

# O Futebol Americano de Minas Gerais na Internet

**Tiago Rodrigues Serra**

Curso de Pós Graduação em Business Intelligence – Pontifícia Universidade Católica de Minas Gerais (PUC-MG) – IEC  
Belo Horizonte – MG – Brasil

contato@tiagoserra.com

Orientador: Prof. Cristiano Carvalho

cristiano.dcc@gmail.com

**Abstract.** *Currently the volume of information generated on the Internet increases exponentially every second. Making the task of measuring this very difficult information with metrics that can guarantee and standardize the results that can be compared in any way. As a problem, that content will not be found in any standard. They are texts, videos, photos, audio, formats are varied. Due to this, the value of this information may be lost or not correctly measured. With this assumption, the challenge was to measure the relevance of social networks for two events, the end of the mining Championship Football 2016 between Get Eagles and Minas Locomotiva, compared to the game between the two major football teams of Belo Horizonte, Atlético and Cruzeiro.*

**Resumo.** *Atualmente o volume de informação gerado na internet aumenta exponencialmente a cada segundo. Isto torna tarefa de mensurar essa informação muito difícil, ensejando o uso de métricas que possam garantir e padronizar os resultados para que possam ser tabulados e comparados. Como dificultador, esse conteúdo não será encontrado dentro de algum padrão. São textos, vídeos, fotos, áudios. Os formatos são variados. Devido a isso, o valor dessa informação pode ser perdido ou não mensurado corretamente. Com este pressuposto, o desafio foi medir a relevância nas mídias sociais para dois eventos, a final do campeonato mineiro de Futebol Americano de 2016 entre Get Eagles e Minas Locomotiva frente ao jogo entre os dois maiores times de futebol Belo Horizonte, Atlético e Cruzeiro.*

## **1. Introdução**

Este trabalho tem como objetivo avaliar a aplicabilidade dos resultados da obtenção, mineração de dados e estudo da relevância do tema proposto; Final do Campeonato Mineiro de Futebol Americano entre Minas Locomotiva e Get Eagles frente ao Clássico do Futebol Mineiro entre Cruzeiro e Atlético. A proposta é comparar o impacto dos dois eventos na internet, em blogs especializados, sites e mídias sociais (Twitter, Facebook). Baseado nas etapas do processo de KDD (Knowledge Discovery in Databases). Processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis. O processo inicia a partir dos dados armazenados em um banco de dados, Fayyad (1996), no qual se divide resumidamente em: dados, informação e conhecimento. Utilizando aplicações que permitam a automatização total ou parcial dessas etapas.

Vale ressaltar que mineração de dados não foi criada com objetivo de substituir as técnicas atuais de análise de dados. Ela utiliza como base para a maioria de seus trabalhos os experimentos da Estatística, Inteligência Artificial, Máquina de Estado e Banco de dados para construir seu modelo, Douglas (2016).

## **2. Desenvolvimento**

Segundo as etapas do KDD (Knowledge Discovery in Databases), seleção, pré-processamento, transformação e ou formatação, mineração dos dados e interpretação dos resultados ilustradas na Figura 1, tem como requisito primordial o conhecimento do campo onde será feita a pesquisa, conhecimento preciso sobre o objetivo desejado. Sobre o termo “não trivial”, fica demonstrada a complexidade na execução e manutenção dos processos de KDD, o termo “interativo” indica a relevância de se ter um elemento controlando o processo, já o termo “iterativo” sugere a possibilidade de repetições em qualquer uma das etapas do processo e finalmente o “conhecimento útil” que é aquele onde o objetivo foi alcançado, trazendo consigo benefícios as aplicações de KDD, Douglas (2016).

## **3. Etapas**

De acordo com Fayyad (1996), o processo de KDD tem início na análise do domínio da aplicação e dos objetivos a serem realizados, sendo este processo dividido em 5 fases: Seleção de dados, Pré-processamento; Transformação, Mineração, Análise e Assimilação de resultados.

Nestas etapas a ordem fica bem determinada, podendo ocorrer a recursividade entre elas de acordo com os resultados dos experimentos.

### 3.1 Seleção

A primeira etapa, a seleção de onde serão coletados os dados para a pesquisa.

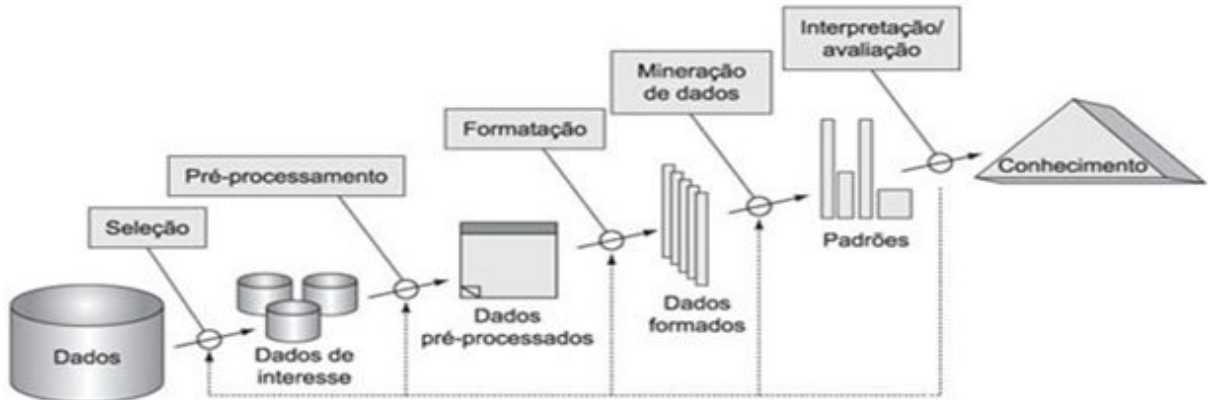


Figura 1. Etapas do KDD (Knowledge Discovery in Databases).

Fonte: Daniel Teofilo, 2016

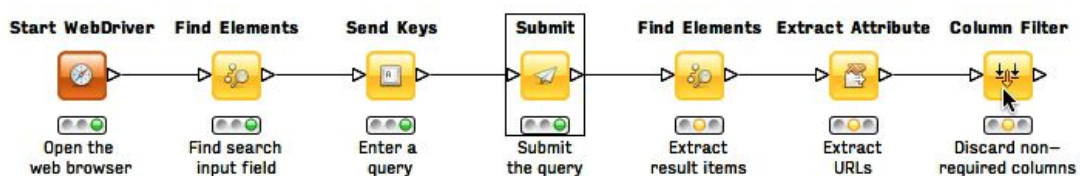
Na seleção definimos os grupos de dados e períodos que serão consumidos para a obtenção de dados realmente relevantes para a pesquisa. Reconhecemos e definimos fontes pertinentes para o tema proposto. Por fim, levamos em consideração que seja encontrado conteúdo para os dois eventos pesquisados, as fontes foram definidas, como observado na tabela seguinte.

Tabela 1. Endereços dos sites e mídias sociais

URL
<a href="http://globo.com/globoesporte">http://globo.com/globoesporte</a>
<a href="http://otempo.com.br">http://otempo.com.br</a>
<a href="http://em.com.br">http://em.com.br</a>
<a href="http://facebook.com">http://facebook.com</a>
<a href="http://twitter.com">http://twitter.com</a>

O período definido para a coleta de dados, foi o seguinte; 5 dias anteriores e posteriores a cada evento do tema, ou seja entre os dias 07/06/2016 e 23/06/2016.

O Knime, foi a plataforma usada para minerar os dados, buscados dentro dos sites e mídias sociais citadas anteriormente. Na busca de dados nos sites relacionados para busca e no Facebook, o Knime foi utilizado juntamente ao Selenium node, que faz um trabalho de Web Browser.



O futebol Americano de Minas Gerais na Internet (Agosto 2016)

Figura 2. Utilizando o Selenium node no Knime para coletar dados na web.

Fonte: Selenium, 2016

Fonte	Tags	Texto	ComFoto	ComVideo
portalotempo	#MinasLocomotiva / #Locomotiva	7	5	0
PortalUAI	#MinasLocomotiva / #Locomotiva	9	9	2
Globoesportecom (MG)	#MinasLocomotiva / #Locomotiva	35	31	6
portalotempo	#GetEagles / #Eagles	5	5	0
PortalUAI	#GetEagles / #Eagles	16	16	6
Globoesportecom (MG)	#GetEagles / #Eagles	41	39	11
portalotempo	#MinasBowl	16	15	0
PortalUAI	#MinasBowl	24	22	2
Globoesportecom (MG)	#MinasBowl	59	50	21

Figura 3. Exemplo de um dos resultados de conteúdos únicos dentro do Facebook para um dos dias do período da pesquisa.

Fonte	Tags	Texto
mineirao	#Atlético #Atletico / #Cruzeiro / #Clássico #Classico	93
Globoesportecom (MG)	#Atlético #Atletico / #Cruzeiro / #Clássico #Classico	387
tvglobominas	#Atlético #Atletico / #Cruzeiro / #Clássico #Classico	219
portaluai	#Atlético #Atletico / #Cruzeiro / #Clássico #Classico	248
otempo	#Atlético #Atletico / #Cruzeiro / #Clássico #Classico	211
Outros	#Atlético #Atletico / #Cruzeiro / #Clássico #Classico	988

Figura 4. Exemplo de resultado para compartilhamentos primários no twitter para um dos dias do período da pesquisa.

Na coleta de dados no Twitter, o Knime também foi a ferramenta utilizada, com o Twitter API node.

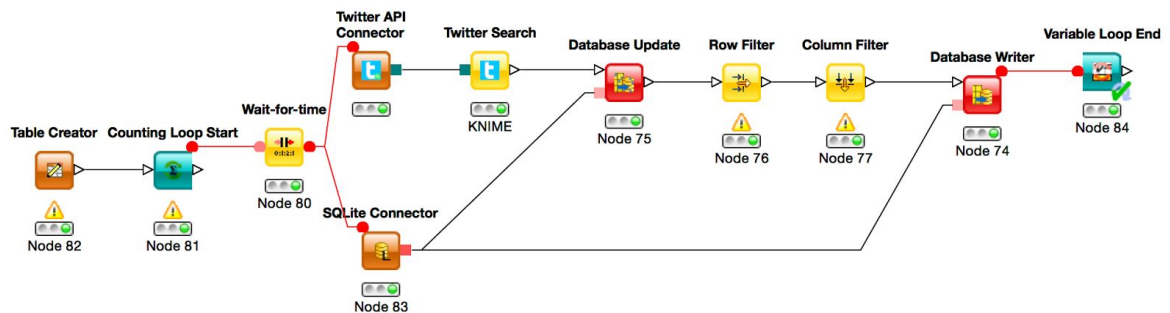


Figura 5. Utilizando o Twitter API node para coletar dados no Twitter.

Fonte: Knime, 2016

Fonte	Tags	Texto
mineirao	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	25
manchanfl	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	31
Globoesportecom (MG)	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	97
100jardas	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	13
theplayoffsbr	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	15
tvglobominas	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	83
NFLPrints	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	11
portaluai	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	31
otempo	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	29
Outros	#MinasLocomotiva / #Locomotiva / #MinasBowl / #GetEagles / #Eagles	132

**Figura 6. Exemplo de resultado para compartilhamentos primários no twitter para um dos dias do período da pesquisa.**

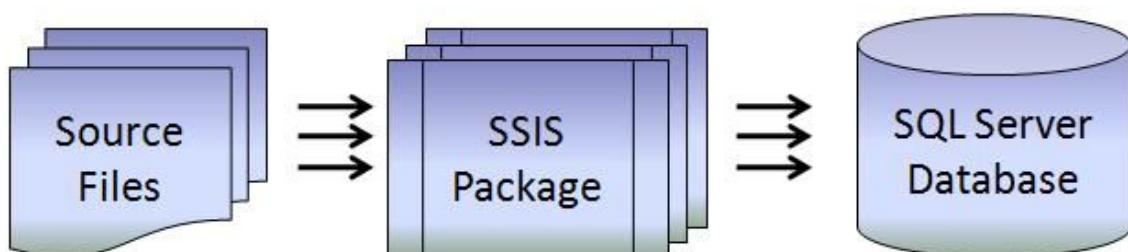
### 3.2 Pré processamento

Para processar os dados, foram utilizadas três ferramentas, Knime (Figura 2), Pentaho Data Integration, uma ferramenta de ETL (Extração, Tratamento e carregamento de dados) que faz parte da plataforma Pentaho.

e Microsoft SQL Server Integration Services, plataforma para a criação de integração de dados em nível corporativo e soluções de transformações de dados. Utilizando o SGBD Sistema de Gerenciamento de Banco de Dados, Microsoft SQL Server 2012 para armazenar desde os dados no pré-processamento, onde os dados são previamente separados em um estágio ainda sem tratamento, apenas como forma de acesso mais rápido, sem precisar buscar nas fontes (Sites, blogs e mídias sociais) diretamente. Como Fayyad (1996) e John (1997), as tarefas da mineração de dados são os tipos de descoberta que se pretende realizar em uma base de dados, isto é, são as informações que se deseja extrair. Para determinar qual tarefa a ser resolvida, deve-se ter um bom conhecimento do domínio da aplicação e saber o tipo de informação que se quer obter.

### 3.3 Formatação de dados

Nesse estágio, o processo de ETL, que destina-se à extração, transformação e carga dos dados de uma ou mais bases de dados de origem para uma ou mais bases de dados de destino, Gama. (2008), é aplicado para tratar, formatar os dados de para a utilização dentro das outras etapas do KDD.



## Figura 7. Fluxo de dados no processo de ETL

Fonte: Microsoft, 2016

### 3.4 Identificando padrões

Das tarefas importantes dentro da mineração de dados, temos a tarefa de associação, que tem como objetivo encontrar a existência de algum padrão do tipo X para Y, onde, a quantificação de X que se implica em Y onde X e Y são conjuntos completamente distintos. Por exemplo, uma pessoa que sempre faz o pedido do produto A, também faz o pedido do produto B. Conforme Berry e Linoff (1997), data mining é a exploração e análise, de forma automática ou semi-automática, de grandes bases de dados com objetivo de descobrir padrões e regras. O objetivo principal do processo de data mining é fornecer as corporações informações que a possibilitem montar melhores estratégias de marketing, vendas, provendo informação para tomadas de decisões referentes a vários setores envolvidos nos dois eventos, melhorando assim os seus negócios.

## 4. Resultados

Com o objetivo de demonstrar a utilização das funcionalidades e utilidades dos processos de Web Mining. Os dados coletados são referentes aos dois eventos, o jogo de futebol entre os times Cruzeiro Esporte Clube e Clube Atlético Mineiro, em 12 de Junho de 2016 no estádio “Mineirão” em Belo Horizonte Evento 1 e ao jogo de futebol americano entre os times Minas Locomotiva e Get Eagles, realizado no mesmo estádio em 18 de Junho de 2016 Evento 2.

### 4.1 Para as mesmas fontes de dados referentes ao Evento 1

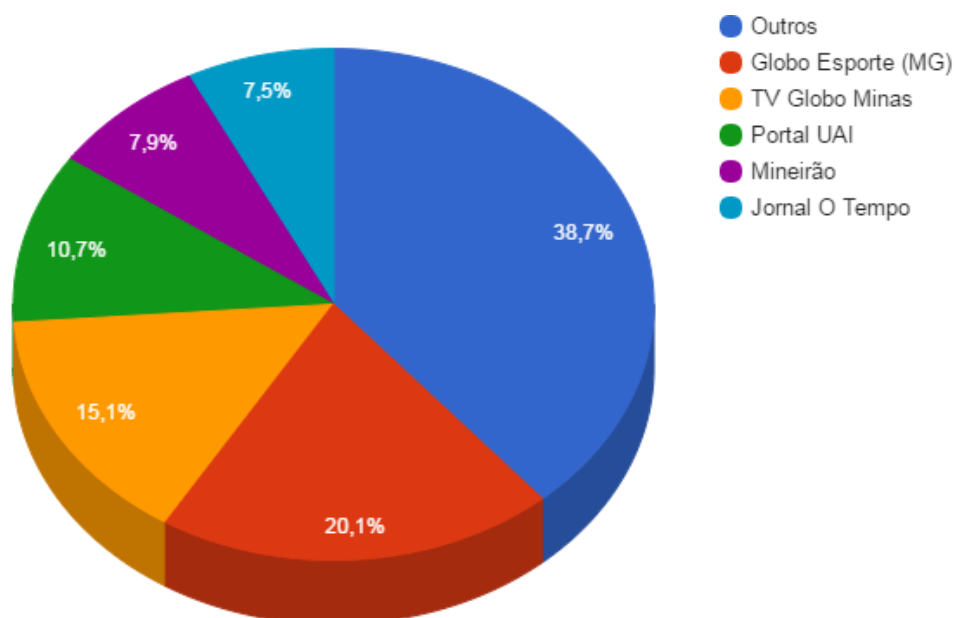
Dados coletados durante os cinco dias antecedentes, o dia do Evento 1 e cinco dias sucedentes ao Evento 1. Com a pesquisa feita nos cinco sites com mais ocorrências e também em seus respectivos perfis dentro das mídias sociais e grupo (Outros) que representa todos os perfis que foram pesquisados mas que tinham um número de ocorrências menor que os cinco primeiros. Com um total de 2637 ocorrências (Veja Tabela 1). Esses dados serão a base para comparação com o Evento 2.

**Tabela 2. Ocorrências por fonte de dados Evento 1**

Fontes	Ocorrências
Outros	1021
Globo Esporte (MG)	529
TV Globo Minas	398
Portal UAI	283
Mineirão	209
Jornal O Tempo	197
Total:	2637

No gráfico seguinte constam os dados destacados com o percentual referente a cada uma das fontes de dados.





**Figura 8. Resultado do levantamento para o Evento 1, com os termos: Atlético, Cruzeiro e Clássico.**

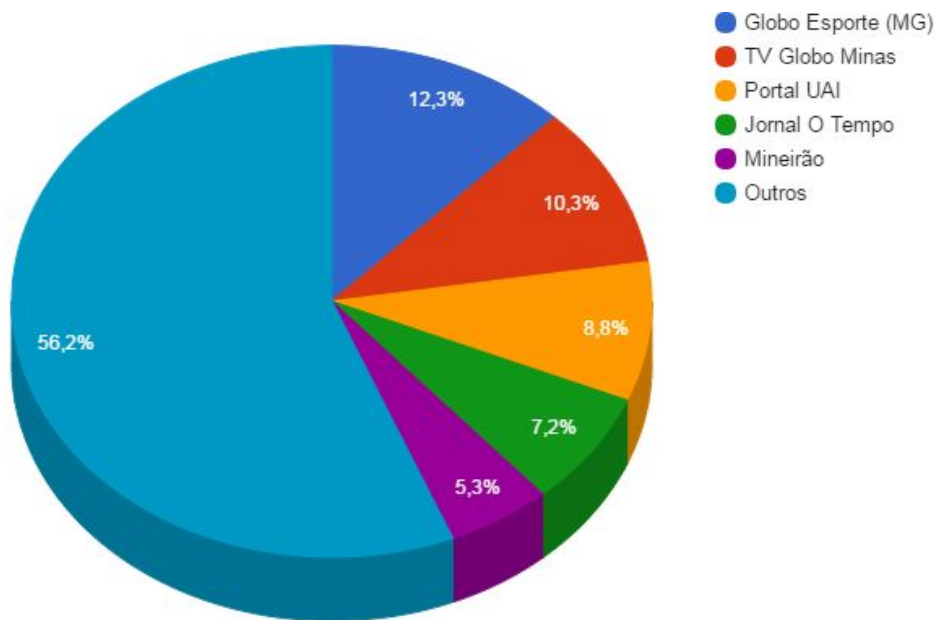
#### 4.2 Para as mesmas fontes de dados referentes ao Evento 2

Dados coletados durante os cinco dias antecedentes, o dia do Evento 1 e cinco dias sucedentes ao Evento 1. Com a pesquisa feita nos cinco sites com mais ocorrências e também em seus respectivos perfis dentro das mídias sociais e grupo (Outros) que representa todos os perfis que foram pesquisados mas que tinham um número de ocorrências menor que os cinco primeiros. Com um total de 1159 ocorrências (Veja Tabela 2) . Como visto na Figura 7, o Evento 2, apresentou resultados bastante interessantes, pois mesmo com o pouco espaço nos meios de comunicação, ainda pouco conhecido por grande parte das pessoas da cidade e região, no grupo (Outros) o resultado foi maior que a metade do Evento 1

**Tabela 3. Ocorrências por fonte de dados Evento 2**

Fontes	Ocorrências
Outros	651
Globo Esporte (MG)	142
TV Globo Minas	119
Portal UAI	102
Jornal O Tempo	84
Mineirão	61
Total:	1159

No gráfico seguinte constam destacados com o percentual referente a cada uma das fontes de dados.



**Figura 9. Resultado do levantamento para o Evento 1, com os termos:: Minas Locomotiva, Locomotiva, Get Eagles, Eagles e Minas Bowl.**

#### **4.3 Para as mesmas fontes de dados referentes aos Evento 1 e Evento 2 .**

Dados coletados durante os cinco dias antecedentes, o dia dos Evento 1 e Evento 2, cinco dias sucedentes aos Evento 1 e Evento 2. Com a pesquisa feita nos cinco sites com mais ocorrências e também em seus respectivos perfis dentro das mídias sociais e grupo (Outros) que representa todos os perfis que foram pesquisados mas que tinham um número de ocorrências menor que os cinco primeiros. Com um total de 6138 ocorrências (Veja Tabela 3) . Na maioria dos casos, as ocorrências do Evento 2 chegam a praticamente  $\frac{1}{3}$  do Evento 1. Novamente, mostra que o Evento 2 se mostrou bastante relevante, devido ao espaço de divulgação ser incomparavelmente menor ao do Evento 1.

**Tabela 4. Ocorrências por fonte de dados (Evento 1 e Evento 2)**



Fonte	Ocorrências
Globo Esporte (MG) (Evento 2)	193
Globo Esporte (MG) (Evento 1)	802
TV Globo Minas (Evento 2)	134
TV Globo Minas (Evento 1)	419
Portal UAI (Evento 2)	102
Portal UAI (Evento 1)	327
Jornal O Tempo (Evento 2)	89
Jornal O Tempo (Evento 1)	214
Mineirão (Evento 2)	74
Mineirão (Evento 1)	208
Outros (Evento 2)	714
Outros (Evento 1)	2972
Total	6318

No gráfico seguinte consta os dados destacados com o percentual referente a cada uma das fontes de dados, para os dois eventos. Mostrando a força entre eles na internet.

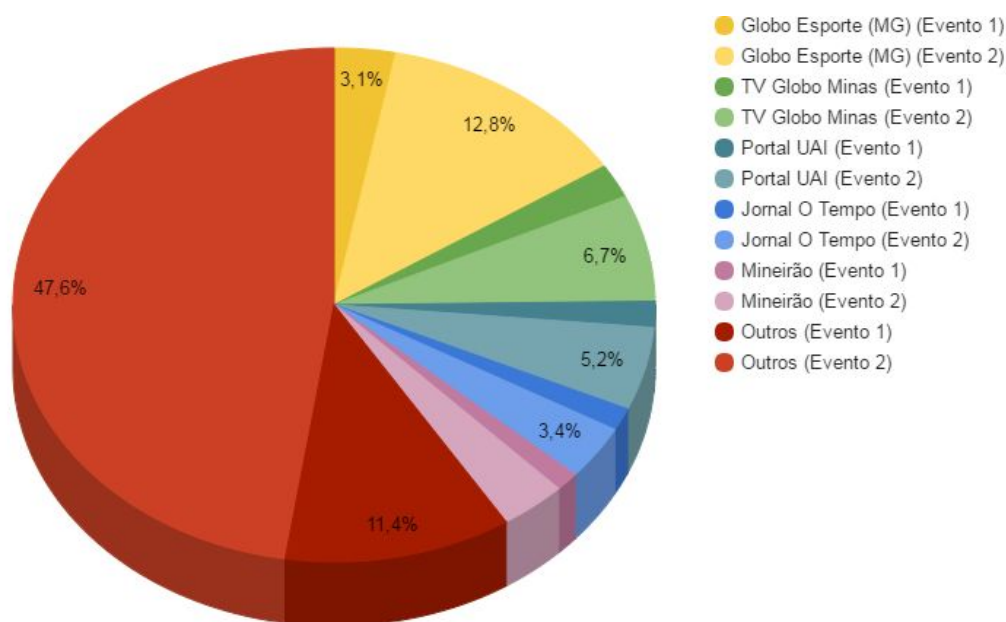


Figura 10. Resultado do levantamento para o Evento 1 e Evento 2, com os termos: Minas Locomotiva, Locomotiva, Get Eagles, Eagles, Minas Bowl, Atlético, Cruzeiro, clássico.

#### 4.5 Dados separados por grupo de faixa etária referentes ao Evento 1

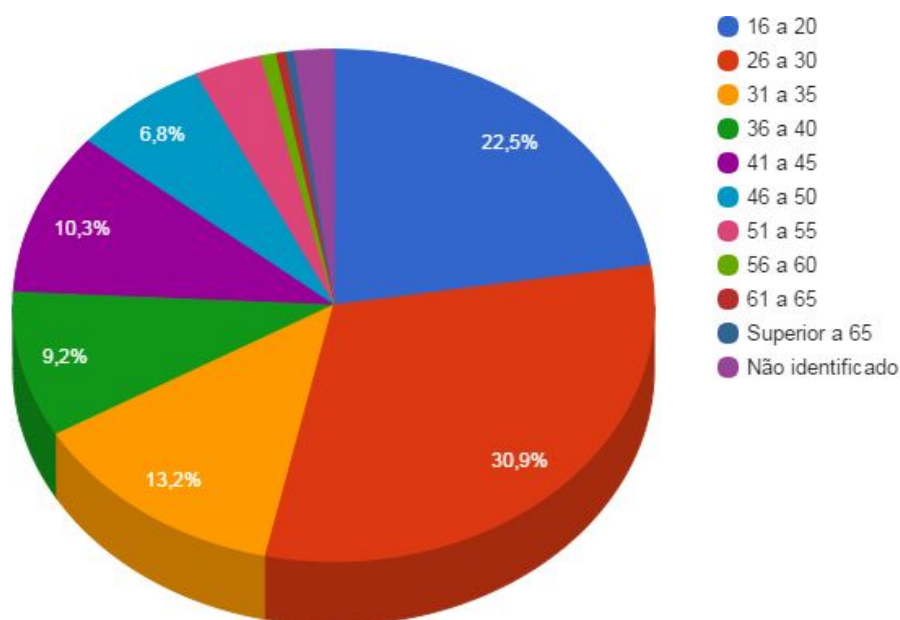
Dados coletados durante os cinco dias antecedentes, o dia do Evento 1 e cinco dias sucedentes ao Evento 1. Produzindo informações sobre a faixa etária dos perfis encontrados com publicações referentes ao Evento 1 .

O futebol Americano de Minas Gerais na Internet (Agosto 2016)

**Tabela 5. Ocorrências por faixa etária Evento 1**

Fontes	Ocorrências
10 a 15	42
16 a 20	2988
26 a 30	4104
31 a 35	1752
36 a 40	1224
41 a 45	1362
46 a 50	906
51 a 55	452
56 a 60	103
61 a 65	63
Superior a 65	58
Não identificado	269
Total	13323

No gráfico seguinte os dados são ilustrados e destacados com o percentual referente a cada uma das fontes de dados, para os dois eventos. Mostrando a força entre eles na internet.



**Figura 11. Resultado do levantamento para o Evento 1, por faixa etária.**

#### **4.6 Dados separados por grupo de faixa etária referentes ao Evento 2**

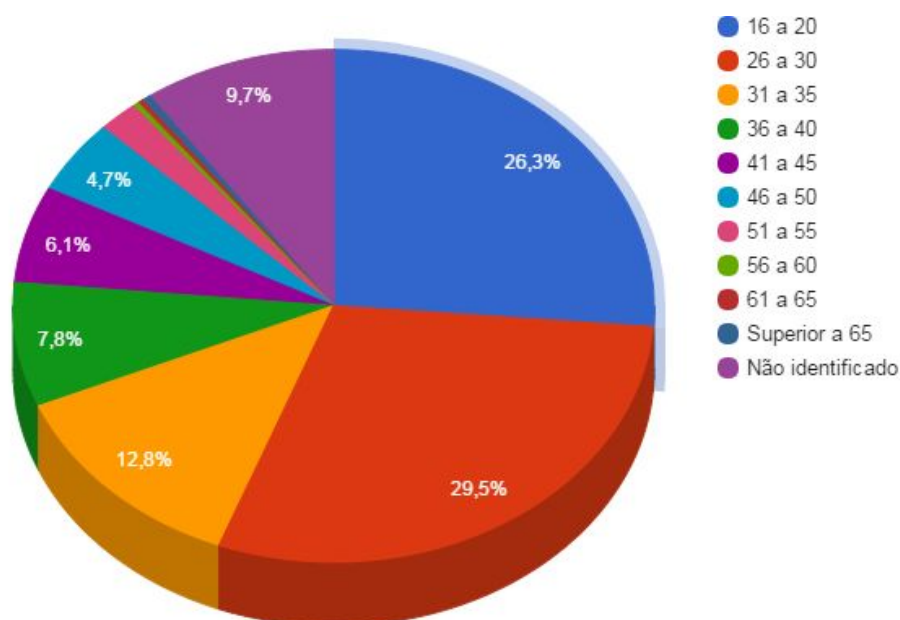
Dados coletados durante os cinco dias antecedentes, o dia do Evento 2 e cinco dias sucedentes ao Evento 2. Produzindo informações sobre a faixa etária dos perfis

encontrados com publicações referentes ao Evento 2 . Comparando com o Evento 1, seria interessante uma coleta de informações mais ampla sobre a idade dos autores das publicações encontradas ou até mesmo aumentar o universo de pesquisa, isso, devido a grande diferença entre as quantidades pesquisadas.

**Tabela 6. Ocorrências por faixa etária Evento 2**

Fontes	Ocorrências
10 a 15	1
16 a 20	181
26 a 30	203
31 a 35	88
36 a 40	54
41 a 45	42
46 a 50	32
51 a 55	14
56 a 60	2
61 a 65	2
Superior a 65	3
Não identificado	67
Total	689

No gráfico seguinte os dados são ilustrados e destacados com o percentual referente a cada uma das fontes de dados, para os dois eventos. Mostrando a força entre eles na internet.



**Figura 12. Resultado do levantamento para o Evento 2, por faixa etária.**

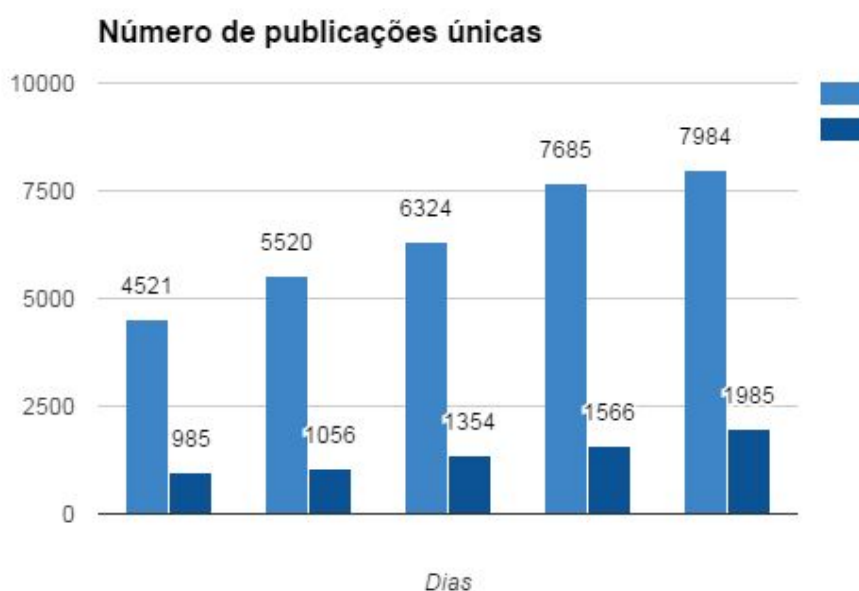
#### 4.7 Dados por cinco dias anteriores aos eventos Evento 1 e Evento 2

Dados coletados durante os cinco dias antecedentes aos Evento 1 e Evento 2. Produzindo informações sobre o número de publicações únicas referentes ao Evento 1 e Evento 2. Em mais um caso de comparação, o Evento 2 manteve um crescimento relevante em comparação ao Evento 1, como pode ser visto na Figura 9, mostrando ser potencialmente interessante para fins publicitários.

**Tabela 7. Ocorrências nos 5 dias anteriores Evento 1 e Evento 2**

Período	Evento 1	Evento 2
5 Dias	4521	985
4 Dias	5520	1056
3 Dias	6324	1354
2 Dias	7685	1566
1 Dias	7984	1985

No gráfico seguinte os dados são destacados com informações sobre publicações únicas nos 5 dias anteriores aos Evento 1 e Evento 2.



**Figura 13. Resultado do levantamento por publicações únicas nos 5 dias anteriores, para os Evento 1 e Evento 2**

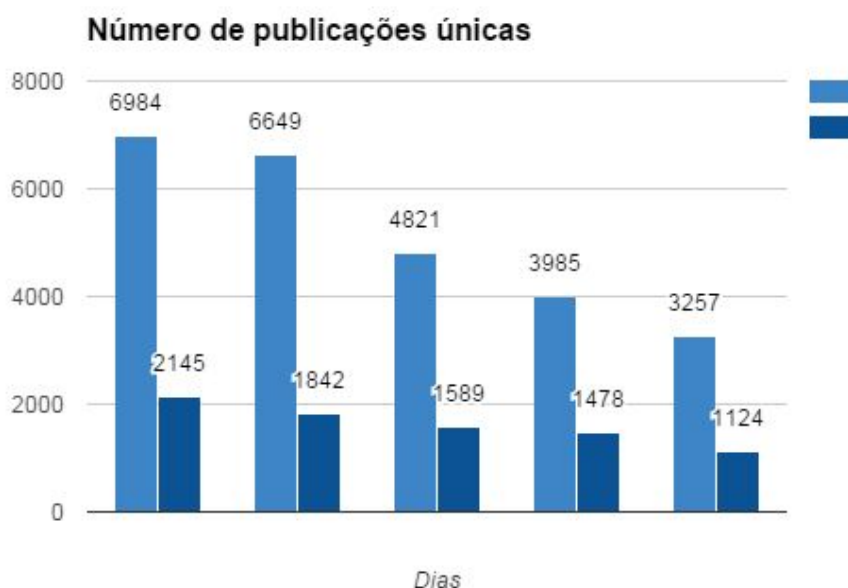
#### 4.8 Dados por cinco dias posteriores aos eventos Evento 1 e Evento 2

Dados coletados durante os cinco dias posteriores aos Evento 1 e Evento 2. Produzindo informações sobre o número de publicações únicas referentes ao Evento 1 e Evento 2 .

**Tabela 8. Ocorrências nos 5 dias anteriores Evento 1 e Evento 2**

Período	Evento 1	Evento 2
5 Dias	442	142
4 Dias	561	256
3 Dias	742	714
2 Dias	1451	921
1 Dias	2541	1312

No gráfico seguinte os dados são destacados com informações sobre publicações únicas nos 5 dias posteriores aos Evento 1 e Evento 2.



**Figura 14. Resultado do levantamento por publicações únicas nos 5 dias posteriores, para os Evento 1 e Evento 2**

#### **4.9 Dados por densidade geográfica das publicações para o Evento 1**

Dados coletados durante os cinco dias anteriores e posteriores ao Evento 1. Produzindo informações sobre o número de publicações únicas, separados e destacados por densidade geográfica, para as cinco cidades mais representadas.

**Tabela 9. Percentual de publicações por cidade, para os Evento 1**

Cidade	Percentual
Belo Horizonte	42
São Paulo	14
Pouso Alegre	7
Rio de Janeiro	6
Porto Alegre	3
Outros	28

No gráfico seguinte os dados são destacados com informações sobre publicações únicas por cidade, para o Evento 1.



**Figura 15. Resultado do levantamento de publicações únicas por cidade para o Evento 1**

#### **4.10 Dados por densidade geográfica das publicações para o Evento 2**

Dados coletados durante os cinco dias anteriores e posteriores aos Evento 2. Produzindo informações sobre o número de publicações únicas, separados e destacados por densidade geográfica, para as cinco cidades mais representadas. Não por acaso, os dados mostram exatamente as cidades onde o esporte em questão, o Futebol Americano tem mais força no Brasil.

**Tabela 10. Percentual de publicações por cidade, para os Evento 2**

Cidade	Percentual
Belo Horizonte	39
São Paulo	19
Rio de Janeiro	9
Curitiba	7
Porto Alegre	4
Outros	22

No gráfico seguinte os dados são destacados com informações sobre publicações únicas por cidade, para o Evento 2.



**Figura 16. Resultado do levantamento de publicações únicas por cidade para o Evento 2**

#### 4.10 Dados por tipo de mídia anexada às publicações referentes ao Evento 1

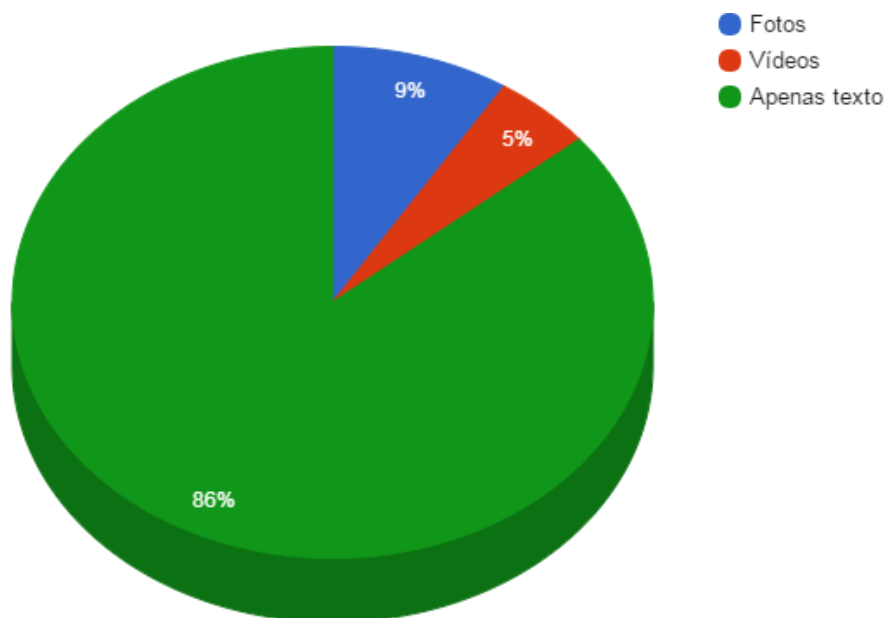
Dados coletados durante os cinco dias anteriores e posteriores aos Evento 1. Produzindo informações sobre os tipos de mídia anexados à cada publicação única.

**Tabela 11. Percentual de publicações por cidade, para o Evento 1**

Tipo	Percentual
Fotos	9
Vídeos	5
Áudio	0
Apenas texto	86



No gráfico seguinte os dados são destacados com informações sobre mídias anexadas à cada publicações única, sobre o Evento 1.



**Figura 17. Resultado do levantamento de tipos de mídia anexado em cada publicação única, referente ao Evento 1.**

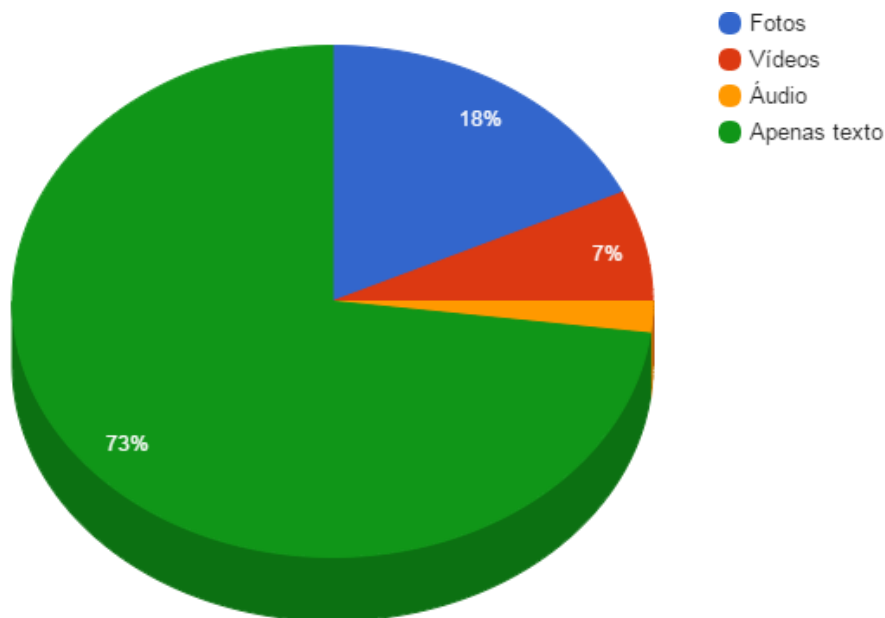
#### 4.11 Dados por tipo de mídia anexada às publicações referentes ao Evento 2

Dados coletados durante os cinco dias anteriores e posteriores aos Evento 2. Produzindo informações sobre os tipos de mídia anexados à cada publicação única. Como o esporte é uma novidade para muitos no Brasil, grande parte dos conteúdos são encontrados na internet, em sites Norte Americanos, o que faz o público ser mais ativo com as mídias sociais.

**Tabela 12. Percentual de publicações por cidade, para o Evento 2**

Tipo	Percentual
Fotos	18
Vídeos	7
Áudio	2
Apenas texto	73

No gráfico seguinte os dados são destacados com informações sobre mídias anexadas à cada publicações única, sobre o Evento 2.



**Figura 11. Resultado do levantamento de tipos de mídia anexado em cada publicação única, referente ao Evento 2.**

## 5. Conclusão

Levando-se em conta o que foi observado, após o levantamento, mineração e geração da informação, seja em forma de gráficos ou tabelas. Conclui-se que o Evento 1, jogo válido pela final do campeonato mineiro de futebol americano, mostrou-se potencialmente interessante para fins publicitários, dentro das mídias sociais e internet em geral, em comparação ao Evento 1, jogo válido pelo campeonato brasileiro de futebol entre o Clube Atlético Mineiro e Cruzeiro Esporte clube, que possui um espaço incomparavelmente maior na mídia, o Evento 2 se mostrou proporcionalmente relevante, gerando um número expressivo de publicações únicas, dentro de um público bem segmentado. Assim como nas publicações separadas por tipo de mídia em anexo, sendo vídeos, fotos, áudios ou somente texto, o evento da final do campeonato mineiro de futebol americano apresentou resultados que mostram um crescimento bem forte e também um padrão interessante para o mercado publicitário, devido ao seu público bem segmentado e marcado pelo número de publicações. Já o gráfico comparativo para fontes segmentadas ou meios de comunicação, ilustra perfeitamente a questão do espaço destinada para cada um dos eventos, servindo como defesa para o fato do Evento 2 ser atraente para investimentos em publicidade. Pois esse gráfico mostra o acompanhamento proporcional do Evento 2 para o Evento 1. Por tanto, com o mesmo espaço nos meios de comunicação, o Evento 2 teria certamente a mesma ou até maior visibilidade que o Evento 1, podendo ser um bom investimento a médio ou longo prazo.

Como trabalhos futuros, seria interessante avaliar o poder aquisitivo desse novo público, podendo assim apontar possíveis produtos e serviços a serem oferecidos.

## Referências

Teófilo, Daniel. “Tecnologia-Tudo é difícil quando não se sabe! - Mineração de Dados”.

Disponível: <http://danielteofilo.files.wordpress.com/2015/02/image.png>  
Acesso: julho/ 2016.

Selenium Nodes. “Put your web browser to work with KNIME”.

Disponível: <http://seleniumnodes.com/img/selenium.png>  
Acesso: julho/ 2016.

Microsoft. “We Loaded 1TB in 30 Minutes with SSIS, 2016”.

Disponível: [https://technet.microsoft.com/en-us/library/dd537533\(v=sql.100\).aspx](https://technet.microsoft.com/en-us/library/dd537533(v=sql.100).aspx)  
Acesso: junho/ 2016.

CAMILO, Cássio Oliveira. (cassio@inf.ufg.br). “Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas”.

Disponível:  
[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: junho/ 2016.

MSND. “Conceitos de mineração de dados”

Disponível: [https://msdn.microsoft.com/pt-br/library/ms174949\(v=sql.120\).aspx](https://msdn.microsoft.com/pt-br/library/ms174949(v=sql.120).aspx)  
Acesso: maio/ 2016

Alfenas, Otávio. “Web Scraping e Web Crawler – Introdução e Passos Iniciais”.

Disponível: <http://netcoders.com.br/blog/web-crawler-introducao-e-passos-iniciais/>  
Acesso: maio/ 2016.

Francischelli, Rodrigo Alexandre. “Estudo sobre a influência da extração de dados na web na descoberta de conhecimento estratégico relevante”.

Disponível: <https://repositorio.ucs.br/handle/11338/1216>  
Acesso: maio/ 2016.