior parte dos ataques ocorreu no Paquistão.

Tabela 13 - Distribuição de ataques por país (Ásia Meridional)

País	Frequência	Porcentagem
Paquistão	12.768	33,7%
Índia	9.940	26,3%
Afeganistão	9.690	25,6%
Sri Lanka	2.982	7,9%
Bangladesh	1.513	4,0%
Nepal	923	2,4%
Maldivas	17	0,0%
Butão	6	0,0%
Mauritius	2	0,0%
Total	37.841	100,0%

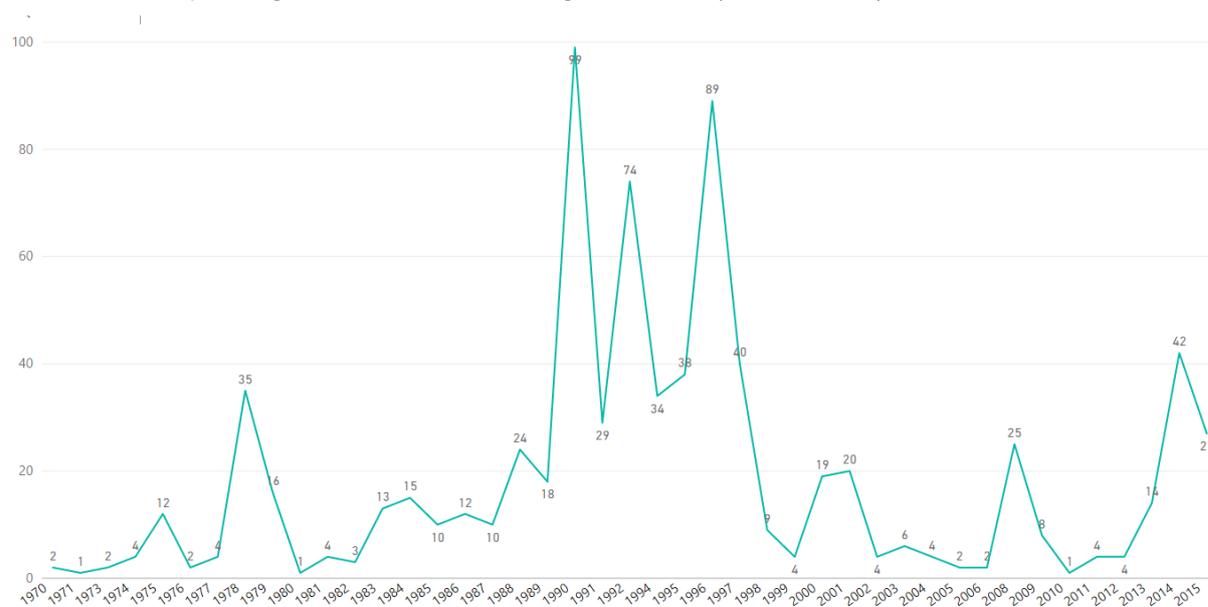
FONTE: Elaborado pela autora (2017)

Em 18.837 (49,8%) ataques nenhum grupo foi responsabilizado, desta forma o líder de ataques nesta região foi o *Taliban* responsabilizado por 5.501(14,5%) ataques.

A média de mortos foi 2,3 por ataque, sendo que o maior número de vítimas fatais registrado foram 518 vítimas, o ataque foi realizado na cidade de Dhading district no Nepal no ano de 2004, o grupo responsável pelo ataque foi o *Communist Party of Nepal- Maoist (CPN-M)*. A média de feridos foi de 3,3 por ataque, o maior número registrado foi de 1.272, na cidade de Colombo no Sri Lanka, o grupo responsável pelo ataque foi o *Liberation Tigers of Tamil Eelam (LTTE)*.

A Ásia Oriental ocupa a 10ª posição dentre as demais regiões analisadas, com 786 registros de ataques. No gráfico 9 onde estes registros são distribuídos de acordo com o ano em que ocorreram nota-se uma grande concentração de ataques na década 90.

Gráfico 9 - Ataques registrados na GTD ao longo dos anos (Ásia Oriental)



FONTE: Elaborado pela autora (2017)

A tabela 14 que apresenta a distribuição de ataques por país mostra que metade dos ataques registrados na Ásia oriental ocorreram do Japão.

Tabela 14 - Distribuição de ataques por país (Ásia Oriental)

País	Frequência	Porcentagem
Japão	385	50,1%
China	242	31,5%
Taiwan	46	6,0%
Coréia do Sul	36	4,7%
Macau	33	4,3%
Hong kong	26	3,4%
North korea	1	0,1%
TOTAL	769	100,0%

FONTE: Elaborado pela autora (2017)

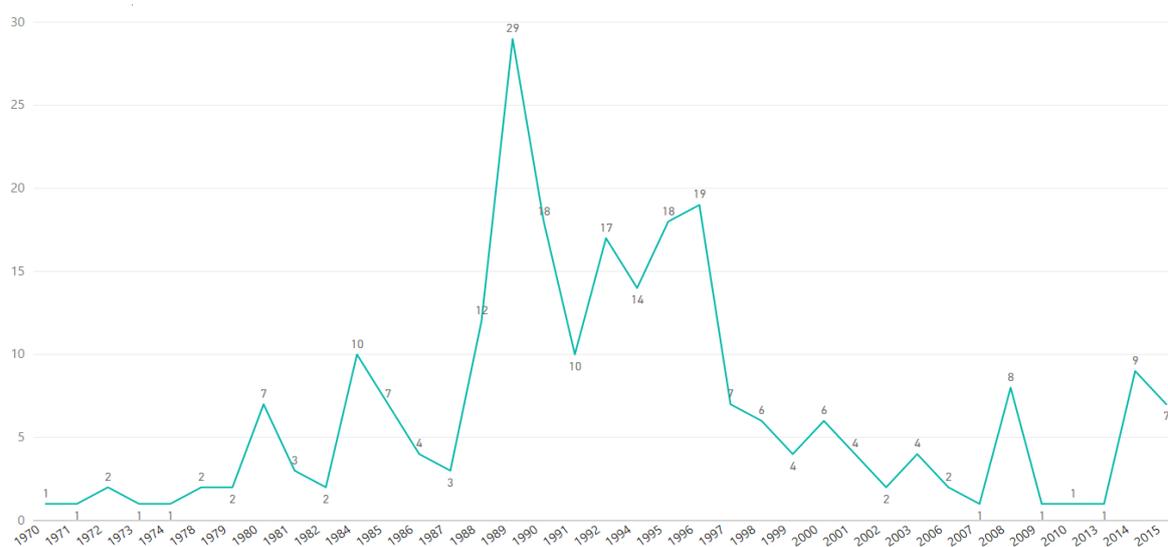
Em 442 (56,2%) casos nenhum grupo responsabilizou-se pelo ataque, desta forma sendo responsável por 82 (10,6%) ataques o *Uighur Separatists* foi o grupo que mais executou ataques nesta região.

A média de feridos foi de 1,5 por ataque sendo que o maior número registrado foi de 184 na cidade de Urumqi na China no ano de 2009; em 38 ataques não foi possível identificar o número de mortos e em 594 (75,6%) não houve vítimas registradas.

Em relação aos feridos a média foi de 12,4 por ataque sendo que o maior número registrado foi de 5.500 na cidade de Tokyo no Japão; em 44 casos não foi possível identificar o número de feridos e em 542 (68,9%) casos não houve feridos registrados.

A Austrália e Oceania apresenta o menor número de ataques dentre as 12 regiões analisadas com 246 registros. O gráfico 10 mostra que a maior parte dos ataques ocorreu no ano de 1989 com 29 registros.

Gráfico 10 - Ataques registrados na GTD ao longo dos anos (Austrália e Oceania)



FONTE: Elaborado pela autora (2017)

A tabela 15 mostra que a Austrália foi o país com o maior número de ataques com 88 registros representando 35,8 % do total de ataques ocorridos nesta região.

Tabela 15 - Distribuição de ataques por país (Austrália e Oceania)

País	Frequência	Porcentagem
Austrália	88	35,8%
Papua-nova guiné	81	32,9%
Nova caledónia	31	12,6%
Nova zelândia	18	7,3%
Fiji	17	6,9%
Ilhas salomão	4	1,6%
Polinésia francesa	3	1,2%
Vanuatu	2	0,8%
Novas hébridas	1	0,4%
Wallis e futuna	1	0,4%
Total	246	100%

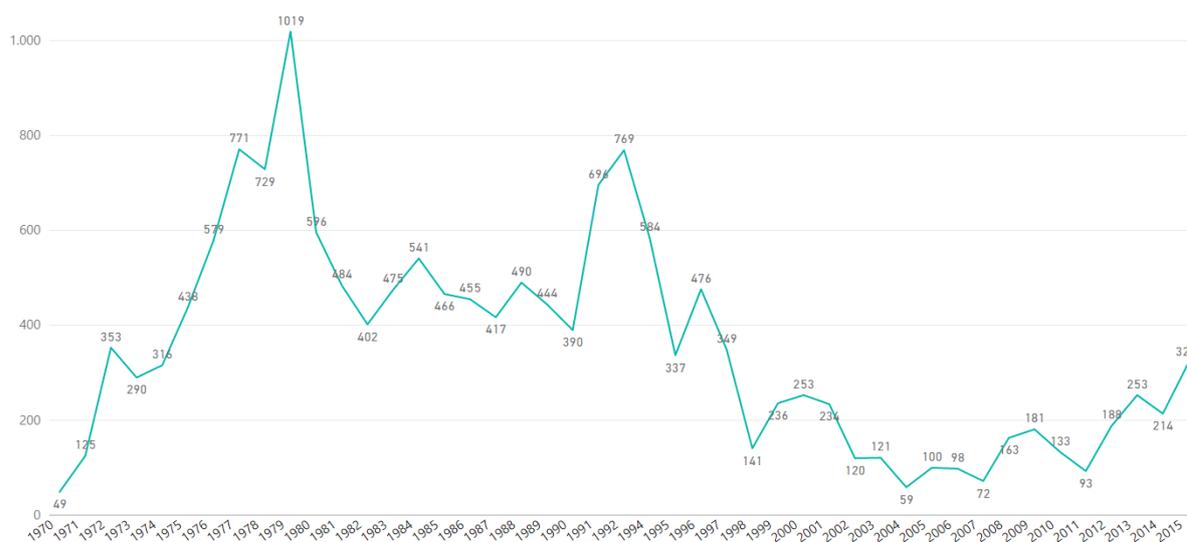
FONTE: Elaborado pela autora (2017)

Em 146 (59,3%) casos não há registro do grupo responsável pelo ataque desconsiderando estes o Bougainville Revolutionary Army (BRA) é responsável por 28 (11,4%) ataques.

A média de vítimas fatais foi de 0,60 por ataque sendo o maior número registrado foi de 17 mortes em dois ataques distintos um ocorrido na cidade de Buin no Papua New Guinea em 1994 e outro na cidade de Goosanah na Nova Caledónia, os responsáveis pelos ataques foram o *Bougainville Revolutionary Army* (BRA) e o *Kanak Separatists* respectivamente.

A Europa Ocidental ocupa a 4ª posição dentre as doze regiões analisadas com 16.020 registros. A distribuição destes no decorrer dos anos é apresentada no gráfico 11, onde observa-se uma maior grande concentração de ataques no ano de 1979 com 1.019 registros, nota-se também que o número de ataques vem diminuindo com o passar dos anos.

Gráfico 11 - Ataques registrados na GTD ao longo dos anos (Europa Ocidental)



FONTE: Elaborado pela autora (2017)

A distribuição de ataques por país é apresentada na tabela 16 onde o Reino Unido aparece como o país que mais sofreu ataques com 4.992 (31,2%) registros seguidos pela Espanha com 3.239 (20,2%) registros.

Tabela 16 - Distribuição de ataques por país (Europa Ocidental)

País	Frequência	Porcentagem
Reino Unido	4.992	31,2%
Espanha	3.239	20,2%
França	2.617	16,3%
Itália	1.545	9,6%

(Continua)

(Conclusão)

Grécia	1.200	7,5%
Alemanha	649	4,1%
Alemanha Ocidental	541	3,4%
Irlanda	274	1,7%
Bélgica	142	0,9%
Portugal	139	0,9%
Chipre	130	0,8%
Holanda	121	0,8%
Suíça	107	0,7%
Áustria	106	0,7%
Suécia	102	0,6%
Dinamarca	40	0,2%
Malta	20	0,1%
Noruega	18	0,1%
Luxemburgo	16	0,1%
Finlândia	15	0,1%
Islândia	4	0,0%
Andorra	1	0,0%
Gibraltar	1	0,0%
Vaticano	1	0,0%
Total	16.020	100,0%

FONTE: Elaborado pela autora (2017)

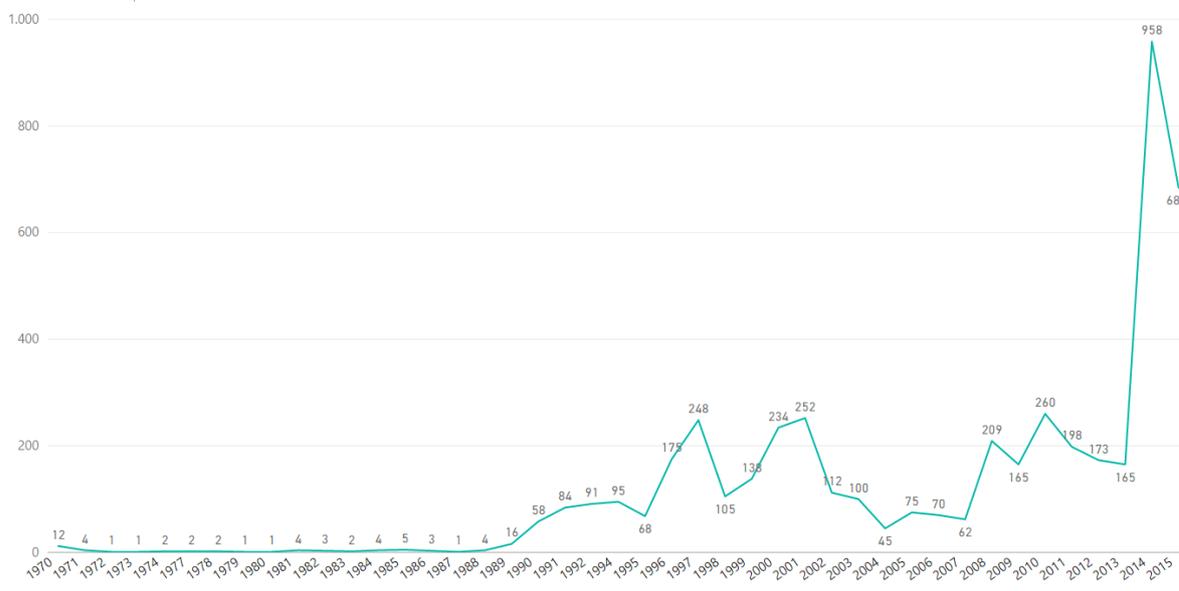
Em 4.712 (29,4%) casos não foi possível identificar o grupo responsável pelo ataque, desconsiderando estes o *Irish Republican Army (IRA)* foi o responsabilizado por 2.667 (16,6%) ataques registrados nesta região.

A média de vítimas fatais foi de 0,42 por ataque sendo que o maior número registrado foi de 270 no ano de 1988 na cidade de Lockerbie no Reino Unido; Em 912 (5,69%) registros não foi possível identificar o número de mortos e em 11.337 (70,7%) não houve vítimas fatais registradas. A média de feridos foi de 1,2 por ataques sendo que o maior número registrado foi de 450 no ano de 2004 em Madrid na Espanha. Em 2.511 (15,6%) registros não foi possível identificar o número de feridos e em 10.976 (68,51%) não houve feridos registrados.

Com 4.892 ataques registrados a Europa Oriental ocupa a 8ª posição dentre as 12 regiões analisadas; a distribuição destes registros pode ser analisada no gráfico 12 onde observa-se que o número de ataques vem

aumentando no decorrer com anos com uma grande concentração no ano de 2014 onde foram registrados 958 ataques.

Gráfico 12 - Ataques registrados na GTD ao longo dos anos (Europa Oriental)



FONTE: Elaborado pela autora (2017)

A tabela 17 apresenta a distribuição de ataques por país, onde a Rússia aparece sendo o país que mais sofreu ataques 2.104, representando 43% do total dos ataques registrados na Europa Oriental.

Tabela 17 - Distribuição de ataques por país (Europa Oriental)

País	Frequência	(Continua)
		Porcentagem
Rússia	2.104	43,0%
Ucrânia	1.583	32,4%
Iugoslávia	203	4,1%
Kosovo	180	3,7%
Bósnia e Herzegovina	159	3,3%
Macedônia	117	2,4%
União soviética	78	1,6%
Albânia	77	1,6%
Croácia	57	1,2%
Bulgária	51	1,0%
Hungria	45	0,9%
Alemanha oriental	38	0,8%
Polônia	34	0,7%
República checa	27	0,6%
Moldávia	20	0,4%
Eslováquia	17	0,3%

(Conclusão)

Estónia	16	0,3%
Letónia	16	0,3%
Bielorrússia	13	0,3%
Sérvia	11	0,2%
Sérvia e Montenegro	11	0,2%
Checoslováquia	10	0,2%
Lituânia	8	0,2%
Romênia	6	0,1%
Eslovênia	6	0,1%
Montenegro	5	0,1%
Total	4.892	100,0%

FONTE: Elaborado pela autora (2017)

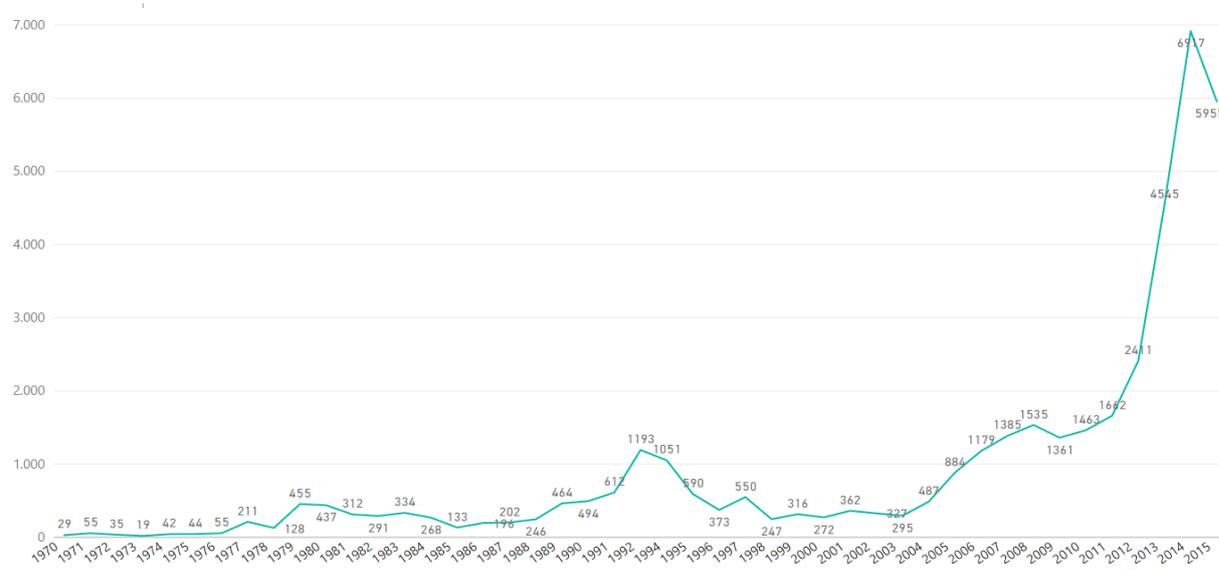
Em 3.229 (66%) casos nenhum grupo responsabilizou-se oficialmente pelo ataque. Desta forma, o *Donetsk People's Republic*, responsável por 600 (12,3%) ataques, foi o grupo que mais cometeu ataques nesta região.

A média de mortos foi de 1,52 por ataque sendo que o maior número registrado foi de 344 no ano de 2004 na cidade de Beslan na Rússia. Em 168 (3,43%) registros não foi possível identificar o número de vítimas fatais e em 3.122 (63,81%) não houve mortos registrados. Em relação aos feridos a média foi de 2,53 por ataque, maior número de ataques foi de 727 coincidentemente no mesmo ataque em que foram registrados 344 mortos; em 252 (5,15%) não foi possível calcular o número de feridos e em 3.015 (61,65%) não houve feridos registrados.

Com 40.422 ataques executados no Oriente Médio e Norte da África representam 25,78 % do total de registros da GTD. Neste sentido está é a região com maior número de ataques dentre as 12 analisadas.

No gráfico 13 onde os registros estão distribuídos de acordo com o ano em que ocorreram nota-se que um aumento significativo de ataques após o ano de 2007 e uma grande concentração no ano de 2014.

Gráfico 13 - Ataques registrados na GTD ao longo dos anos (Oriente Médio e Norte da África)



FONTE: Elaborado pela autora (2017)

Na distribuição dos registros de ataques por país (tabela 18) observa-se que os ataques ocorridos no Iraque (18.770) representam 46,4 % do total de ataques executados nesta região.

Tabela 18 - Distribuição de ataques por país (Oriente Médio e Norte da África)

(Continua)

País	Frequência	Porcentagem
Iraque	18.770	46,4%
Turquia	3.557	8,8%
Argélia	2.720	6,7%
Iémen	2.598	6,4%
Líbano	2.413	6,0%
Israel	2.085	5,2%
Cisjordânia e faixa de gaza	1.990	4,9%
Egito	1.799	4,5%
Líbia	1.643	4,1%
Síria	1.468	3,6%
Irã	667	1,7%
Arábia Saudita	193	0,5%
Bahrein	182	0,5%
Jordânia	92	0,2%
Tunisia	92	0,2%
Kuwait	74	0,2%

(Conclusão)

Marrocos	36	0,1%
Emirados árabes unidos	22	0,1%
Catar	7	0,0%
Iêmen do norte	6	0,0%
Saara ocidental	5	0,0%
Iêmen do sul	2	0,0%
International	1	0,0%
<b>Total</b>	<b>40.422</b>	<b>100,0%</b>

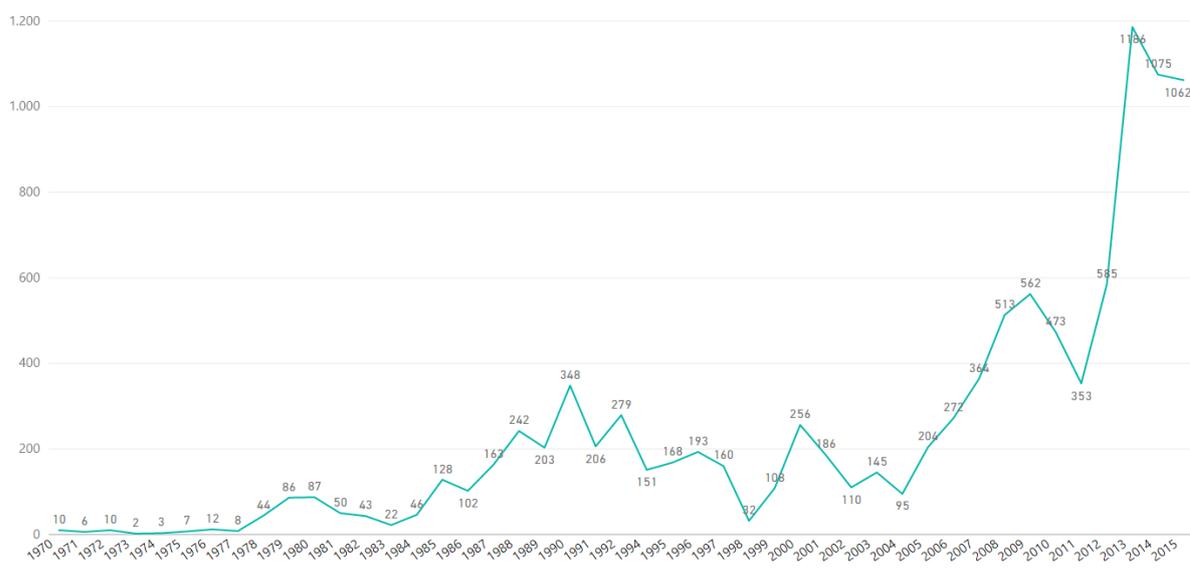
FONTE: Elaborado pela autora (2017)

Em 24.664 registros o grupo que executou o ataque é desconhecido, desta forma o grupo responsável pelo maior número de ataques nesta região é o *Islamic State of Iraq and the Levant (ISIL)* responsabilizado por 2.824 (7%) ataques.

A média de vítimas fatais foi de 2,75 sendo que o maior número registrado foi de 1.500 mortes no ano de 2014 na cidade de Tikrit no Iraque, é importante ressaltar que este é o maior de número de vítimas fatais registrado não só desta região, mas sim dentre as 12 consideradas nesta pesquisa; em 1.682 (4,18%) casos não foi possível calcular o número de mortes ocorridas e em 16.488 (41%) não houve mortos registrados. Em relação aos feridos, a média foi de 4,8 por ataque sendo que o maior número registrado foi de 750 em dois ataques distintos ocorridos no ano de 2007 no Iraque, nas cidades de Qahtaniya e Jazeera, em 2.768 (68,8%) não foi possível calcular o número de feridos e em 1.517 (37,71%) não houve feridos registrados.

O Sudeste Asiático ocupa a 6ª dentre as 12 regiões apresentadas possui 10.360 ataques registrados, o gráfico 14 mostra a distribuição destes no decorrer dos anos, onde nota-se oscilações no número de ataque nas décadas de 80 e 90 um aumento significativo nos 3 últimos anos (2013 à 2015).

Gráfico 14 - Ataques registrados na GTD ao longo dos anos (Sudeste Asiático)



FONTE: Elaborado pela autora (2017)

A distribuição dos ataques por país é apresentada na tabela 19, onde observa-se que 5.576 (53,8%) ocorreram em um único país (Filipinas).

Tabela 19 - Distribuição de ataques por país (Sudeste Asiático)

País	Frequência	Porcentagem
Filipinas	5.576	53,8%
Tailândia	3.338	32,2%
Indonésia	714	6,9%
Myanmar	347	3,3%
Camboja	258	2,5%
Malásia	76	0,7%
Laos	22	0,2%
Timor-leste	10	0,1%
Vietnã	10	0,1%
Singapura	7	0,1%
Brunei	1	0,0%
Vietnã do sul	1	0,0%
Total	10.360	100%

FONTE: Elaborado pela autora (2017)

Desconsiderando os 4.990 (48%) registros onde não foi possível identificar o grupo responsável pelo ataque, o *New People's Army (NPA)* foi o que mais cometeu ataques, sendo responsável por 2.239 (21,6%).

A média de mortes foi de 1,39 por ataque sendo que o maior número registrado foi de 116 no ano de 2004 na cidade de Manila nas Filipinas; em 223 casos (2,15%) não foi possível calcular o número de vítimas fatais ocorridas e em 5.502 (53%) não houve mortes registradas.

Em relação ao número de feridos a média foi de 2,4 por ataque sendo que o maior número registrado foi de 400 no ano de 2012 na cidade de Hat Yai na Tailândia, em 360 (3,47%) não foi possível calcular o número de feridos ocorridos e em 6.277 (60,5%) não houve feridos registrados.

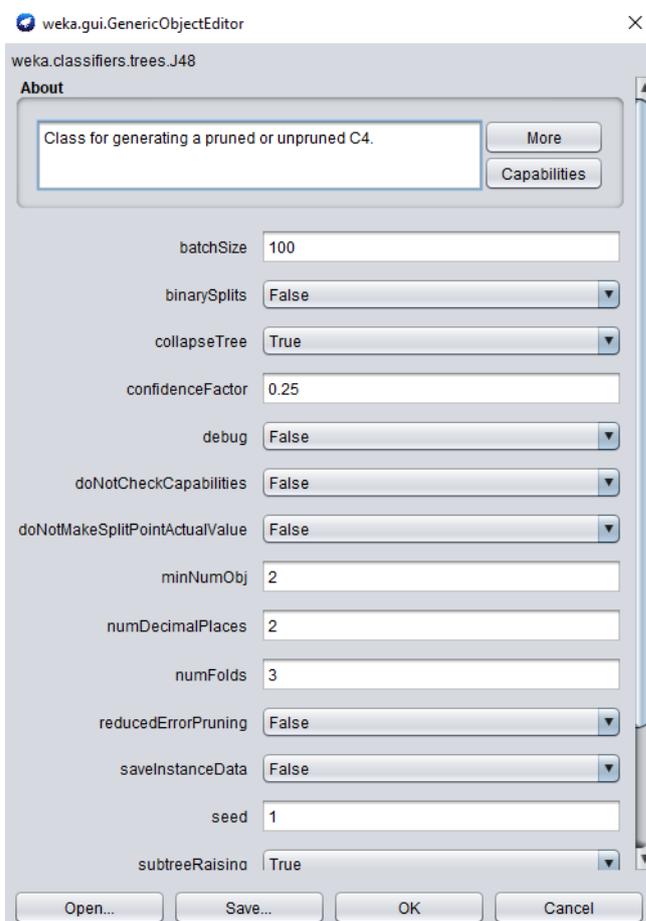
## 4.2 TAREFAS DE MINERAÇÃO DE DADOS

Esta seção apresenta os resultados obtidos através dos experimentos relacionados as tarefas classificação, associação e agrupamento utilizando os algoritmos J48, Apriori e K-means.

### 4.2.1 J48

O algoritmo J48, que é a implementação do c4.5 no Weka, foi executado com os parâmetros *default* apresentados na figura 6.

Figura 6 - Parâmetros default J48 (C4.5)



FONTE: Dados da pesquisa utilizando o software Weka (2017)

No primeiro experimento realizado por meio do c4.5 o atributo “sucesso”, classificado com 0 quando o alvo pretendido não foi atingido e como 1 quando o alvo pretendido foi atingido, foi utilizado como atributo-meta. O tempo de execução foi de em 0,23 segundos; a taxa de acerto foi de 93,78 % sendo que 17.470 instâncias foram classificadas corretamente e 1.158 incorretamente, conforme apresentado na matriz de confusão (tabela 20), onde a soma das diagonais representam as instâncias classificadas corretamente e incorretamente.

Tabela 20 - Matriz de confusão gerada no experimento 1 utilizando o J48 (C4.5)

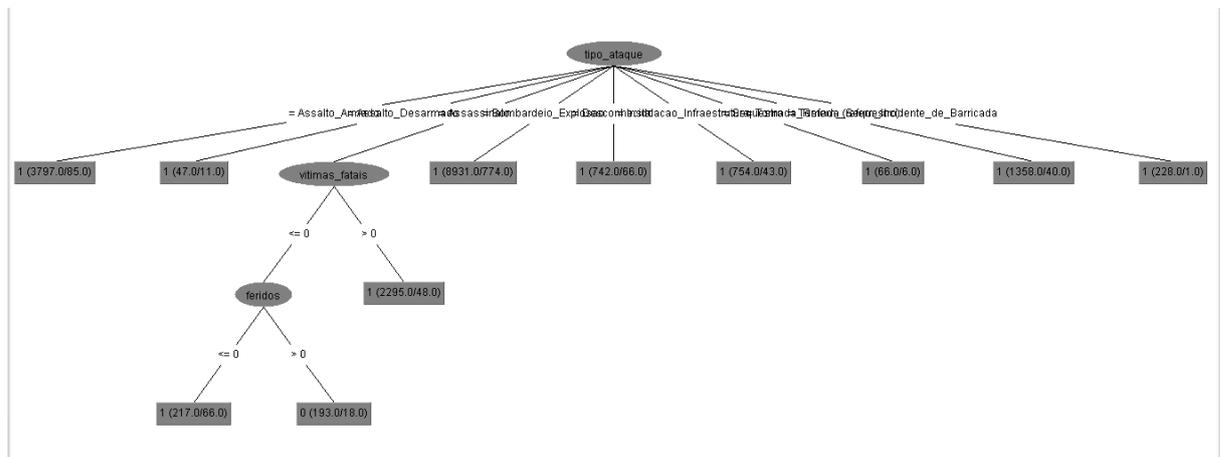
```

=== Confusion Matrix ===
      a    b  <-- classified as
    175 1140 |    a = 0
     18 17295 |    b = 1
  
```

FONTE: Dados da pesquisa utilizando o software Weka (2017)

A árvore de decisão (figura 7) gerada é de tamanho 14 e contém 11 folhas. Como o *weka* não permite a edição da árvore e nem a exportação da mesma no formato png ou jpeg parte de sua visualização ficou comprometida, no entanto pode ser vista selecionando o campo “*Auto Scale*” no próprio *software*, ou por meio da visualização da árvore no formato de texto (figura 8). Nota-se o “*Tipo\_Ataque*” na raiz da árvore, sendo este o atributo de maior influência seguido por “*Vitimas\_fatais*” e “*Feridos*” quando o tipo de ataque foi assassinato.

Figura 7 - Árvore de decisão J48, experimento 1 (C4.5)



FONTE: Resultados da pesquisa utilizando o software Weka (2017)

Figura 8 - Árvore de decisão no tipo texto J48 (C4.5), experimento 1

```

J48 pruned tree
-----

tipo_ataque = Assalto_Armado: 1 (3797.0/85.0)
tipo_ataque = Assalto_Desarmado: 1 (47.0/11.0)
tipo_ataque = Assassinato
|  vitimas_fatais <= 0
|  |  feridos <= 0: 1 (217.0/66.0)
|  |  feridos > 0: 0 (193.0/18.0)
|  vitimas_fatais > 0: 1 (2295.0/48.0)
tipo_ataque = Bombardeio_Explosao: 1 (8931.0/774.0)
tipo_ataque = Desconhecido: 1 (742.0/66.0)
tipo_ataque = Instalacao_Infraestrutura: 1 (754.0/43.0)
tipo_ataque = Sequestro: 1 (66.0/6.0)
tipo_ataque = Tomada_Refem_(Sequestro): 1 (1358.0/40.0)
tipo_ataque = Tomada_refem_Incidente_de_Barricada: 1 (228.0/1.0)

```

FONTE: Resultados da pesquisa utilizando o software Weka (2017)

No segundo experimento realizado ainda com o mesmo algoritmo, o atributo “País”, foi utilizado como atributo meta. O tempo de execução foi de 0,11 segundos; a taxa de acerto de 92,10 %, sendo que 17.147 foram classificadas corretamente e 1.471 incorretamente. A matriz de confusão disposta na tabela 21 foi adaptada para melhor análise nota-se que em 4 (Guiana, Guiana Francesa, Suriname e Uruguai) dos 14 países que fazem parte desta região todas as instâncias foram classificadas corretamente.

Tabela 21 - Matriz de confusão gerada no experimento 2 com a aplicação J48 (C4.5)

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	classified as
	729	0	0	1	73	0	0	0	0	0	1	0	0	0	a=Argentina
	0	281	0	0	30	0	0	0	0	0	2	0	0	1	b=Bolivia
	0	0	205	3	59	0	0	0	0	0	0	0	0	0	c=Brazil
	2	0	2	2234	92	0	0	0	0	0	3	0	0	1	d=Chile
	20	21	0	21	7954	1	0	0	0	1	51	0	0	8	e=Colombia
	0	0	0	0	41	173	0	0	0	0	1	0	0	0	f=Ecuador
	0	0	0	0	1	0	0	0	0	0	0	0	0	0	g=Falkland_Islands
	0	0	0	0	0	0	0	7	0	0	0	0	0	0	h=French_Guiana
	0	0	0	0	2	0	0	0	23	0	0	0	0	0	i=Guyana
	0	0	0	1	27	0	0	0	0	60	1	0	0	0	j=Paraguay
	6	3	0	15	873	1	0	0	0	0	5187	0	0	0	k=Peru
	0	0	1	0	19	0	0	0	0	0	0	46	0	0	l=Suriname
	0	0	0	0	2	0	0	0	0	0	0	0	73	0	m=Uruguay
	0	2	0	2	79	0	0	0	0	0	1	0	0	185	n=Venezuela
	757	307	208	2277	9252	175	0	7	23	61	5247	46	73	195	TOTAL
	96%	92%	99%	98%	86%	99%	0%	100%	100%	98%	99%	100%	100%	95%	

FONTE: Resultados da pesquisa utilizando o software Weka adaptado pela autora. (2017)

A árvore de decisão gerada é de tamanho 3.234 e possui 3.202 folhas, tornando assim impossível sua visualização devido ao grande número de folhas

sobrepostas umas às outras. Por meio da árvore no formato de texto é possível notar que o atributo cidade exerce maior influência, observa-se também que os atributos “tipo\_de\_ataque” e “armas\_utilizadas” aparecem com maior frequência no restante da árvore.

O terceiro experimento realizado através do J48 foi realizado com os atributos numéricos discretizados conforme mencionado na seção de “preparação de dados.” A árvore de decisão gerada é de tamanho 29 e possui 25 folhas (figura 9). O tempo de execução do algoritmo foi de 0,56 segundos. Como a visualização da árvore também ficou comprometida apresenta-se a árvore no tipo texto onde nota-se a atributo “vitimas\_fatais” como o de maior influência. A taxa de acerto foi de 93,77 % sendo que 17.468 instâncias foram classificadas corretamente e 1.160 incorretamente, conforme apresentado na matriz de confusão (tabela 22), onde a soma das diagonais representam as instâncias classificadas corretamente e incorretamente.

Figura 9 - Árvore de decisão no tipo texto J48 (C4.5), experimento 3

```

-----
vitimas_fatais = A
|  tipo_ataque = Assalto_Armado: 1 (1228.0/59.0)
|  tipo_ataque = Assalto_Desarmado
|  |  armas_utilizadas = Armas_Falsas: 0 (1.0)
|  |  armas_utilizadas = Armas_Fogo: 1 (0.0)
|  |  armas_utilizadas = Biologica: 0 (1.0)
|  |  armas_utilizadas = Corpo_Corpo: 1 (22.0/1.0)
|  |  armas_utilizadas = Desconhecido: 1 (0.0)
|  |  armas_utilizadas = Explosivos_Bombas_Dinamite: 1 (0.0)
|  |  armas_utilizadas = Incendiaria: 1 (0.0)
|  |  armas_utilizadas = Outro: 1 (2.0)
|  |  armas_utilizadas = Quimica: 0 (14.0/6.0)
|  |  armas_utilizadas = Sabotagem_Equipamento: 1 (0.0)
|  |  armas_utilizadas = Veiculos_(exceto_carros_caminhoes_bomba): 1 (0.0)
|  tipo_ataque = Assassinato
|  |  feridos = A: 1 (205.0/66.0)
|  |  feridos = B: 0 (192.0/18.0)
|  |  feridos = C: 0 (0.0)
|  |  feridos = D: 1 (1.0)
|  tipo_ataque = Bombardeio_Explosao: 1 (7016.0/712.0)
|  tipo_ataque = Desconhecido: 1 (289.0/41.0)
|  tipo_ataque = Instalacao_Infraestrutura: 1 (620.0/35.0)
|  tipo_ataque = Sequestro: 1 (54.0/6.0)
|  tipo_ataque = Tomada_Refem_(Sequestro): 1 (904.0/29.0)
|  tipo_ataque = Tomada_refem_Incidente_de_Barricada: 1 (187.0/1.0)
vitimas_fatais = B: 1 (6195.0/155.0)
vitimas_fatais = C: 1 (28.0)
vitimas_fatais = D: 1 (1669.0/26.0)

```

FONTE: Resultados da pesquisa utilizando o software Weka (2017)

Tabela 22 - Matriz de confusão gerada no experimento 3 utilizando o J48 (C4.5)

```

=== Confusion Matrix ===
      a    b  <-- classified as
178 1137 |    a = 0
 23 17290 |    b = 1

```

FONTE: Resultados da pesquisa utilizando o software Weka (2017)

No quarto experimento realizado ainda com J48 e atributos numéricos discretizados, o atributo “País”, foi utilizado como atributo meta. O tempo de execução foi de 0,12 segundos; a taxa de acerto de 95,95 %, sendo que 17.874 foram classificadas corretamente e 754 incorretamente. A matriz de confusão disposta na tabela 23 foi adaptada para melhor análise.

Tabela 23 - Matriz de confusão gerada no experimento 4 com a aplicação J48 (C4.5)

a	b	c	d	e	f	g	h	i	j	k	l	m	n	classified as
729	0	0	1	73	0	0	0	0	0	1	0	0	0	a=Argentina
0	281	0	0	31	0	0	0	0	0	2	0	0	1	b=Bolivia
0	0	205	3	61	0	0	0	0	0	0	0	0	0	c=Brazil
2	0	2	2224	103	0	0	0	0	0	2	0	0	1	d=Chile
19	18	0	16	7960	1	0	0	0	1	54	0	0	8	e=Colombia
0	0	0	0	41	173	0	0	0	0	1	0	0	0	f=Ecuador
0	0	0	0	1	0	7	0	0	0	0	0	0	0	g=Falkland_Islands
0	0	0	0	2	0	0	23	0	0	0	0	0	0	h=French_Guiana
0	0	0	0	1	0	0	0	0	0	0	0	0	0	i=Guyana
0	0	0	1	28	0	0	0	0	60	1	0	0	0	j=Paraguay
9	3	0	14	877	1	0	0	0	0	5181	0	0	0	k=Peru
0	0	1	0	19	0	0	0	0	0	0	46	0	0	l=Suriname
0	0	0	0	2	0	0	0	0	0	0	0	73	0	m=Uruguay
0	2	0	2	78	0	0	0	0	0	1	0	0	186	n=Venezuela
759	304	208	2261	9277	175	7	23	0	61	5243	46	73	196	TOTAL
96%	92%	99%	98%	86%	99%	0%	100%	0%	98%	99%	100%	100%	95%	

FONTE: Resultados da pesquisa utilizando o software Weka (2017)

A árvore de decisão gerada é de tamanho 3.194 e possui 3.172 folhas, tornando assim impossível sua visualização devido ao grande número de folhas sobrepostas umas às outras. Por meio da árvore no formato de texto é possível notar que o atributo cidade exerce maior influência, observa-se também que os atributos “tipo\_de\_ataque” e “armas\_utilizadas” aparecem com maior frequência no restante da árvore da mesma forma com obtido no experimento 2.

O quadro 3 apresenta os resultados de ambos os experimentos

Quadro 3 - Experimentos realizados com o J48 (C 4.5)

	Experimento 1 (Atributo meta = "Sucesso")	Experimento 2 (Atributo meta = "país")	Experimento 3 (Atributo meta = "sucesso" e atributos numéricos discretizados)	Experimento 4 (Atributo meta = "país" e atributos numéricos discretizados)
Instâncias classificadas corretamente	17.470 (93,78 %)	17.157 (92,10%)	17.468 (93,77 %)	17.148 (92,05 %)
Instâncias classificadas incorretamente	1.158 (6,22 %)	1.471 (7,90%)	1.160 (6,22%)	1.480 (7.945 %)
Estatística de kappa	0,218	0,883	0,2202	0,8822
Erro absoluto médio	0,1134	0,0152	0,1122	0,0155
Erro quadrado médio da raiz	0,2384	0,0887	0,2371	0,0892
Erro absoluto relativo	86,36%	15,50%	85.48%	15.75 %
Erro relativo ao quadrado da raiz	93,08%	40,05%	92.56%	40.27 %
Número de instâncias	18.628	18.628	18.628	18.628

FONTE: Resultados da pesquisa utilizando o software Weka adaptado pela autora. (2017)

O experimento 2 obteve melhores resultados em relação à classificação das instâncias, sendo que este classificou corretamente 17.157 instâncias corretamente.

Considerando a estatística de Kappa o quarto experimento obteve o melhor resultado, considerando que este teve o valor aproximado de 1, ou seja, teve a maior confiança de observação.

O erro absoluto médio e o erro quadrado médio da raiz indicam que o experimento 4 gerou menor diferença entre os valores atuais e os preditos, portanto, considerando estes parâmetros o experimento 4 também teve resultados mais satisfatórios.

O erro absoluto relativo e erro relativo ao quadrado da raiz de apresentam melhores resultados nos experimentos 2 e 4 indicando que estes obtiveram

mais precisão para previsão numérica do que 1º e 3º experimentos. É importante ressaltar que os experimentos 1 e 3 obtiveram um erro absoluto relativo é elevado 86,36% e 85,48% ou seja, o modelo estatístico gerado por meio do C4.5 não é preciso. Neste sentido pode-se afirmar que a utilização do atributo “sucesso” não obteve resultados satisfatórios nos experimentos realizados através do j48.

Como os resultados dos experimentos realizados com o atributo “sucesso” não obtiveram resultados satisfatórios, para o experimento realizado com a amostra relacionada aos incidentes ocorridos no Brasil, optou-se por pelo atributo “armas\_utilizadas” como meta, por ser o atributo que apresenta menor número de possibilidades.

A árvore gerada é de tamanho 10 e possui 9 folhas (figura 10), está apresentada no formato de texto devido ao comprometimento da imagem gerada pelo Weka. O tempo de execução foi de 0 (zero) segundos. Foram classificadas 224 instâncias corretamente (83,89%) e 43 (16,10%) incorretamente, conforme a matriz de confusão apresentada na tabela 24. Nota-se que o “tipo\_ataque” é o único atributo que sofre influência na construção da árvore.

Figura 10 - Árvore de decisão no tipo texto J48 (c4.5), experimento realizado com a amostra referente aos incidentes ocorridos no Brasil

J48 pruned tree

```

-----

tipo_ataque = Assalto_Armado: Armas_Fogo (50.0/2.0)
tipo_ataque = Assalto_Desarmado: Corpo_Corpo (4.0)
tipo_ataque = Assassinato: Armas_Fogo (64.0/17.0)
tipo_ataque = Bombardeio_Explosao: Explosivos_Bombas_Dinamite (65.0/1.0)
tipo_ataque = Desconhecido: Desconhecido (17.0)
tipo_ataque = Instalacao_Infraestrutura: Incendiaria (17.0/4.0)
tipo_ataque = Sequestro: Desconhecido (3.0)
tipo_ataque = Tomada_Refem_(Sequestro): Desconhecido (44.0/15.0)
tipo_ataque = Tomada_refem_Incidente_de_Barricada: Armas_Fogo (3.0/2.0)

```

FONTE: Resultados da pesquisa utilizando o software Weka

Tabela 24 - Matriz de confusão gerada no experimento com a amostra que representa os incidentes ocorridos no Brasil utilizando o J48 (C4.5)

a	b	c	d	e	f	g	<--	classified as	
0	0	0	1	0	0	0		a	= Armas_Falsas
0	93	2	13	2	2	0		b	= Armas_Fogo
0	5	4	0	0	1	0		c	= Corpo_Corpo
0	6	0	50	0	0	0		d	= Desconhecido
0	5	0	1	64	0	0		e	= Explosivos_Bombas_Dinamite
0	4	0	0	0	13	0		f	= Incendiaria
0	0	0	0	0	1	0		g	= Quimica
<b>0</b>	<b>113</b>	<b>6</b>	<b>65</b>	<b>66</b>	<b>17</b>	<b>0</b>			
0	82%	67%	77%	97%	76%	0%			

FONTE: Resultados da pesquisa utilizando o software Weka adaptado pela autora. (2017)

Os demais resultados são apresentados no quadro 4; onde nota-se um baixo erro absoluto e estatística de Kappa próxima de 1, sendo resultados satisfatórios para a tarefa de mineração proposta.

Quadro 4 - Resultados do experimento realizado com a amostra relacionada aos incidentes ocorridos no Brasil através do algoritmo J48 (C 4.5)

Instâncias classificadas corretamente	224 (83,89 %)
Instâncias classificadas incorretamente	43 (16.10 %)
Estatística de kappa	0.7705
Erro absoluto médio	0,0694
Erro quadrado médio da raiz	0.1944
Erro absoluto relativo	34.20 %
Erro relativo ao quadrado da raiz	61.20 %
Número de instâncias	267

FONTE: Resultados da pesquisa utilizando o software Weka adaptado pela autora. (2017)

#### 4.2.2 Apriori

O algoritmo Apriori proposto por Agrawal, Imielinksyi e Swami (1993), citados por Castro e Ferrari (2016) é o algoritmo mais utilizado para descobrir regras de associação. Para isto, executa múltiplas passagens sobre o banco de dados de transações, é capaz de trabalhar com um número grande de atributos, obtendo como resultado, várias alternativas combinatórias entre eles, a partir da realização de buscas sucessivas em toda a base de dados e, apesar disso, os autores apontam o ótimo desempenho em termos de processamento desse algoritmo. O Apriori foi executando alterando o parâmetro *numRules* para 1.000, possibilitando a visualização de todas as regras geradas. Os demais parâmetros foram utilizados em *default* conforme a figura 11.

Figura 11 - Parâmetros default Apriori



FONTE: Dados da pesquisa utilizando o software Weka (2017)

A partir da execução do Apriori foram geradas 242 regras, onde foi possível identificar algumas regras redundantes, tais como: cidade=Lima 2358 ==> pais=Peru 2358 e grupo\_responsavel=Shining\_Path\_(SL) vitimas\_fatais=C 1924 ==> pais=Peru 1923, é evidente que todos os habitantes de Lima moram

no Peru e o grupo *Shining Pathm* cometeu atentados somente no Peru até o momento, neste sentido estas regras podem ser consideradas irrelevantes.

A fim de evitar esta redundância de regras o Apriori foi executado novamente desconsiderando os atributos cidade e grupo responsável. Foram geradas 159 regras com confiança entre 99% e 90% sendo que:

- a) vitimas\_fatais=B feridos=A armas\_utilizadas=Armas\_Fogo ==> sucesso=1
- b) vitimas\_fatais=B feridos=A tipo\_ataque=Assassinato ==> sucesso=1
- c) vitimas\_fatais=B tipo\_ataque=Assassinato  
armas\_utilizadas=Armas\_Fogo ==> sucesso=1
- d) vitimas\_fatais=B feridos=A ==> sucesso=1
- e) feridos=A alvo=Governo\_Geral ==> sucesso=1
- f) vitimas\_fatais=B tipo\_ataque=Assassinato ==> sucesso=1
- g) feridos=A tipo\_ataque=Assalto\_Armado  
armas\_utilizadas=Armas\_Fogo ==> sucesso=1

Nota-se que ambas as regras apresentadas estão relacionadas ao atributo “sucesso”, ou seja, apesar da eliminação dos atributos “cidade” e “grupo responsável” ainda há redundância nas regras geradas, isto se deve ao fato de o atributo sucesso ser igual a 1 em 17.313 (93%) casos.

Desta forma o atributo “Sucesso”, foi retirado e o algoritmo Apriori executado novamente. Foram geradas 48 regras onde também observa-se redundância nas regras onde são mencionados os países peru e colômbia, considerando que estes juntos representam 76% do total de incidentes registrados, tais regras também tornam-se irrelevantes. Desta forma, o atributo “país” também foi eliminado no da tarefa de associação realizada neste estudo. Por fim, o Apriori foi executado novamente gerando 19 regras (apêndice 3) das quais podem ser consideradas relevantes:

- a) feridos=A tipo\_ataque=Bombardeio\_Explosao ==> vitimas\_fatais=A (conf =0.95);
- b) feridos=A tipo\_ataque=Assassinato ==> vitimas\_fatais=B (conf=0.91);
- c) tipo\_ataque=Assassinato armas\_utilizadas=Armas\_Fogo==> vitimas\_fatais=B conf:(0.91);

- d)  $\text{vitimas\_fatais}=B \text{ tipo\_ataque}=\text{Assalto\_Armado} \implies$   
 $\text{armas\_utilizadas}=\text{Armas\_Fogo}$  (conf =0.95).

O quadro 5 apresenta as regras destacadas com as suas devidas conclusões:

Quadro 5 - Regras e conclusões através do Apriori (América do Sul)

Regra	Interpretação
$\text{feridos}=A \text{ tipo\_ataque} = \text{Bombardeio\_Explosao} \implies \text{vitimas\_fatais}=A$	Quando o número de feridos foi zero e o tipo de ataque foi bombardeio e/ou explosão , o número de mortes (vitimas_fatais) também foi igual a zero
$\text{feridos}=A \text{ tipo\_ataque}=\text{Assassinato} \implies \text{vitimas\_fatais}=B$	Quando o número de feridos foi zero e o tipo de ataque foi assassinato, o número de óbitos foi entre 1 e 50;
$\text{tipo\_ataque}=\text{Assassinato} \text{ armas\_utilizadas}=\text{Armas\_Fogo} \implies \text{vitimas\_fatais}=B$	Quando o tipo de ataque foi assassinato e foram utilizadas armas de fogo, o número de vítimas fatais foi entre 1 e 50.
$\text{vitimas\_fatais}=B \text{ tipo\_ataque}=\text{Assalto\_Armado} \implies \text{armas\_utilizadas}=\text{Armas\_Fogo}$	Quando o número de óbitos foi entre 1 e 50 e tipo de ataque foi assalto armado, foram utilizadas armas de fogo.

FONTE: Dados da pesquisa utilizando o software Weka adaptado pela autora (2017)

Em relação ao experimento a ser realizado com a amostrada relacionada aos incidentes ocorridos no Brasil, o atributo “grupo\_responsável” foi eliminado pois dos 42 grupos que foram responsabilizados por ataques no país 25 cometeram apenas um ataque e em 181 casos nenhum grupo responsabilizou-se pelo ataque. Da mesma que no experimento anterior o atributo Sucesso também foi eliminado desta tarefa de classificação pois representa 233 dos 267 incidentes registrados.

Desta forma, o Apriori gerou 19 regras (Apêndice 4) das quais destacam-se:

- a)  $\text{feridos}=A \text{ armas\_utilizadas}=\text{Explosivos\_Bombas\_Dinamite} \implies \text{vitimas\_fatais}=A$  (conf =0,98)
- b)  $\text{tipo\_ataque}=\text{Assalto\_Armado} \implies \text{armas\_utilizadas}=\text{Armas\_Fogo}$  (conf = 0,96)
- c)  $\text{armas\_utilizadas}=\text{Explosivos\_Bombas\_Dinamite} \implies$

vitimas\_fatais=A (conf = 0,95)

- d) feridos=A tipo\_ataque=Assalto\_Armado  
 ==>armas\_utilizadas=Armas\_Fogo (conf = 0,94)

O quadro 6 apresenta as regras destacadas com as suas devidas conclusões:

Quadro 6 - Regras e conclusões através do Apriori (Brasil)

Regra	Conclusão
feridos=A tipo_ataque = Bombardeio_Explosao ==> vitimas_fatais=A	Quando o número de feridos foi zero e o tipo de ataque foi bombardeio e/ou explosão, o número de óbitos também foi zero
tipo_ataque=Assalto_Armado ==> armas_utilizadas=Armas_Fogo	Quando o tipo de ataque foi assalto armado foram utilizadas armas de fogo
armas_utilizadas=Explosivos_Bombas_Dinamite ==> vitimas_fatais=A (conf = 0,95)	Quando foram utilizados explosivos, bombas e dinamites o número de vítimas fatais foi zero
vitimas_fatais=B tipo_ataque=Assalto_Armado ==> armas_utilizadas=Armas_Fogo	Quando o número de vítimas fatais foi maior do que zero e o tipo de ataque foi assalto armado foram utilizadas armas de fogo

FONTE: Dados da pesquisa utilizando o software Weka adaptado pela autora (2017)

Através dos resultados apresentados nota-se certa semelhança em relação ao tipo de ataque e armas de utilizadas e tanto nas regras destacadas quando a amostra utilizada compreende a América do Sul como à amostra que representa os ataques registrados no Brasil.

#### 4.2.3 K-Means

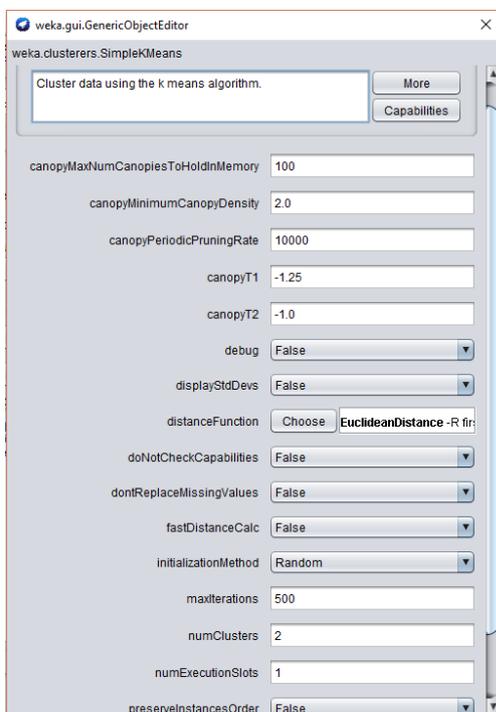
O K-Means é uma heurística de agrupamento não hierárquico que busca minimizar a distância dos elementos a um conjunto de k centros dado por  $\chi=\{x_1,x_2,\dots,x_k\}$  de forma iterativa. O algoritmo depende de um parâmetro (k=número de clusters) definido de forma *ad hoc* pelo usuário. Isto costuma ser um problema, tendo em vista que normalmente não se sabe quantos clusters existem a priori.

O algoritmo do K-Means pode ser descrito da seguinte maneira:

1. Escolher k distintos valores para centros dos grupos (possivelmente, de forma aleatória)
2. Associar cada ponto ao centro mais próximo
3. Recalcular o centro de cada grupo
4. Repetir os passos 2-3 até nenhum elemento mudar de grupo. (LINDEN, 2009).

O algoritmo K-Means foi utilizado em seus parâmetros *default* conforme apresentado na figura 12.

Figura 12 - Parâmetros default K-means



FONTE: Dados da pesquisa utilizando o software Weka (2017)

No primeiro experimento o “sucesso” foi utilizado como atributo meta, os dois clusters gerados estão apresentados no quadro 7.

Quadro 7 - Resultado K-means (América do Sul)

Atributo	CLUSTER 0	CLUSTER 1
País	Colômbia	Peru
Cidade	Desconhecido	Lima
Grupo responsável	Shining_Path_(SL)	Desconhecido
Vítimas_fatais	B	A
Feridos	A	A
Alvo	Cidadãos_Privados_Propriedades	Empresas
Tipo de ataque	Assalto_Armado	Bombardeio_Explosão
Armas utilizadas	Armas_fogo	Explosivos_Bombas_Dinamite

FONTE: Dados da pesquisa utilizando o software Weka (2017)

Nota-se que os incidentes onde não houve feridos associados há armas de fogo, assaltos armados e vítimas fatais entre 1 e 50, semelhante ao apresentado pelo algoritmo Apriori. Da mesma forma, nos ataques causados através de bombardeio ou explosão não houve mortos nem feridos.

No entanto este experimento classificou incorretamente 9.220 (49%) instâncias. Por isso o K-Means foi executado novamente utilizando “armas utilizadas” como atributo meta e eliminado o atributo “sucesso” da mineração de dados foram obtidos os mesmos grupos que no experimento anterior, sendo que as instâncias classificadas incorretamente foram reduzidas para 6.896 (37%).

O mapa apresentado na figura 13, elaborado na ferramenta Microsoft Power BI® destaca os pontos onde os resultados apresentados pelo *K-Means* bem como pelo Apriori foram identificados.

Figura 13 - Resultado Apriori e K-means apresentados no mapa



FONTE: Dados da pesquisa utilizando o *Power BI* (2017)

No último experimento realizado, o K-Means, foi utilizado para realizar o agrupar os dados referentes aos ataques ocorridos no Brasil, nenhum dos atributos possui distribuição adequada para a realização da tarefa de agrupamento. Neste sentido, para que não houvesse um grande número de instâncias agrupadas incorretamente o experimento foi realizado sem atributo meta. Foram utilizados três *clusters* para que as instâncias fossem agrupadas da melhor forma mais uniforme (Quadro 8).

Quadro 8 - Resultado K-means utilizando a amostra referente aos incidentes ocorridos no Brasil  
(Continua)

Attribute	Cluster 0	Cluster 1	Cluster 2
Cidade	Rio_de_janeiro	Sao_paulo	Rio_de_janeiro
Alvo	Cidadãos_ Privados_ Propriedades	Empresas	Cidadãos_ Privados_ Propriedades
Tipo_ataque	Assassinato	Tomada_refem_ (sequestro)	Bombardeio_explosao
Armas_utilizadas	Armas_fogo	Armas_fogo	Explosivos_bombas_ Dinamite
Sucesso	1	1	1

	(Conclusão)		
Vitimas_fatais	B	A	A
Feridos	A	A	A

FONTE: Dados da pesquisa utilizando o software Weka (2017)

É possível observar que assim como nos experimentos realizados com os dados referentes aos incidentes ocorridos na América do Sul o resultado apresentado pelo K-Means é semelhante as regras obtidas com o Apriori quando o Cluster 0 onde houve mortos, não houve feridos e o tipo de ataque foi assassinato. Da mesma forma no cluster 2 não houve mortos nem feridos e o tipo de ataque foi “bombardeiro\_explosão”. Nota-se também que “Armas\_Fogo” e o sucesso 1, “Rio\_de\_Janeiro”, feridos “A” e “cidadãos\_privados” prevalecem entre os grupos indicando a prevalência destas instâncias nos incidentes registrados.

## 5 CONSIDERAÇÕES FINAIS

O terrorismo é um problema que exige grande esforço cooperativo por parte dos governantes, como forma de compartilhar informações pertinentes relacionadas aos ataques registrados. Neste sentido a existência de uma base de dados como a GTD é uma das principais fontes a serem consultadas. Apesar disso, há poucas publicações relacionadas a GTD, indicando desconhecimento da existência desta base de dados por parte da comunidade científica.

É importante ressaltar que para que as informações disponíveis na GTD tenham alguma utilidade para a sociedade em geral, faz-se necessário que estas sejam tratadas de forma adequada. Cabe aqui destacar a importância da escolha do algoritmo, considerando em mineração de dado, a obtenção de resultados precisos depende da utilização do algoritmo adequado.

O fato de o terrorismo não possuir uma tipificação universal, devido à abrangência de definições existentes, pode gerar inconsistência nas informações divulgadas, bem como em análises criminais já realizadas. No caso do Brasil tal fato se torna ainda mais preocupante devido à ausência de uma lei que se preocupa especificamente de ataques terroristas até março de 2016. Neste contexto, também é preocupante o veto de alguns artigos que tratam como ato terrorista atos cibernéticos, considerando que estes são os mais comuns diante das frequentes inovações tecnológicas atuais.

### 5.1 VERIFICAÇÃO DOS OBJETIVOS PROPOSTOS

Para realizar uma análise criminal dos incidentes terroristas ocorridos na América do Sul registrados na *Global Terrorism Database* (GTD), foi necessário alcançar três objetivos específicos, sendo eles:

- a) escolher o algoritmo adequado para realizar uma efetiva mineração nos dados presentes na base: com base nos resultados dos experimentos realizados apresentados na seção 4 é válido afirmar que ambos os algoritmos utilizados foram adequados as tarefas de mineração de dados propostas neste estudo. Em relação ao J48, ainda que não tenha apresentado satisfatórios quando utilizado o “sucesso” como atributo meta, devido ao alto erro relativo absoluto e baixa estatística de kappa. O

algoritmo obteve bons resultados quanto aos demais experimentos e classificou uma boa quantidade de instâncias corretamente, considerando uma escala de 0 a 100 %. O Apriori apresentou bons resultados quando eliminados os atributos que apresentam distribuição desigual o que já era esperado em tarefas de associação. O K-means confirma as informações obtidas através do Apriori possibilitando uma maior confiança nos resultados apresentados. Neste sentido pode-se afirmar que este objetivo foi atingido.

- b) identificar padrões e relações entre os atributos da base de dados: os resultados apresentados através dos algoritmos J48, Apriori bem como pelo K-means indicam fortes relações entre o número de óbitos e feridos, o tipo de ataque e o tipo de arma utilizada. Logo este objetivo também foi alcançado.
- c) comparar os resultados obtidos por meio deste estudo, com outras análises já realizadas na GTD: como mencionado na seção 1.3 que aborda a justificativa da realização desta pesquisa foram poucos os estudos realizados na GTD. Ainda que na fundamentação teórica cite uma análise estatística já realizada que pode ser associada a análise apresentada, não foram encontradas publicações que abordem as tarefas de mineração de dados realizadas da mesma forma que foram aqui abordadas. Desta forma este objetivo foi atingido uma vez que não foram encontrados outros estudos com abordagem semelhante a pesquisa apresentada.

## 5.2 CONTRIBUIÇÕES DA PESQUISA

Devido à carência de publicações realizados pela comunidade científica sobre o tema proposto, o presente estudo tem seu valor, tendo em vista o contexto social em que está inserido e as frequentes oscilações em relação ao número de ataques registrados apresentadas. Desta forma, esta pesquisa justifica os demais estudos tanto científicos quanto práticos de cunho social associadas ao terrorismo que vêm sendo realizados.

Atentando-se aos resultados obtidos tanto na análise estatística quanto nas tarefas de mineração é possível perceber quais aspectos do incidente terrorista possui maior relevância e devem ser observados mais atentamente no caso da análise criminal de um incidente específico. Neste sentido os padrões encontrados através da mineração de dados pode ser utilizado a fim de minimizar a frequência de incidentes terroristas ocorridos.

Por fim, destaca-se a importância do conhecimento prévio de gestão da informação para a realização de estudos como o proposto, pois todo o material disponível pelo START serão apenas dados se não tratados ou, se tratados de maneira inadequada podem gerar informações que não condizem com a verdade e conseqüentemente não irão contribuir para geração de conhecimento.

### 5.3 TRABALHOS FUTUROS

Como são poucas as análises e tarefas de mineração de dados realizadas através da GTD, são muitas as possibilidades de estudos posteriores a este.

Sugere-se principalmente, a realização de estudos que considerem os demais atributos que compõe a GTD, considerando que esta pesquisa se concentra em apenas nove dos cento e trinta e dois atributos pertencentes a base. Neste sentido cabe uma análise que tenha por objetivo identificar quais dos demais atributos tenham maior relação com os atributos utilizados neste estudo considerando que estes são os atributos principais segundo o START.

Sugere-se também a utilização de outros algoritmos para as tarefas de mineração. No caso de pesquisas relacionadas a países com poucos registros de incidentes como no Brasil, os algoritmos que utilizam redes neurais mencionados na fundamentação teórica podem apresentar bons resultados, por exemplo.

Sugere-se, por fim, a realização de análises estatísticas ainda mais detalhadas para que estas possam ser utilizadas por especialistas a fim de servir como auxílio na prevenção de ataques terroristas que possam vir a ocorrer.

## REFERÊNCIAS

ALCÂNTARA, Priscila Drozdek de. Terrorismo: uma abordagem conceitual. In: II workshop de pesquisa em relações internacionais da ufpr, 2., 2015, Curitiba. **Anais....** Curitiba, 2015. p. 1 - 26. Disponível em: <[http://www.humanas.ufpr.br/portal/nepri/files/2012/04/Terrorismo\\_Uma-abordagem-conceitual.pdf](http://www.humanas.ufpr.br/portal/nepri/files/2012/04/Terrorismo_Uma-abordagem-conceitual.pdf)>. Acesso em: 05 mar. 2016.

BASGALUPP, Márcio Porto. **LEGAL-Tree: um algoritmo genético multi-objetivo para indução de árvores de decisão.** 2010. 116 f. Tese (Doutorado) - Curso de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2010. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-12052010-165344/pt-br.php>>. Acesso em: 27 out. 2012.

BENITES, Afonso. Sob pressão internacional, Câmara aprova lei que tipifica o terrorismo. **El País**. Brasília, p. 1-1. 25 fev. 2016. Disponível em: <[http://brasil.elpais.com/brasil/2016/02/24/politica/1456351659\\_569702.html](http://brasil.elpais.com/brasil/2016/02/24/politica/1456351659_569702.html)>. Acesso em: 06 set. 2016.

BRASIL. Lei nº 13260, de 16 de março de 2016. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2015-2018/2016/Lei/L13260.htm](http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Lei/L13260.htm)>. Acesso em: 02 ago. 2016.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de dados: conceitos, tarefas, métodos e ferramentas.** Goiás, 2009. 29 p. Disponível em: <[http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-09.pdf](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf)>. Acesso em: 07 set. 2016.

CARDON, André; MÜLLER, Daniel Nehme. **Introdução as redes neurais artificiais.** 1994. 21 f. Monografia (Especialização) - Curso de Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1994. Disponível em: <[http://www.inf.ufrgs.br/~danielnm/docs/intro\\_rna.pdf](http://www.inf.ufrgs.br/~danielnm/docs/intro_rna.pdf)>. Acesso em: 02 nov. 2016.

CARVALHO, Deborah Ribeiro. **Árvore de decisão / algoritmo genético para tratar o problema de pequenos disjuntos em classificação de dados.** 2005. 173 f. Tese (Doutorado) - Curso de Engenharia Civil, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005. Disponível em: <[http://www.ipardes.gov.br/biblioteca/docs/tese\\_deborah\\_carvalho.pdf](http://www.ipardes.gov.br/biblioteca/docs/tese_deborah_carvalho.pdf)>. Acesso em: 08 out. 2016

CASTANHEIRA, Luciana Gomes. **Aplicação de técnicas de mineração de dados em problemas de classificação de padrões.** 2008. 95 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: <<http://www.ppgee.ufmg.br/defesas/349M.PDF>>. Acesso em: 07 out. 2016.

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016. 559 p.

CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de dados: funcionalidades, técnicas e abordagens**. 2002. Disponível em: <[ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02\\_10\\_cortes.pdf](ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf)>. Acesso em: 10 out. 2016.No prelo

DISSENHA, Rui Carlo. O terrorismo e o direito penal genocida. **Revista da Defensoria Pública da União**, Brasília, v. 1, n. 6, p.160-193, dez. 2013. Disponível em: <<http://revistadadpu.dpu.def.br/index.php/dpu/issue/view/5>>. Acesso em: 10 maio 2017.

DOSUALDO, Daniel Gomes. **Investigação de regressão no processo de mineração de dados**. 2003. 177 f. Dissertação (Mestrado) - Curso de Ciências de Computação e Matemática Computacional, Universidade de São Paulo, São Carlos, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-12112014-101732/pt-br.php+>>. Acesso em: 20 out. 2016.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From *data mining* to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996. Disponível em: <<http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>>. Acesso em: 08 mai. 2016.

FEITOSA, Valdiná Alves; CARVALHO, Carlos Vinícius Delatorres G. de. Contribuição da Inteligência Policial para Análise Criminal. **Conjuntura Econômica Goiânia**. Goiânia, p. 9-13. dez. 2013.

FILHO FERREIRA, Pedro. Introdução ao planejamento e análise estatística de experimentos – c. Universidade Federal de São Carlos. Notas de aula.2009.

FILIPE, Alécio. **Análise criminal: perfil do analista nos EUA**. 2007. 92 f. Monografia (Especialização) - Curso de Gestão de Segurança Pública e Defesa Social, União Pioneira de Integração Social, Brasília, 2007. Disponível em: <[http://www.justica.gov.br/central-de-conteudo/seguranca-publica/monografias/analise\\_criminal\\_perfil\\_analista\\_eua.pdf](http://www.justica.gov.br/central-de-conteudo/seguranca-publica/monografias/analise_criminal_perfil_analista_eua.pdf)>. Acesso em: 08 set. 2016.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2002.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GOTTLIEB, Steven. **Crime Analysis**. California: Alpha Publishing, 1998

GOLDSCHMIDT, Ronaldo; BEZERRA, Eduardo. **Exemplos de aplicações de data mining no mercado brasileiro**. 2016. Disponível em: <<http://computerworld.com.br/exemplos-de-aplicacoes-de-data-mining-no-mercado-brasileiro>>. Acesso em: 07 out. 2016.No Prelo

HALMENSCHLAGER, Carine. **Um algoritmo para indução de árvores e regras de decisão**. 2002. 112 f. Dissertação (Mestrado) - Curso de Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. Disponível em: <<http://hdl.handle.net/10183/2755>>. Acesso em: 28 out. 2016

LASMAR, Jorge Mascarenhas. A legislação brasileira de combate e prevenção do terrorismo quatorze anos após 11 de Setembro: limites, falhas e reflexões para o futuro. **Revista de Sociologia e Política**, Curitiba, v. 23, n. 53, p.47-70, mar. 2015. Disponível em: <<http://www.scielo.br/pdf/rsocp/v23n53/0104-4478-rsocp-23-53-0047.pdf>>. Acesso em: 01 ago. 2016.

LINDEN, Ricardo. Técnicas de Agrupamento. Revista de Sistemas de Informação da Fsma, Visconde de Araújo, v. 1, n. 4, p.18-36, jul. 2009. Disponível em: <[http://www.fsma.edu.br/si/edicao4/FSMA\\_SI\\_2009\\_2\\_Tutorial.pdf](http://www.fsma.edu.br/si/edicao4/FSMA_SI_2009_2_Tutorial.pdf)>. Acesso em: 29 maio 2017.

MAGALHÃES, Luiz Carlos. Análise criminal e mapeamento da criminalidade – GIS. **Âmbito Jurídico**, Rio Grande, v. 9, n. 50, *on-line*, 29 fev. 2008. Disponível em: <[http://www.ambito-juridico.com.br/site/?n\\_link=revista\\_artigos\\_leitura&artigo\\_id=4405&revista\\_caderno=3](http://www.ambito-juridico.com.br/site/?n_link=revista_artigos_leitura&artigo_id=4405&revista_caderno=3)>. Acesso em: 10 mar. 2017.

MARTINEZ, Luís; CASAL, Ricardo; JANEIRO, João. **Sistemas de apoio à decisão clínica**. Porto: Faculdade de Medicina do Porto, 2009. 16 p.

MINAYO, Maria. Cecília de Souza; DESLANDES, Sueli Ferreira; NETO CRUZ, Otávio; GOMES, Romeu. Pesquisa social: teoria, método e criatividade. Petrópolis, RJ: Vozes, 2002.

MOREIRA, Benjamin. **Sistema para análise de dados em mastologia**. 2004. 79 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade do Vale do Itajaí, Itajaí, 2004. Disponível em: <[http://siaibib01.univali.br/pdf/Benjamin\\_Moreira.pdf](http://siaibib01.univali.br/pdf/Benjamin_Moreira.pdf)>. Acesso em: 09 out. 2016.

NATIONAL CONSORTIUM FOR THE STUDY OF TERRORISM AND RESPONSES TO TERRORISM (START). **Global Terrorism Database (GTD)**. Disponível em: <<http://www.start-dev.umd.edu/gtd/>>. Acesso em: 18 AGO. 2016.

NETTO, Sérgio de Oliveira. **A nova lei sobre o terrorismo**. 2016. Disponível em: <<http://www.defesanet.com.br/terror/noticia/21886/Analise---A-nova-lei-sobre-o-terrorismo-do-Brasil/>>. Acesso em: 07 set. 2016.

ONU. Resolução nº 49/60, de 17 de fevereiro de 1995. **Measures to eliminate international terrorism**. Disponível em: <[http://www.un.org/ga/search/view\\_doc.asp?symbol=A/RES/49/60](http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/49/60)>. Acesso em: 09 ago. 2016.

PEIXOTO, Betânia Toti. **Análise criminal**. Brasília: Senasp/mj, 2008. Notas de aula. Disponível em: <[http://www.justocantins.com.br/files/publicacao/AnaliseCriminal\\_completo.pdf](http://www.justocantins.com.br/files/publicacao/AnaliseCriminal_completo.pdf)>. Acesso em: 15 ago. 2016.

RODRIGUES, Dallbian Guimarães. **Estudo comparado entre criminologia ambiental e tradicional: Metodologia de Produção de Conhecimento Científico**. Goiânia: 2012. p. 22. Acesso em 15 ago 2016

SANTOS, Isabelle Dias Carneiro; SANTOS, Isabelle Dias Carneiro. O terrorismo internacional no sul da América do Sul e a segurança internacional. **Revista do Instituto do Direito Brasileiro**, Lisboa, v. 1, n. 6, p.5675-5703, jun. 2013. Disponível em: <[http://www.cidp.pt/publicacoes/revistas/ridb/2013/06/2013\\_06\\_05675\\_05703.pdf](http://www.cidp.pt/publicacoes/revistas/ridb/2013/06/2013_06_05675_05703.pdf)>. Acesso em: 17 maio 2017.

SCHMID, Alex P. **The Routledge handbook of terrorism research**. 1. ed. Estados Unidos: Routledge Taylor & Francis Group, 2011.

SFERRA, Heloisa Helena; CORRÊA, Ângela M. C. Jorge. Conceitos e aplicações de data mining. **Revista de Ciência & Tecnologia**, Piracicaba, v. 11, n. 22, p.19-34, mar. 2003. Disponível em: <[http://www.unifra.br/professores/EDUARDO/Artigo 8.pdf](http://www.unifra.br/professores/EDUARDO/Artigo%208.pdf)>. Acesso em: 18 set. 2016.

VASCONCELOS, Livia Maria Rocha de; CARVALHO, Cedric Luiz de. **Aplicação de Regras de Associação para Mineração de Dados na Web**. Goiás, 2004. 20 p. Disponível em: <[http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_004-04.pdf](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-04.pdf)>. Acesso em: 07 maio 2017

VASCONCELLOS, Fábio. Ataques terroristas no mundo desde 1970. O Globo. Rio de Janeiro, nov. 2015. *Online*. Disponível em: <<http://infograficos.oglobo.globo.com/mundo/ataques-terroristas-no-mundo-desde-1970.html>>. Acesso em: 08 maio 2017

### APÊNDICE A – Atributos GTD

Atributo	Descrição	Tipo
Eventid	ID do registro, segue um sistema de identificação de 12 dígitos, sendo que os 8 primeiros números representam a data "AAAAMMDD", e últimos 4 números é o número do processo sequencial para o dia dado (0001, 0002 etc)	Numérico
Iyear	Ano em que ocorreu o incidente	Data
Imonth	Número do mês em que ocorreu o incidente	Numérico
Iday	Dia numérico do mês em que ocorreu o incidente.	Numérico
Approxdate	Data aproximada do incidente (quando a data exata do incidente não é conhecida ou ainda não está claro)	Nominal
Extended	1 = "Sim" A duração de um incidente prolongado superior a 24 horas. 0 = "Não" A duração de um incidente alargado menos do que 24 horas.	Categórico
Resolution	Registra a data em que o incidente foi resolvido (reféns libertados pelos perpetradores; reféns mortos, resgate bem sucedido, etc.)	Data
Country	Categorização do atributo "country_txt" (escala de 4 - 1004)	Categórico
Country_txt	País ou localização onde ocorreu o incidente (inclui estados não-independentes, dependências e territórios).	Nominal
Region	Categorização do atributo "region_txt" (escala de 1 - 12)	Categórico
Region_txt	Região em que ocorreu o incidente	Nominal
Provstate	Província / Região Administrativa / Estado	Nominal
City	Cidade / Village	Nominal
Latitude	Latitude (com base nas normas WGS1984) da cidade	Numérico
Longitude	Longitude (com base nas normas WGS1984) da cidade	Numérico
Specificity	Identifica a resolução geoespacial dos campos de latitude e longitude	Numérico
Vicinity	1 = "Sim". O incidente ocorreu nas imediações da cidade 0 = "Não" O incidente na própria cidade.	Categórico
Location	Especifica informações adicionais sobre o local do incidente	Nominal

Summary	Resumo da narrativa do incidente, notando a "quando, onde, quem, o quê, como e porquê."	Nominal
Crit1, Crit2, Crit3 *	Registra quais dos critérios de inclusão (além dos critérios necessários) Crit 1 = político, econômico, religioso ou meta social Crit 2 = a intenção para coagir, intimidar ou publicidade a maior audiência (s) Crit 3 = fora direito internacional humanitário	Categórico
Doubtterr	1 = "Sim". Há dúvida sobre se o incidente é um ato de terrorismo. 0 = "Não". Não há essencialmente nenhuma dúvida sobre se o incidente é um ato de terrorismo.	Categórico
Alternative	1 = Insurgência / Ação Guerrilha 2 = Outro tipo de crime 3 = Conflito Inter / Intra-Grupo 4 = Falta de intencionalidade 5 = Atores Estaduais (sistematicamente codificados pós-2012)	Categórico
Alternative_txt	Identifica a categorização mais provável dos incidentes diferentes do terrorismo	Nominal
Multiple	1 = "Sim" O ataque é parte de um incidente múltipla. 0 = "não" O ataque não é parte de um incidente múltipla.	Categórico
Success	1 = "Sim" O incidente foi bem sucedido. 0 = "não" O incidente não foi bem sucedido.	Categórico
Suicide	1 = "Sim". O incidente foi um ataque suicida. 0 = "não" Não há indicação de que o incidente foi um ataque suicida.	Categórico
Attacktype1, Attacktype2, Attacktype3	Categorização do atributo "Attacktype1_txt" (escala de 1 - 9), até 3 tipos de tipo de ataque podem ser registrados para cada incidente.	Categórico
Attacktype1_txt, Attacktype2_txt, Attacktype3_txt	Método geral de ataque. Até três tipos de ataque podem ser registrado para cada incidente.	Nominal
Targtype1, Targtype2, Targtype3	Categorização do atributo "Targtype1_txt" (escala de 1 - 22)	Categórico
Targtype1_txt, Targtype2_txt, Targtype3_txt	Descreve o alvo do ataque	Nominal
Targsubtype1, Targsubtype2, Targsubtype3	Categorização do atributo "Targtype1_txt" (escala de 1 - 92)	Categórico
Targsubtype1_txt, Targsubtype2_txt,	Descreve a categoria de destino mais específica e fornece o próximo nível de	Nominal

Targsubtype3_txt	designação para cada tipo de alvo	
Corp1, Corp2, Corp3	Nome da entidade corporativa ou agência do governo que foi alvo.	Nominal
Target1, Target2, Target3	Pessoa específica, construção, instalação, etc., que foi alvo e / ou vitimados	Nominal
Natlty1, Natlty2, Natlty3	Categorização do atributo "Natlty1_txt" (obedece a mesma classificação do atributo "country")	Catagórico
Natlty1_txt, Natlty2_txt, Natlty3_txt	Nacionalidade do alvo que foi atacado (não é necessariamente o mesmo que o país em que ocorreu o incidente).	Nominal
Gname, Gname2, Gname3	Grupo que realizou o ataque (até 3 por ataque)	Nominal
Gsubname, Gsubname2, Gsubname3	Qualificadores adicionais ou detalhes sobre o nome do grupo que realizou ataque	Nominal
Ingroup, Ingroup2, Ingroup3		
Motive	Relata o motivo do ataque quando o mesmo é explícito	Nominal
Guncertain1, Guncertain2, Guncertain3	1 = "Sim" A atribuição (s) autor pelo incidente são suspeitos. 0 = "não" A atribuição (s) autor para o incidente não são suspeitos.	Catagórico
Nperps	Número total de terroristas participantes do incidente	Numérico
Nperpcap	Número de autores levados em custódia	Numérico
Claimed, Claim2, Claim3	1 = "Sim" Um grupo ou indivíduo reivindicou a responsabilidade pelo ataque; 0 = "não" Nenhuma reivindicação de responsabilidade foi feita	Catagórico
Claimmod, Claimmode2, Claimmode3	Modo utilizado pelos requerentes para reivindicar a responsabilidade. 1 = Carta 2 = Call (pós-incidente) 3 = Call (pré-incidente) 4 = <i>E-mail</i> 5 = Nota à esquerda na cena 6 = Vídeo 7 = postados <i>site</i> , <i>blog</i> , mídias sociais 8 = Pedido pessoal 9 = Outro 10 = Desconhecido	Catagórico

Claimmode_txt, Claimmode2_txt Claimmode3_txt,	Modo utilizado pelos requerentes para reivindicar a responsabilidade.	Nominal
Compclaim	1 = "Sim" Existem reivindicações concorrentes de responsabilidade pelo ataque. 0 = "não" Não há reivindicações conflitantes de responsabilidade pelo ataque. -9 = "Desconhecido" Não há indicação de reivindicações concorrentes.	Categórico
Weaptype1, Weaptype2, Weaptype3, Weaptype4	Categorização do atributo "Weaptype1_txt" (escala de 1 - 13)	Categórico
Weaptype1_txt, Weaptype2_txt, Weaptype3_txt; Weaptype4_txt	Tipo de arma utilizada no ataque (até 4 por incidente)	Nominal
Weapsubtype1, Weapsubtype2, Weapsubtype3, Weapsubtype4	Categorização dos atributo "Weaptype1_txt", "Weapsubtype2_txt", "Weapsubtype3_txt" e "Weapsubtype4" (escala de 1 - 27)	Categórico
Weapsubtype1_txt Weapsubtype2_txt Weapsubtype3_txt Weapsubtype4_txt	Específica a maioria dos tipos de armas identificadas nos campos "Weaptype1_txt", "Weaptype2_txt", "Weaptype3_txt" e Weaptype4_txt" respectivamente	Nominal
Weapdetail	Qualquer informação pertinente sobre o tipo de arma (s) utilizado no incidente	Nominal
Nkill	Número total de mortes	Numérico
Nkillus	Número de norte-americanos que morreram	Numérico
Nkillter	Número de mortes de perpetradores	Numérico
Nwound	Número total de feridos	Numérico
Nwoundus	Número de norte-americanos feridos	Numérico
Nwoundte	Número de perpetradores feridos	Numérico
Property	1 = "Sim" Houve danos materiais. 0 = "não" Não houve danos materiais. -9 = "Desconhecido" Não se sabe houve danos materiais.	Categórico
Propextent	1 = catastrófica (provável > \$ 1 bilhão) 2 = Major (probabilidade > \$ 1 milhão, mas < \$ 1 bilhão) 3 = Menor (provavelmente < \$ 1 milhão) 4 = Desconhecido	Categórico
Propextent_txt	Descrição do atributo "Propextent"	Nominal
Propvalue	Valor de danos à propriedade (em U\$)	Numérico

Propcomment	Medidas não-monetários ou imprecisas dos danos a propriedade	Nominal
Ishostkid	1 = "Sim" As vítimas foram tomados como reféns ou sequestrados. 0 = "não" As vítimas não foram tomados como reféns ou sequestradas. -9 = "Desconhecido" Não se sabe se as vítimas foram tomados como reféns ou sequestrados.	Categórico
Nhostkid	Número total de reféns / vítimas de sequestro	Numérico
Nhostkidus	Número total de reféns / vítimas de sequestro americanos	Numérico
Nhours	Duração do incidente	Numérico
Ndays	Duração do sequestro / incidente refém que passaram de 24 horas	Numérico
Divert	País que os seqüestradores desviaram um veículo /país que as vítimas do seqüestro foram mantidas	Nominal
Kidhijcountry	País em que o incidente foi resolvido ou terminado	Nominal
Ransom	1 = "Sim" O incidente envolveu uma demanda de resgate monetária. 0 = "não" O incidente não envolveu uma demanda de resgate monetária. -9 = "Desconhecido" Não se sabe se o incidente envolveu uma demanda de resgate monetária.	Categórico
Ransomamt	Valor (em U\$) solicitado para o resgate	Numérico
Ransomamtus	Valor (em U\$) de recursos norte-americanos solicitados para o resgate	Numérico
Ransompaid	Valor total do resgate pago (U\$)	Numérico
Ransompaidus	Valor total do resgate pago (U\$) por meio de recursos norte-americanos	
Ransomnote	Detalhes específicos relativos a um resgate	Nominal
Hostkidoutcome	1 = tentativa de resgate 2 = Reféns liberados pelos perpetradores 3 = Os reféns escaparam (não durante a tentativa de resgate) 4 = Reféns mortos (não durante a tentativa de resgate) 5 = Salvamento bem sucedido 6 = Combinação 7 = Desconhecido	Categórico
Hostkidoutcome_txt	Descrição do atributo "Hostkidoutcome"	Nominal
Nreleased	Número de reféns que sobreviveram ao	Numérico

	incidente	
Addnotes	Detalhes adicionais relevantes sobre o ataque	Nominal
Scite1, Scite2, Scite3	Fonte que foi utilizada para compilar informações sobre o incidente	Nominal
Dbsource	Esforço de coleta de dados original na qual cada evento foi registrado.	Nominal
Int_log	1 = "Sim" a nacionalidade do grupo agressor difere do local do ataque 0 = "Não" a nacionalidade do grupo agressor é o mesma do local de ataque. -9 = "Desconhecido" Não se sabe se o ataque foi logisticamente internacional ou nacional; a nacionalidade do grupo agressor é desconhecida.	Categórico
Int_ideo	1 = "Sim" A nacionalidade do grupo agressor difere da nacionalidade do alvo (s) / vítima (s). 0 = "não" Qualquer e todas as nacionalidades do grupo agressor são os mesmos que os nacionalidades do alvo (s) / vítima (s). -9 = "Desconhecido" A nacionalidade do grupo agressor ou da nacionalidade do alvo / vítima é desconhecida	Categórico
Int_misc	1 = "Sim" o local do ataque difere da nacionalidade do alvo / vítima. 0 = "não" o local do ataque é o mesmo que as nacionalidades do alvo/ vítima -9 = "Desconhecido" A nacionalidade do alvo / vítima é desconhecida.	Categórico
Int_any	1 = "Sim" O ataque é internacional (Quando um dos atributos Int_log, Int_ideo, Int_misc é igual a 1) 0 = "Não" O ataque não é internacional - 9 = Não se sabe se o ataque foi internaciona	Categórico
Related	Quando um ataque é parte de um incidente coordenado em várias partes, os IDs GTD dos incidentes relacionados são listados neste campo, separados por vírgulas.	Numérico

**APÊNDICE B – Distribuição de ataques por país (África Subariana)**

País	Frequência	Porcentagem
Somália	2.890	21,5%
Nigéria	2.888	21,5%
África do sul	1.957	14,6%
Sudão	688	5,1%
Quênia	517	3,8%
Burundi	504	3,8%
Angola	491	3,7%
República Democrática do Congo	459	3,4%
Uganda	366	2,7%
Mali	319	2,4%
Moçambique	272	2,0%
República centro-africana	202	1,5%
Camarões	180	1,3%
Ruanda	156	1,2%
Namíbia	151	1,1%
Etiópia	150	1,1%
Senegal	118	0,9%
Níger	116	0,9%
Sudão do Sul	111	0,8%
Zimbábue	98	0,7%
Serra leoa	97	0,7%
Rodésia	83	0,6%
Chade	80	0,6%
Costa do Marfim	60	0,4%
Zâmbia	60	0,4%
Tanzânia	50	0,4%
Togo	48	0,4%
Zaire	48	0,4%
Libéria	33	0,2%
Lesoto	29	0,2%
Madagascar	26	0,2%
Guiné	24	0,2%
Djibouti	21	0,2%
República do Congo	21	0,2%
Gana	18	0,1%
Mauritânia	18	0,1%
Suazilândia	16	0,1%
Botswana	10	0,1%
Eritreia	10	0,1%
Burkina faso	9	0,1%
Guiné-bissau	9	0,1%
Benin	8	0,1%

Comores	5	0,0%
Gabão	4	0,0%
Malawi	4	0,0%
Gambia	3	0,0%
República Popular do Congo	3	0,0%
Guiné equatorial	2	0,0%
Seicheles	2	0,0%
<b>Total</b>	<b>13.434</b>	<b>100,0%</b>

Fonte: A autora (2017)

## APÊNDICE C – Regras geradas a partir do algoritmo *Apriori* (América do Sul)

<p>1. alvo=Empresas armas_utilizadas=Explosivos_Bombas_Dinamite 2128 ==&gt;  tipo_ataque=Bombardeio_Explosao 2105 &lt;conf:(0.99)&gt; lift:(2.06) lev:(0.06) [1084] conv:(46.16)</p>
<p>2. vitimas_fatais=A feridos=A tipo_ataque=Bombardeio_Explosao 6189 ==&gt;  armas_utilizadas=Explosivos_Bombas_Dinamite 6061 &lt;conf:(0.98)&gt; lift:(2.03) lev:(0.17) [3081]  conv:(24.88)</p>
<p>3. alvo=Empresas tipo_ataque=Bombardeio_Explosao 2153 ==&gt;  armas_utilizadas=Explosivos_Bombas_Dinamite 2105 &lt;conf:(0.98)&gt; lift:(2.03) lev:(0.06) [1068]  conv:(22.79)</p>
<p>4. vitimas_fatais=A tipo_ataque=Bombardeio_Explosao 7016 ==&gt;  armas_utilizadas=Explosivos_Bombas_Dinamite 6852 &lt;conf:(0.98)&gt; lift:(2.03) lev:(0.19) [3474]  conv:(22.05)</p>
<p>5. vitimas_fatais=A feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 6209 ==&gt;  tipo_ataque=Bombardeio_Explosao 6061 &lt;conf:(0.98)&gt; lift:(2.04) lev:(0.17) [3084] conv:(21.69)</p>
<p>6. feridos=A tipo_ataque=Bombardeio_Explosao 6523 ==&gt;  armas_utilizadas=Explosivos_Bombas_Dinamite 6349 &lt;conf:(0.97)&gt; lift:(2.02) lev:(0.17) [3209]  conv:(19.33)</p>
<p>7. vitimas_fatais=A armas_utilizadas=Explosivos_Bombas_Dinamite 7057 ==&gt;  tipo_ataque=Bombardeio_Explosao 6852 &lt;conf:(0.97)&gt; lift:(2.03) lev:(0.19) [3468] conv:(17.83)</p>
<p>8. feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 6548 ==&gt;  tipo_ataque=Bombardeio_Explosao 6349 &lt;conf:(0.97)&gt; lift:(2.02) lev:(0.17) [3209] conv:(17.04)</p>
<p>9. tipo_ataque=Bombardeio_Explosao 8931 ==&gt; armas_utilizadas=Explosivos_Bombas_Dinamite  8640 &lt;conf:(0.97)&gt; lift:(2.01) lev:(0.23) [4340] conv:(15.86)</p>
<p>10. armas_utilizadas=Explosivos_Bombas_Dinamite 8967 ==&gt; tipo_ataque=Bombardeio_Explosao  8640 &lt;conf:(0.96)&gt; lift:(2.01) lev:(0.23) [4340] conv:(14.23)</p>
<p>11. feridos=A tipo_ataque=Bombardeio_Explosao armas_utilizadas=Explosivos_Bombas_Dinamite  6349 ==&gt; vitimas_fatais=A 6061 &lt;conf:(0.95)&gt; lift:(1.66) lev:(0.13) [2401] conv:(9.31)</p>
<p>12. vitimas_fatais=B tipo_ataque=Assalto_Armado 2274 ==&gt; armas_utilizadas=Armas_Fogo 2168  &lt;conf:(0.95)&gt; lift:(2.77) lev:(0.07) [1385] conv:(13.94)</p>
<p>13. feridos=A tipo_ataque=Bombardeio_Explosao 6523 ==&gt; vitimas_fatais=A 6189 &lt;conf:(0.95)&gt;  lift:(1.65) lev:(0.13) [2429] conv:(8.25)</p>

14. feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 6548 ==> vitimas_fatais=A 6209 <conf:(0.95)> lift:(1.65) lev:(0.13) [2435] conv:(8.16)
15. feridos=A tipo_ataque=Bombardeio_Explosao 6523 ==> vitimas_fatais=A armas_utilizadas=Explosivos_Bombas_Dinamite 6061 <conf:(0.93)> lift:(2.45) lev:(0.19) [3589] conv:(8.75)
16. feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 6548 ==> vitimas_fatais=A tipo_ataque=Bombardeio_Explosao 6061 <conf:(0.93)> lift:(2.46) lev:(0.19) [3594] conv:(8.36)
17. vitimas_fatais=B tipo_ataque=Assassinato 2295 ==> feridos=A 2099 <conf:(0.91)> lift:(1.24) lev:(0.02) [400] conv:(3.03)
18. feridos=A tipo_ataque=Assassinato 2305 ==> vitimas_fatais=B 2099 <conf:(0.91)> lift:(2.74) lev:(0.07) [1332] conv:(7.43)
19. tipo_ataque=Assassinato armas_utilizadas=Armas_Fogo 2135 ==> vitimas_fatais=B 1941 <conf:(0.91)> lift:(2.73) lev:(0.07) [1230] conv:(7.31)

### APÊNDICE D – Regras geradas a partir do algoritmo *Apriori* (Brasil)

<p>1. feridos=A tipo_ataque=Bombardeio_Explosao 42 ==&gt;  armas_utilizadas=Explosivos_Bombas_Dinamite 42 &lt;conf:(1)&gt; lift:(4.45) lev:(0.12) [32]  conv:(32.56)</p>
<p>2. vitimas_fatais=A feridos=A tipo_ataque=Bombardeio_Explosao 41 ==&gt;  armas_utilizadas=Explosivos_Bombas_Dinamite 41 &lt;conf:(1)&gt; lift:(4.45) lev:(0.12) [31]  conv:(31.79)</p>
<p>3. tipo_ataque=Bombardeio_Explosao 56 ==&gt; armas_utilizadas=Explosivos_Bombas_Dinamite 55  &lt;conf:(0.98)&gt; lift:(4.37) lev:(0.16) [42] conv:(21.71)</p>
<p>4. vitimas_fatais=A tipo_ataque=Bombardeio_Explosao 53 ==&gt;  armas_utilizadas=Explosivos_Bombas_Dinamite 52 &lt;conf:(0.98)&gt; lift:(4.37) lev:(0.15) [40]  conv:(20.54)</p>
<p>5. feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 45 ==&gt; vitimas_fatais=A 44  &lt;conf:(0.98)&gt; lift:(1.58) lev:(0.06) [16] conv:(8.6)</p>
<p>6. feridos=A tipo_ataque=Bombardeio_Explosao 42 ==&gt; vitimas_fatais=A 41 &lt;conf:(0.98)&gt;  lift:(1.58) lev:(0.06) [15] conv:(8.02)</p>
<p>7. feridos=A tipo_ataque=Bombardeio_Explosao armas_utilizadas=Explosivos_Bombas_Dinamite  42 ==&gt; vitimas_fatais=A 41 &lt;conf:(0.98)&gt; lift:(1.58) lev:(0.06) [15] conv:(8.02)</p>
<p>8. feridos=A tipo_ataque=Bombardeio_Explosao 42 ==&gt; vitimas_fatais=A  armas_utilizadas=Explosivos_Bombas_Dinamite 41 &lt;conf:(0.98)&gt; lift:(4.57) lev:(0.12) [32]  conv:(16.52)</p>
<p>9. tipo_ataque=Assalto_Armado 50 ==&gt; armas_utilizadas=Armas_Fogo 48 &lt;conf:(0.96)&gt;  lift:(2.31) lev:(0.1) [27] conv:(9.74)</p>
<p>10. armas_utilizadas=Explosivos_Bombas_Dinamite 60 ==&gt; vitimas_fatais=A 57 &lt;conf:(0.95)&gt;  lift:(1.54) lev:(0.07) [19] conv:(5.73)</p>
<p>11. tipo_ataque=Bombardeio_Explosao 56 ==&gt; vitimas_fatais=A 53 &lt;conf:(0.95)&gt; lift:(1.53)  lev:(0.07) [18] conv:(5.35)</p>
<p>12. tipo_ataque=Bombardeio_Explosao armas_utilizadas=Explosivos_Bombas_Dinamite 55 ==&gt;  vitimas_fatais=A 52 &lt;conf:(0.95)&gt; lift:(1.53) lev:(0.07) [18] conv:(5.25)</p>
<p>13. feridos=A tipo_ataque=Assalto_Armado 34 ==&gt; armas_utilizadas=Armas_Fogo 32  &lt;conf:(0.94)&gt; lift:(2.26) lev:(0.07) [17] conv:(6.62)</p>
<p>14. feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 45 ==&gt;</p>

tipo_ataque=Bombardeio_Explosao 42 <conf:(0.93)> lift:(4.45) lev:(0.12) [32] conv:(8.89)
15. vitimas_fatais=A feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 44 ==> tipo_ataque=Bombardeio_Explosao 41 <conf:(0.93)> lift:(4.44) lev:(0.12) [31] conv:(8.69)
16. tipo_ataque=Bombardeio_Explosao 56 ==> vitimas_fatais=A armas_utilizadas=Explosivos_Bombas_Dinamite 52 <conf:(0.93)> lift:(4.35) lev:(0.15) [40] conv:(8.81)
17. armas_utilizadas=Explosivos_Bombas_Dinamite 60 ==> tipo_ataque=Bombardeio_Explosao 55 <conf:(0.92)> lift:(4.37) lev:(0.16) [42] conv:(7.9)
18. vitimas_fatais=A armas_utilizadas=Explosivos_Bombas_Dinamite 57 ==> tipo_ataque=Bombardeio_Explosao 52 <conf:(0.91)> lift:(4.35) lev:(0.15) [40] conv:(7.51)
19. feridos=A armas_utilizadas=Explosivos_Bombas_Dinamite 45 ==> vitimas_fatais=A tipo_ataque=Bombardeio_Explosao 41 <conf:(0.91)> lift:(4.59) lev:(0.12) [32] conv:(7.21)