

**UNIVERSIDADE DE CAXIAS DO SUL
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS
SISTEMAS DE INFORMAÇÃO**

CAMILA LONGHI

**ANÁLISE DE DADOS PARA APOIO A TOMADA DE DECISÃO
COM ÊNFASE NO MARKETING DIGITAL**

**CAXIAS DO SUL
2016**

CAMILA LONGHI

**ANÁLISE DE DADOS PARA APOIO A TOMADA DE DECISÃO
COM ÊNFASE NO MARKETING DIGITAL**

**Trabalho de Conclusão de Curso
apresentado ao Curso de Sistemas de
Informação da Universidade de Caxias
do Sul, como requisito parcial para
obtenção do grau de Bacharel em
Sistemas de Informação.
Orientador: Prof.^a Dr.^a Helena G. Ribeiro**

CAXIAS DO SUL

2016

A minha família, pai, mãe e irmã, por todo apoio em todos estes anos de universidade e a meu querido marido por toda atenção e suporte em todos os momentos.

Agradeço a minha orientadora por toda atenção e paciência neste processo e principalmente a todos ensinamentos a mim apresentados, que foram de grande valor e ajuda para conclusão deste trabalho.

“Conhecimento não é aquilo que você sabe, mas o que você faz com aquilo que você sabe. ”

Aldous Huxley

RESUMO

Este estudo de caso tem como objetivo o auxílio à tomada de decisão para o Marketing Digital da empresa ALLPS Alimentos Saudáveis, por meio da mineração e análise dos dados coletados em redes sociais sobre assuntos relacionados a alimentação. A ALLPS é um clube de assinaturas voltada para pessoas que possuem alguma restrição alimentar. Conhecer seu público alvo, as necessidades deste público e tomar as decisões corretas referente ao marketing é um desafio para as empresas. Nos últimos anos, as redes sociais vêm crescendo de forma rápida e se tornando um banco de dados rico em informações. Após a coleta e análise destes dados por meio da aplicação de algoritmos de mineração de dados, os dados foram transformados em informações apresentadas a ALLPS Alimentos, pretendendo-se, ao final deste estudo de caso, auxiliar a empresa a obter uma nova visão de seu negócio.

Palavras-Chave: Marketing Digital; Redes Sociais; Mineração; Dados; Análise de Dados; *Twitter*.

ABSTRACT

The aim of this paper is to aid decision-making towards Digital Marketing of the company known as ALLPS Alimentos Saudáveis, through the mining and analysis of the data collected in social networks on subjects related to alimentation. ALLPS is a sign-up club for people who suffer from eating disorders or diet restrictions. Knowing its target audience, its needs and making the right marketing decisions is a challenge for most companies. In recent years, social networks have been growing rapidly and becoming an information-rich database. After the collection and analysis of this data through the application of data mining algorithms, the data was transformed into information that was then presented to ALLPS Alimentos in order to assist the company in obtaining a new vision of its own business.

Key words: Digital Marketing; Social Networks; Mining; Data; Data Analysis; *Twitter*.

LISTA DE SIGLAS

API: Application Programming Interface

BBS: Bulletin Board System

CRISP-DM: Cross Industry Standard Process for Data Mining

DHP: Direct Hashing and Pruning

EUA: Estados Unidos da América

GSP: Generalized Sequential Patterns

HTTP: Hypertext Transfer Protocol

ICQ: I (ai), C (si), Q (kiu), "I seek You"

IP: Internet Protocol

KD: Knowledge Discovery

KDD: Knowledge Discovery in Databases

KDT: Knowledge Discovery in Texts

MDR: Multifactor Dimensionality Reduction

MSN: Microsoft Service Network

PT: Português

REST: Representational State Transfer

RSS: Really Simple Syndication

SAS: Statistical Analysis System

SMS: Short Message Service

SWOT: Strengths, Weakness, Opportunities and Threats

SPSS: Statistical Package for the Social Science

TCC: Trabalho de Conclusão de Curso

WEKA: Waikato Environment for Knowledge Analysis

LISTA DE FIGURAS

Figura 1 - Marketing de Massa e Marketing One-to-One.	18
Figura 2 - Representação de Rede Social	20
Figura 3 - Abordagens do KD.....	28
Figura 4 - Etapas do processo de KDD.....	29
Figura 5 - Processo de KDT.....	30
Figura 6 - Etapas do Processo de Indexação Automática.....	32
Figura 7 - Algoritmo de Stemming para Língua Portuguesa	34
Figura 8 - Algoritmo Apriori	38
Figura 9 - Algoritmo K-Means	40
Figura 10 - Plano de Trabalho.....	49
Figura 11 - Análise SWOT	50
Figura 12 - Análise SWOT - ALLPS	52
Figura 13 - Aplicação Java – Max. Tweets.....	57
Figura 14 - Tela Inicial - Aplicação Java.....	58
Figura 15 - Arquivo ARFF	59
Figura 16 - Tela Inicial NodeXL.....	60
Figura 17 - Captura de Tweets.....	61
Figura 18 - Importação de Dados.....	62
Figura 19 - Autorização Twitter	63
Figura 20 - Código de Autorização.....	64
Figura 21 - Autorização NodeXL	65
Figura 22 – Dados retornados Nós.	66

Figura 23 - Dados retornados Vértices	67
Figura 24 - Tela Inicial Netlytic.....	68
Figura 25 - Datasets Netlytic.....	69
Figura 26 - Configuração Dataset Instagram	70
Figura 27 - Configuração Dataset Facebook	70
Figura 28 - Arquivo ARFF	72
Figura 29 - Conversão CSV.....	72
Figura 30 - Aplicação fórmula Excel	73
Figura 31 – Preparação dos dados para o Apriori	74
Figura 32 - Preparação de Dados para o K-Means	75
Figura 33 - Arquivo ARFF - Apriori	76
Figura 34 - Arquivo ARFF - K-Means.....	77
Figura 35 - Tela Inicial WEKA.....	80
Figura 36 - Seleção algoritmo Apriori.....	81
Figura 37 - Parâmetros Apriori.....	83
Figura 38 - Resultados Apriori	84
Figura 39 - Tela Inicial WEKA.....	85
Figura 40 - Seleção algoritmo K-Means.....	86
Figura 41 - Parâmetros iniciais K-Means	87
Figura 42 - Parâmetros K-Means.....	88
Figura 43 - Resultados NodeXL – Diet	94
Figura 44 - Resultados NodeXL - Light.....	95
Figura 45 - Resultados NodeXL – Glúten	96
Figura 46 - Resultados NodeXL - Lactose	97

Figura 47 - Nuvem de Tags - Gluten Free.....	98
Figura 48 - Nuvem de Tags - Clube do Zero	99
Figura 49 - Nuvem de Tags - ALLPS Alimentos.....	99
Figura 50 - Network Facebook - Glúten Free	100
Figura 51 - Network Facebook - Clube do Zero	101
Figura 52 - Network Facebook - ALLPS Alimentos	102
Figura 53 - Resultado Instagram – Diet.....	103
Figura 54 - Resultado Instagram – Glúten	104
Figura 55 - Resultado Instagram – Lactose	105
Figura 56 - Resultado Instagram - Light	106
Figura 57 - Tela Inicial - Tagul.....	107
Figura 58 - Importação texto	108
Figura 59 - Limpeza de Palavras.....	109
Figura 60 - Resultado - Nuvem de Tags	110
Figura 61 - Análise SWOT atualizada - ALLPS	113

LISTA DE TABELAS

Tabela 1 - Principais campos de um tweet.....	23
Tabela 2 - API's.....	24
Tabela 3 - Stopwords	32
Tabela 4 - Ferramentas de Mineração de Dados.....	41
Tabela 5 - Probabilidade de Kappa	44
Tabela 6 - Perguntas para análise SWOT.....	51
Tabela 7 - Processamento de Língua Natural.....	53
Tabela 8 - Combinações mais conhecidas pela ALLPS.....	53
Tabela 9 - Alimentos ALLPS	54
Tabela 10 - Campos Twitter	56
Tabela 11 - Resultado Alimentos – K-Means	91
Tabela 12 - Resultados Apriori.....	92

SUMÁRIO

1.	INTRODUÇÃO	12
1.1.	OBJETIVO.....	13
1.2.	CONSIDERAÇÕES INICIAIS	14
2.	MARKETING DIGITAL.....	16
2.1.	ESTRATÉGIAS DE MARKETING.....	16
2.2.	REDES SOCIAIS.....	18
2.2.1.	Histórico	20
2.2.2.	Análise de Dados em Redes Sociais.....	21
2.2.3.	Twitter	21
2.2.4.	Instagram.....	25
3.	DESCOBERTA DE CONHECIMENTO.....	28
3.1.	O PROCESSO DE KDT	30
3.1.1.	Coleta de Dados (Extração de Informação).....	31
3.1.2.	Pré-Processamento dos Dados	31
3.1.3.	Indexação	31
3.1.4.	Mineração de Dados (Data Mining)	34
3.1.5.	Avaliação	35
3.1.6.	Consolidação do Conhecimento	35
3.2.	MINERAÇÃO DE DADOS	36
3.2.1.	Métodos de Mineração	36
3.2.2.	Associação (association)	37
3.2.2.1	Algoritmo de Associação - Apriori	37
3.2.3.	Análise de Agrupamento (clustering).....	39
3.2.3.1	Algoritmo de Agrupamento - K-Means.....	39
3.3.	FERRAMENTAS PARA MINERAÇÃO DE DADOS	40
4.	TRABALHOS RELACIONADOS	42

4.1.	A UTILIZAÇÃO DAS REDES SOCIAIS: UM ESTUDO DE CASO SOBRE O FACEBOOK COMO ESTRATÉGIA DE MARKETING NO BANCO BRADESCO S.A	42
4.2.	ESTUDO DE PERFIL DE USUÁRIOS DE REDES SOCIAIS ATRAVÉS DA MINERAÇÃO DE DADOS	44
5.	ESTUDO DE CASO	46
5.1.	ALLPS ALIMENTOS SAUDÁVEIS	46
5.2.	PLANO DE TRABALHO E IMPLANTAÇÃO	47
5.3.	ENTENDIMENTO DO NEGÓCIO	49
5.4.	COLETA DE DADOS.....	55
5.4.1.	Aplicação em JAVA	55
5.4.2.	NodeXL.....	59
5.4.3.	Netlytic	68
5.5.	PRÉ-PROCESSAMENTO	71
5.5.1.	Arquivos gerados pela Aplicação JAVA.....	71
5.5.2.	Arquivos gerados pelo NodeXL	78
5.5.3.	Arquivos gerados pelo Netlytic	78
5.6.	APLICAÇÃO DAS TÉCNICAS DE MINERAÇÃO DE DADOS	78
5.6.1.	Algoritmo de Associação - Apriori.....	79
5.6.2.	Algoritmo de Agrupamento - K-Means	85
5.7.	ANÁLISE DOS DADOS.....	89
5.7.1.	Resultado Algoritmos de Mineração	89
5.7.2.	Resultado NodeXL	92
5.7.3.	Resultado Netlytic	98
5.7.4.	Nuvem de Tags	107
6.	CONCLUSÃO	112
6.1.	TRABALHOS FUTUROS	114
	REFERÊNCIAS.....	116

1. INTRODUÇÃO

Segundo Silva et al (2013), um dos maiores desafios das empresas nos dias atuais é conhecer seu público alvo para compreender e entender como utilizar o Marketing Digital. Com o crescimento da internet, as empresas têm focado seus trabalhos em práticas que adotam essa nova forma de marketing. É por meio da internet que as empresas estão buscando coletar dados importantes e relevantes que auxiliem seu negócio.

O Marketing Digital “são todas as ações planejadas pelas empresas visando ampliar os negócios de forma sustentada” (KENDZERSKI, 2009, p.23). Seguindo este pensamento, uma das principais estratégias do Marketing Digital é entender como criar um relacionamento com o seu cliente e desenvolver campanhas focadas diretamente ao seu público-alvo. (SIQUEIRA, 2011).

Segundo Diniz (2013), como conhecer seu público-alvo em uma rede na qual o compartilhamento de informações ultrapassa barreiras e possui modificações constantes? Como controlar essa informação que foge ao controle das empresas? Essas respostas estão nas redes sociais e ganharam espaço no compartilhamento de interesses, colaboração social e tem se tornado alvo de estudos de dados nos últimos anos.

Conforme Raposo (2010), é em busca destes dados e pensando na melhor forma de auxiliar empresas na implementação do marketing digital que se utiliza a Mineração de Dados (Data Mining). É por meio dessa mineração e análise dos dados retirados das redes sociais que se gera informações para auxílio empresas na tomada de decisão.

A mineração de dados, segundo Braga (2005) é uma etapa do processo de descoberta do conhecimento, o qual consiste em coletar, armazenar, tratar, analisar e transformar dados de um determinado período e contexto, a fim de transformá-los em informação, seja pelo meio de grafos, gráficos, nuvens de *tags* e etc.

As redes sociais possuem uma grande quantidade de publicações vistas computacionalmente como dados. Conforme Bevenuto (2010), estes dados podem ser extraídos e estruturados por intermédio de API's de sistemas desenvolvidos por empresas e por profissionais da área de banco de dados e análise de dados. Algumas

API's apresentam dados mais fáceis de serem estruturados, outras são um pouco mais complexas. Após obter esses dados, é realizada a aplicação de algoritmos de mineração de dados específicos, e assim eles são transformados em informações, que podem auxiliar as empresas na tomada de decisões.

1.1. OBJETIVO

O objetivo geral deste estudo de caso é analisar e minerar os dados disponíveis em redes sociais sobre assuntos relacionados a alimentação para transformá-los em informações que auxiliem as empresas a tomar decisões com foco no Marketing Digital, mediante a um estudo de caso para uma empresa específica.

Tem-se como objetivos específicos:

- a) Coletar informações e estruturar conforme o negócio da empresa;
- b) Escolher ferramentas aptas para a coleta de dados em redes sociais;
- c) Efetuar a coleta de dados, avaliar e estruturar os dados coletados;
- d) Organizar os dados coletados e transformar os dados em informações;
- e) Escolher e selecionar métodos/algoritmos de mineração de dados;
- f) Avaliar o processo e os resultados obtidos a fim de extrair aprendizagens e conhecimento.

1.2. CONSIDERAÇÕES INICIAIS

Neste estudo de caso, no capítulo 2, são apresentados, por meio de referências bibliográficas, entre outras, conceitos de Marketing Digital e suas estratégias, bem como a importância do marketing para as empresas. Também neste capítulo, são apresentadas algumas das principais redes sociais e como é realizar a análise de dados que são gerados nessas redes sociais. No capítulo 3 são apresentadas as etapas do processo de Descoberta do Conhecimento bem como as principais técnicas de mineração de dados e seus algoritmos e ferramentas, com o intuito de realizar uma análise de dados coletadas conforme necessidade da empresa que foi utilizada como estudo de caso deste trabalho.

Já no capítulo 4, são apresentados alguns trabalhos relacionados a este estudo, em um deles é trabalhado o marketing digital por meio das redes sociais e no outro é apresentado o uso da mineração de dados para conhecimento de perfil no *Facebook*. No capítulo 5 apresenta-se a proposta deste estudo de caso e seu desenvolvimento, onde conhece-se um pouco da empresa ALLPS Alimentos Saudáveis, na qual foram focados as necessidades e resultados. Após, apresenta-se o desenvolvimento de cada etapa do processo e então os resultados conquistados com este estudo.

2. MARKETING DIGITAL

Neste capítulo são apresentadas, por meio de referencial teórico, as estratégias de marketing e o uso de dados das redes sociais para a aplicação do marketing digital.

2.1. ESTRATÉGIAS DE MARKETING

O que é Marketing? “O marketing envolve a identificação e a satisfação das necessidades humanas e sociais, [...] podendo-se dizer que ele ‘supre necessidades lucrativamente’ (KOTLER; KELLER, 2006, p. 4). O marketing, em geral, é uma atividade ou uma ação focada na satisfação e desejos dos clientes, gerando lucro para quem fornece e para quem recebe.

Segundo Morais (2013), o marketing em geral por ser dividido em duas formas de aplicação, dependendo da estratégia necessária, que são conhecidos por Marketing Digital e Marketing Tradicional.

O que é Marketing Digital? Qual a diferença entre Marketing Digital e o Marketing Tradicional? Conforme Adolpho (2010), marketing é sempre marketing, seja digital ou analógico, ou seja, sua finalidade é sempre a mesma mudando apenas suas estratégias.

No Marketing Digital encontra-se as estratégias trabalhadas de forma *online*, que são divididas e conhecidas como *pull* e *push*. Conforme informado por Gualda (2012), na estratégia *pull*, o consumidor procura os conteúdos, como em uma pesquisa no *Google*, *blogs*, entre outros, não existindo restrição quanto ao tipo e tamanho do conteúdo pesquisado, no qual é o próprio interessado que vai atrás dessa informação. Na estratégia *push*, é a empresa que “empurra” informações aos clientes, por *e-mails*, postagens patrocinadas em redes sociais e etc., possibilitando assim uma visibilidade mais direta das informações, ou seja, uma visibilidade forçada. Apesar disso, o *push* pode trazer pontos negativos as empresas, pois este tipo de estratégia de marketing pode ser rejeitado pela maioria dos consumidores, por diversas vezes, e ser considerada invasivo.

A internet tem-se tornado um dos maiores meios de comunicação, devido a sua agilidade e rapidez de alcance de marketing, atingindo uma grande massa de consumidores. É por meio dela que pequenas e grandes empresas vêm fazendo a divulgação de seu nome e de produtos e, assim, desenvolvendo seu marketing *on-line*.

Conforme publicação realizada pela Folha de São Paulo, Marroquín (2014) cita que o novo consumidor está mais bem informado, mais conectado e está interessado nos compromissos das empresas. "O novo consumidor está conectado e utiliza a internet para pesquisar produtos e serviços do seu interesse. Cerca de 80% a 90% deles buscam opiniões sobre produtos e reputação de empresas principalmente nas redes sociais e sites de busca" (TERRA, 2015). Além disso, a famosa propaganda boca-a-boca é, há anos, responsável pelo sucesso e insucesso de diversos produtos do mercado, e o consumidor busca pelos melhores produtos e quer opinar sobre eles.

Segundo Rebelatto et al. (2014), é recorrente encontrar empresas que estão iniciando suas operações e decidem ingressar no mercado com soluções demasiadamente abrangentes - por medo de errarem ao fazer algum tipo de segmentação e deixarem de atingir potenciais consumidores. Por isso, dentro do Marketing Digital, encontra-se uma subdivisão de Marketing, conhecidos como Segmentado, em Massa e *One-to-One*.

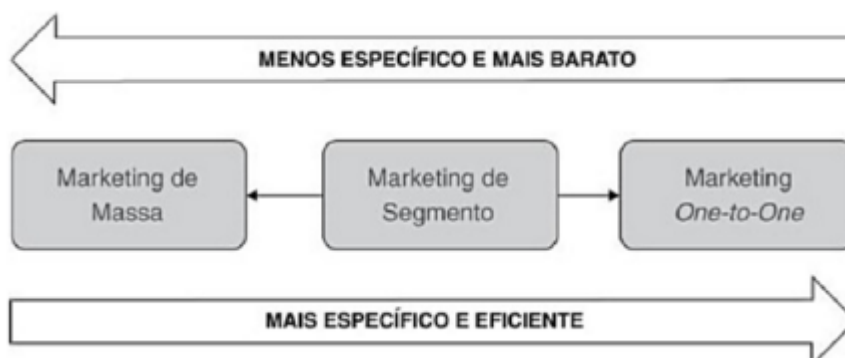
Segundo Salvador (2014), Marketing Segmentado parte de um princípio de que não existe produto que atenda ou agrade ao público como um todo, assim atingido apenas um segmento específico de público.

Kotler e Keller (2006) definem o Marketing em Massa como aquele executado pelas empresas que se dedicavam à produção, distribuição e promoção de um único ou de poucos produtos. O Marketing em Massa funcionou bem enquanto existiam poucos produtos e poucos concorrentes. Essa estratégia era muito proveitosa para as corporações que a executavam, porque apesar de atender seus consumidores apenas parcialmente, implicava no geral em menores custos às empresas, segundo Kotler e Keller (2006).

"O Marketing *One-to-One*, deve ser utilizado para conseguir uma relação personalizada com os clientes, de forma a conhecê-los o suficiente para poder oferecer a cada categoria os bens e serviços mais adequados. Esta ferramenta é fundamental na fidelização dos clientes" (HORTINHA. 2001, p. 259)

A Figura 1 ilustra o contraste entre o Marketing em Massa (que atinge todo o público, de forma única) e um Marketing *One-to-One* (que tem em seu foco um público mais seletivo, pensando em seu individual).

Figura 1 - Marketing de Massa e Marketing One-to-One.



Fonte: Boone e Kurtz (2009, p. 238)

O mercado é um conjunto de pessoas com disposição para consumir. Para Boone e Kurtz (2009), são os executivos de marketing que devem mapear os fatores que afetam a decisão de compra, selecionar os grupos de consumidores com base nesses perfis, e ajustar as estratégias de marketing para atender mais assertivamente estes segmentos.

No decorrer deste estudo de caso, serão efetuados estudos de práticas de análise e mineração de dados para apoio ao Marketing *One-to-One*, focando a coleta de dados e análise para um grupo específico de público.

2.2. REDES SOCIAIS

Rebelatto, destaca as redes sociais seguinte forma

Em um mundo conectado à internet, cada vez mais as redes sociais assumem um papel importante no cotidiano das empresas e indivíduos. Nos últimos anos, os consumidores ganharam espaço em expor opiniões, compartilhar experiências e influenciar milhares de pessoas, e boa parte disso se deve ao crescimento das chamadas redes sociais. (2014, p. 2031).

As redes sociais, representam “[...] um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados” (MARTELETO, 2001, p.72).

Ao falar de redes sociais, pode-se falar de estreitamento no convívio das pessoas e de suas opiniões. As pessoas têm compartilhado mais informações e opiniões sobre produtos, lugares, problemas, doenças, entre outros. Devido a isso, este compartilhamento tem gerado um número de dados que podem se transformar em informações valiosas as empresas e auxiliá-las a entender e desenvolver uma estratégia de marketing. Para isso, é necessário conhecer seu produto, conhecer seu público e, a partir deste ponto, tomar as decisões corretas para o sucesso do marketing e consequentemente da sua empresa.

O conhecido como “novo consumidor” assiste menos a televisão, consequentemente a menos propagandas, ouve menos rádio e tem escolhido ler notícias por meio de portais e redes sociais na internet. As conversas mudaram de lugar, as discussões sobre diversos assuntos estão centralizadas nas redes sociais. Assim, o marketing digital apareceu para facilitar a vida das pessoas e das empresas.

Em um curto espaço de tempo, é possível divulgar e atingir pessoas de forma abrangente com um custo menor, comparado ao custo investido pelas empresas em marketing televisivo e/ou marketing em rádios. Além disso, o compartilhamento e acesso a essas informações tem acontecido de forma mais ágil e rápida, não apenas com a utilização de computadores, mas também de celulares e *tablets*, que estão cada dia mais acessíveis e populares.

Considerando essas colocações sobre as redes sociais, Goldenrberg et al. (2001), afirma que o estudo sobre o marketing boca-a-boca envolve diretamente as interações dentro da uma rede social. Essa interação tem gerado um aumento de interesse das organizações nas oportunidades que este conteúdo pode conter.

Este estudo de caso, tem como intuito capturar os dados gerados por essas redes sociais. A análise de redes pode ser apresentada por uma definição simples, porém eficaz “conjunto de relações ou ligações sociais entre um conjunto de atores (e também atores ligados entre si) ” (EMIRBAYER e GOODWIN, 1994, p. 1449). A Figura 2 apresenta de forma ilustrativa como são ligados os dados gerados pelas pessoas a Internet, mostrando que todos os dados ficam centralizados.

Figura 2 - Representação de Rede Social



Fonte: iStock

2.2.1. Histórico

Os primeiros relatos de serviços que possuem características parecidas com as utilizadas nas redes sociais, tem seus primórdios no ano 1969, com o desenvolvimento da *dial-up* e o do CompuServe — um serviço comercial de conexão à internet em nível internacional muito propagado nos EUA. Em 1971, surgiu o primeiro envio de e-mail, sendo seguido pela criação do BBS, ou seja, *Bulletin Board System*, um sistema criado por estudantes de Chicago, para convidar amigos para eventos, usando linhas telefônicas e *modem* para transmissão de dados.

Enfim, as redes sociais surgiram a partir de 1990, conforme GUEDES (2012), surgindo os primeiros chats como ICQ, MSN, Skype e serviços de bate papo em sites. A partir dos anos 2000, surgiram redes sociais como o *Facebook*, *Twitter*, *Instagram*, *MySpace*, *Orkut*, sendo que as duas últimas citadas já são obsoletas e não são mais utilizadas.

2.2.2. Análise de Dados em Redes Sociais

A análise de dados em redes sociais está estabelecendo um novo paradigma sobre pesquisa social, Marteleto diz

Para estudar como os comportamentos ou as opiniões dos indivíduos dependem das estruturas nas quais eles se inserem, a unidade de análise não são os atributos individuais (classe, idade, gênero), mas o conjunto de interações que os indivíduos estabelecem através de suas interações uns com os outros. (2001, p. 71-81).

Segundo Marteleto (2001) as redes sociais possuem espaços informais e são iniciadas a partir do compartilhamento de interesses e/ou valores entre os usuários. Diariamente, assuntos gerais como política, globalização, experiências são motivações para o desenvolvimento das redes.

Nos próximos capítulos estão descritas as principais características do *Twitter* e *Instagram*, algumas nas redes sociais mais utilizadas no Brasil para Marketing Digital, segundo Drubsky (2015).

2.2.3. *Twitter*

O *Twitter* é uma rede social conhecida mundialmente e tem em sua característica principal a semelhança com um *blog*, mas permite apenas postagens de até 140 caracteres. A rede social foi fundada em março de 2006 por Jack Dorsey, Evan Williams e Biz Stone como um projeto paralelo. Conforme Smaal (2010), a ideia inicial foi de Dorsey em uma reunião de discussão de ideias (brainstorming) em que ele falava sobre um serviço de troca de status, como um SMS.

Na época de seu lançamento, a rede obteve um enorme número de usuários, dentre estes, diversas celebridades e grandes empresas, aproveitando a vantagem de escrever em poucas palavras mensagens que vão direto ao ponto. Com o passar dos anos, o *Twitter* vem perdendo um número significativo de seguidores, mas ainda tem um valor alto de mercado, estipulado em 11 bilhões de dólares, segundo matéria da Folha de São Paulo, e por este motivo ainda há usuários, marcas e empresas que fazem questão de estarem presentes nesta rede social por conhecerem seu valor.

No *Twitter*, as postagens são chamadas de *tweets* e podem ser direcionadas diretamente a outros perfis ou até mesmo em mensagens privadas, caso seja necessário. As empresas ou marcas anunciam seus produtos nestes *tweets* geralmente acompanhados por links que encaminham os usuários a outras páginas, com mais informações. O *Twitter* pode ser considerado um canal de captação de novos clientes, ou até mesmo um canal de comunicação com os clientes atuais.

O site conta ainda com o recurso das *hashtags*, que são os marcadores das postagens que permitem efetuar buscas com a palavra ou até mesmo frases utilizadas. O *Twitter* pode ser considerado uma ótima ferramenta de negócios, desde que a empresa mantenha sempre seu conteúdo atualizado e mantenha seu público interessado em seus posts. Por ser uma ferramenta considerada simples e fácil de utilizar, é importante que as empresas tenham uma atenção e cuidado com seu conteúdo e com o retorno aos seus clientes.

Um caso de uso do *Twitter* de sucesso é o da microempresa Camisetaria. Conforme Gonçalves (2012), no início de sua história a Camisetaria apenas vendia algumas camisetas com estampas diferentes, pela internet. Hoje a empresa possui um grande estoque de camisetas, com diversas estampas e modelos para seus clientes. A Camisetaria utiliza o *Twitter* para divulgar seus lançamentos de estampa e outros acessórios e ainda, aceita ideias de seus clientes, na qual as melhores se tornam camisetas de qualidade, além de tornar a empresa uma amiga do cliente, com um relacionamento próximo. Ainda, o cliente recebe créditos na loja, quando sua ideia é escolhida, além de um prêmio em dinheiro. A Camisetaria conquistou diversos clientes fiéis no *Twitter* e hoje seu perfil @Camisetaria é seguido por mais de 92,2 mil *followers*, forma como os seguidores do *twitter* são chamados.

Para realizar uma análise dos dados gerados no *Twitter*, por meio dos *tweets*, é necessário entender como um *tweet* é composto e quais são os campos disponíveis após a extração por meio de API's. A Tabela 1, apresenta as partes de um *tweet*.

Tabela 1 - Principais campos de um *tweet*

Campo	Tipo	Descrição
<i>hashtags</i>	<i>Array</i> de objetos	Representa a <i>hashtag</i> de um <i>tweet</i> .
<i>coordinates</i>	Coordenadas	Contém as coordenadas geográficas do local do usuário ou aplicação responsável pelo <i>tweet</i> .
<i>created_at</i>	<i>String</i>	Data e hora da criação do <i>tweet</i> .
<i>current_user_retweet</i>	Objeto	Contém o identificador do usuário que realizou um <i>retweet</i> .
<i>entities</i>	<i>Entities</i>	Contém diversas informações, como por exemplo, <i>hashtags</i> , urls, usuários mencionados etc.
<i>favorite_count</i>	<i>Integer</i>	Contador de quantas vezes o <i>tweet</i> recebeu <i>like</i> .
<i>favorited</i>	<i>Boolean</i>	Indica se o <i>tweet</i> recebeu <i>like</i> do usuário autenticado.
<i>id</i>	<i>Int64</i>	Identificador único de um usuário do <i>Twitter</i> .
<i>id_str</i>	<i>String</i>	<i>String</i> que representa o identificador único de um usuário do <i>Twitter</i> , recomenda-se utilizar esse campo ao invés do campo <i>id</i> .
<i>In_reply_to_screen_name</i>	<i>String</i>	Caso o <i>tweet</i> em questão seja um <i>retweet</i> , esse campo conterá o usuário responsável pelo <i>tweet</i> original.
<i>In_reply_to_user_id</i>	<i>Int64</i>	Caso o <i>tweet</i> em questão seja um <i>retweet</i> , esse campo conterá o identificador do usuário responsável pelo <i>tweet</i> original.
<i>In_reply_to_user_id_str</i>	<i>String</i>	Caso o <i>tweet</i> em questão seja um <i>retweet</i> , esse campo conterá a

		<i>String</i> do identificador do usuário responsável pelo <i>tweet</i> original.
<i>lang</i>	<i>String</i>	Contém a linguagem em que foi postado o <i>tweet</i> (Inglês, português etc.).
<i>retweet_count</i>	<i>Integer</i>	Contador de quantos <i>retweets</i> teve um <i>tweet</i> .
<i>text</i>	<i>String</i>	Contém a mensagem do <i>tweet</i> .

Fonte: *Twitter* (2016).

Após entender quais os campos compõem um *tweet*, o próprio *Twitter* disponibiliza algumas API's para coleta de dados. Na Tabela 2, estão apresentadas as API's disponíveis e suas vantagens e desvantagens.

Tabela 2 - API's

API	Vantagens	Desvantagens
<i>REST API</i>	Implementação simples. Não necessita de autenticação.	Retorna apenas os últimos dados inseridos. Não permite que sejam efetuadas pesquisas.
Streaming API	Pesquisa em tempo real. Não possui limitação de tráfego.	Necessita de autenticação. Não permite a captura de <i>tweets</i> antigos. Necessita de uma conexão HTTP persistente. Proibição de múltiplas conexões simultâneas. Taxa limite medida pelo endereço IP requisitante.

<i>Search API</i>	Coleta das mensagens dos usuários em um período recente. Não necessita de autenticação.	Restrição de coleta de mensagens postadas num período de 6 a 9 dias anteriores à data de coleta. Tamanho máximo dos parâmetros de busca de mil caracteres. Restringe o número de buscas pela complexidade e frequência. Taxa limite medida pelo endereço IP requisitante.
-------------------	--	--

Fonte: *Twitter* (2016).

2.2.4. *Instagram*

Segundo Ribeiro (2015), o *Instagram* é uma das redes sociais *online* mais utilizadas na atualidade. É por meio dessa ferramenta que os usuários vêm compartilhando imagens e vídeos, além de permitir que seus usuários tirem fotos e façam vídeos pelo próprio aplicativo, podendo compartilhar automaticamente em outras redes sociais, como *Twitter*, *Facebook*, além de aplicar filtros digitais nas imagens geradas.

O *Instagram* foi criado em 2010 por Kevin Systrom, na época a ferramenta era chamada de *Burbn* e tinha como base a publicação de imagens organizadas por geolocalização e chat, mas no seu início o aplicativo não conquistou muitos usuários. Após um investimento de 500 mil dólares na ferramenta e a inclusão de Mike Kriegle como sócio da empresa, foi criada a versão *Instagram*, disponibilizada apenas para smartphones com sistema operacional iOS (Apple) e após sua popularização, foi liberada a versão para Windows Phone (Microsoft) e Android (Google).

A principal funcionalidade do *Instagram* é o compartilhamento de fotos e vídeos, conforme Bergström e Bäckman (2013). Nas legendas das fotos e dos vídeos postados

é possível incluir *hashtags* que caracterizem a foto compartilhada, como exemplo #comida, #sol, #verão, #praia entre outros.

É possível então seguir outros usuários e acompanhar suas imagens e caracterizações, além de ser seguido por amigos e conhecidos. Todo conteúdo gerado é visível por todos os usuários da rede, permitindo que se explore outros conteúdos e interesses. Mas o *Instagram* também possui a funcionalidade de se publicar fotos para apenas alguns usuários, de forma que apenas seus seguidores possam visualizar.

As *hashtags* do *Instagram* funcionam da mesma forma que no *Twitter*, podendo ser utilizadas tanto na legenda das imagens, como em seus comentários. Ao clicar em uma *hashtag* o *Instagram* realiza uma busca e retorna ao usuário diversas postagens que utilizaram a mesma *hashtag* de caracterização. Pode-se considerar a *hashtag* uma forma de organização, representação e recuperação de dados e informações, podendo-se analisar outras atribuições que acompanham a mesma.

Sendo assim, é possível compreender o quanto o *Instagram* pode ser útil a uma marca ou produto. Mesmo que a proposta inicial do *Instagram* não seja voltada aos negócios, é possível adaptá-la para as necessidades das empresas, efetuar postagens estratégicas e analisar as *hashtags* de emoção utilizadas pelos usuários.

O site Forrester, por Elliott (2014), publicou uma que considera o *Instagram* a rede social que mais promove a comunicação entre empresa/marca e consumidor atualmente. Algumas ferramentas do mercado, como NodeXL, criada por um grupo de pesquisadores de mídias sociais da *Social Media Research Foundation*, uma organização pública da Califórnia e o Netlytic desenvolvida por uma comunidade aberta, com o intuito de executar análise de dados de diversas redes sociais, prometem efetuar a análise dos dados gerados pelo *Instagram*, ainda não testados amplamente.

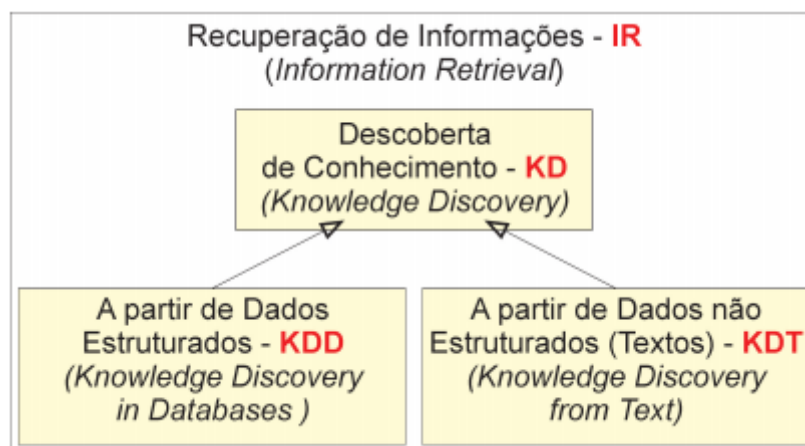
3. DESCOBERTA DE CONHECIMENTO

O aumento de dados gerados e compartilhados, especialmente pela internet, possibilita às empresas o acesso a registros valiosos sobre seus produtos e clientes. Inicialmente estes dados são apenas dados, ou seja, não geram informações consistentes para as empresas. Porém, se bem estruturados, podem trazer conhecimento e auxiliar as empresas nas tomadas de decisões.

Neste contexto surge a Descoberta de Conhecimento ou *Knowledge Discovery* (KD), que é “um processo de análise de dados ou informações, cujo principal objetivo é fazer com que as pessoas possam adquirir novos conhecimentos a partir da manipulação de grandes quantidades de dados” (MORAIS e AMBRÓSIO, 2007, p. 2).

Segundo Moraes e Ambrósio (2007), existem duas abordagens utilizadas no KD, a Descoberta de Conhecimento em Dados Estruturados e a Descoberta de Conhecimento em Dados não Estruturados conforme Figura 3.

Figura 3 - Abordagens do KD



Fonte: Moraes e Ambrósio (2007).

O processo de Descoberta de Conhecimento em Banco de Dados (KDD) consiste em etapas de um processo serem seguidos, que podem ser visualizados de uma forma geral na Figura 4. Este processo compreende todo o ciclo que um dado percorre até se

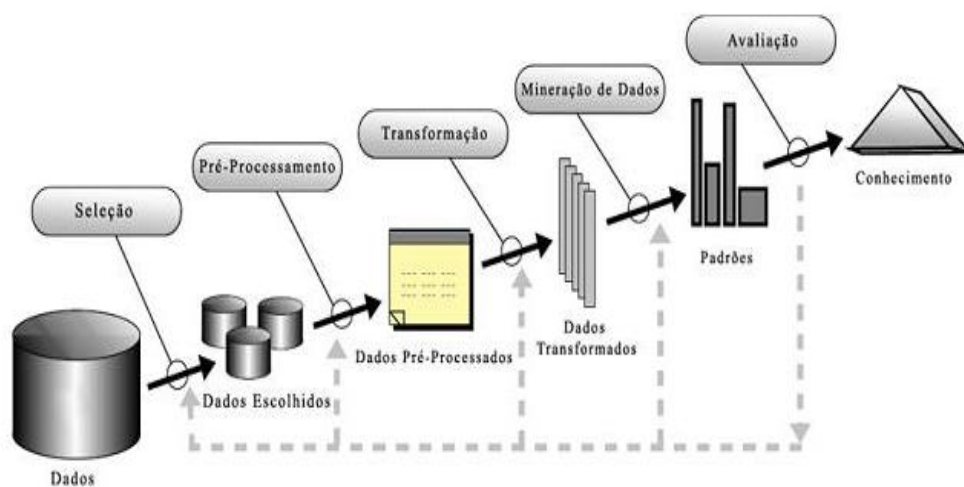
tornar uma informação, ou seja, um conhecimento que permite que os dados sejam melhorados ao longo do tempo.

Conforme Moraes e Ambrósio citam:

O principal objetivo desta área está relacionado à descoberta de correlacionamentos e dados implícitos em registros de bancos de dados, através do estudo e desenvolvimento de processos de extração de conhecimento. Seu principal objetivo é encontrar conhecimento a partir de um conjunto de dados para ser utilizado em algum processo decisório. (2007, p.2).

Cada etapa compreende: conhecer o negócio, selecionar os dados, limpar os dados e pré-processá-los, transformá-los, minerá-los, avaliá-los e consolidá-los em informações. Embora haja uma ordem, uma sequência a ser seguida, segundo Fayyad et al. (1996) este processo é interativo e iterativo, isto é, as decisões tomadas pelos usuários ao longo do processo podem ocasionar a repetição de alguma etapa.

Figura 4 - Etapas do processo de KDD

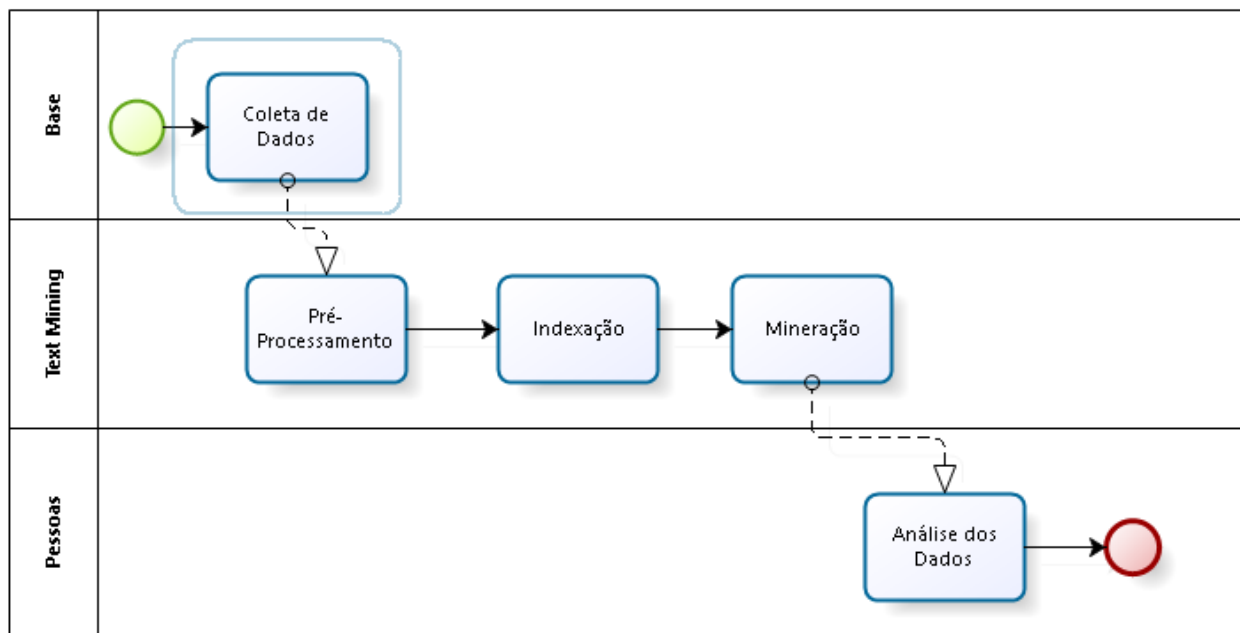


Fonte: FAYYAD et al., 1996

O Processo de KDT (*Knowledge Discovery in Texts*) é muito similar ao processo de KDD, principalmente das etapas de mineração e avaliação (SILVA, 2012, p. 24). A diferença entre KDD e KDT é que o KDT se baseia em “técnicas específicas para tratamento de textos que devem ser utilizadas a fim de se obter conhecimentos implícitos em banco de dados textuais” (BARION; LAGO, 2008, p. 126). Neste processo, os dados

obtidos nas coletas são cruzados a procura de padrões, para que assim se obtenha o conhecimento. O KDT requer uma série de etapas de processamento para a estruturação dos dados. Após o pré-processamento dos dados é que é possível realizar a análise dos mesmos. A Figura 5 apresenta o fluxo do processo de KDT.

Figura 5 - Processo de KDT



Fonte: Autor (2016)

3.1. O PROCESSO DE KDT

O processo de KDT envolve alguns passos que necessitam de conhecimento e decisões tomadas pelos usuários. Estes passos são descritos e explicados a seguir, de forma que se possa entender como funciona todo processo.

3.1.1. Coleta de Dados (Extração de Informação)

É nesta etapa que serão selecionados os textos conforme o domínio do problema. “Considerando os objetivos que se deseja alcançar com o processo, o primeiro passo é eleger o conjunto de textos que será utilizado” (CECI et al., 2010).

Segundo Fayyad et al. (1996), para que a etapa de coleta dos dados seja eficiente, é necessário um analista com conhecimento sobre o domínio apoiando na tarefa para encontrar os objetos necessários. Em geral, “esse procedimento se estabelece basicamente em três ambientes distintos: no diretório de pastas do disco rígido; em tabelas de diferentes bancos de dados e na Internet”. (ARANHA, 2007, p. 41).

Segundo Filho (2014), nesta etapa, pode-se utilizar de *web crawlers*. *Web crawlers* são programas que visitam sites que possuem as informações desejadas e extraem os textos necessários para as próximas etapas do processo.

3.1.2. Pré-Processamento dos Dados

“O objetivo desta fase é eliminação de termos não relevantes (*stopwords*), redução das palavras aos seus radicais (*stemming*), correções ortográficas e outros aspectos morfológicos e também sintáticos que as expressões textuais possuem. ” (CECI et al., 2010). Nesta etapa, é fundamental o Processamento de Linguagem Natural (PLN) pois trata-se da descoberta de estruturas implícitas como por exemplo, a estrutura sintática (RAJMAN; BESANÇON, 1997). De acordo com Ceci et al., (2010), nesta etapa ocorre à transformação do texto em vetores, tabelas, matrizes, etc. A partir deste ponto a informação encontra-se estruturada e pronta para a fase de indexação.

3.1.3. Indexação

É na fase de indexação que são identificadas as similaridades entre os termos coletados. A indexação tem como objetivo identificar a similaridade de significado entre

termos, considerando variações morfológicas e a sinonímia, tendo como resultado um índice para os dados coletados (MORAIS e AMBRÓSIO, 2007).

Segundo Moraes e Ambrósio (2007), com um índice adequado aumenta-se significativamente o desempenho das leituras e buscas na base de dados. Durante a fase de indexação também são aplicadas as técnicas de normalização morfológica.

Na Figura 6 é apresentado o processo de indexação automática, com suas principais fases: a identificação dos termos, a remoção de *stopwords* e a normalização morfológica.

Figura 6 - Etapas do Processo de Indexação Automática



Fonte: Wives (2002).

As *stopwords* devem ser removidas da coleção textual, pois são palavras comuns e sem significado para o resultado final. Após a remoção das *stopwords*, será possível ter uma base de dados mais enxuta o que torna as tarefas de mineração de dados mais ágil, segundo EL-KHAIR (2006). Na Tabela 3 apresenta-se uma relação de *stopwords* que podem ser removidas.

Tabela 3 - *Stopwords*

A	As	Um	Uma	Uns	Umas	O	Os	Onte m
Amanh ã	Agora	Muit o	Muitos	Pouco	Pouca s	Perto	Longe	Eu
Tu	Eles	Você	Outro	Eles	Comig o	Contig o	Um	Dois

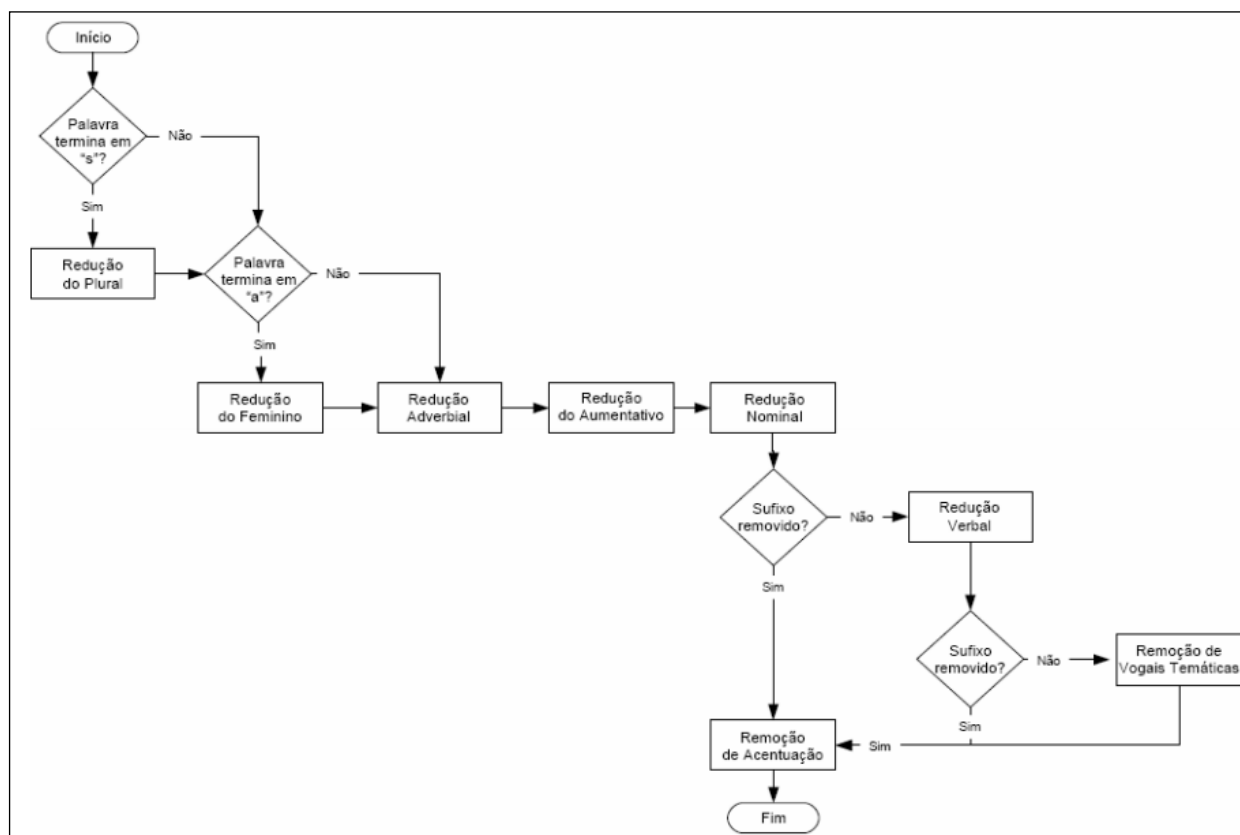
Primeir o	Segund o	Meio	Duplo	Durant e	Neste	Este	Aquel e	Aquilo
Da	Do	Ficar	Permanece r	Que	Ou	Mas	Nem	E

Fonte: Autor (2016)

Na normalização morfológica, segundo Wives (2012), dependendo do caso de mineração de textos, é interessante eliminar as variações morfológicas de uma palavra, isso durante o processo de indexação. Os prefixos e os sufixos das palavras são removidos e os radicais resultantes são adicionados ao índice. A normalização morfológica é uma técnica de identificação de radicais e é denominada de *stemming*, que em inglês significa reduzir uma palavra ao seu radical (ou raiz).

O *stemming*, além de eliminar prefixos e sufixos, características de gênero, número e grau das palavras também são eliminados. Segundo Moraes e Ambrósio (2007), isso resulta em várias palavras sendo reduzidas para um único termo, o que pode reduzir tamanho de um índice em até 50%.

Um algoritmo de *stemming* para língua portuguesa, apresentado por Orengo e Huyck (2001), foi implementado em linguagem C e compõe 8 etapas, conforme a Figura 7. Cada etapa apresenta um conjunto de regras que são examinadas em sequência e apenas uma regra que cada etapa pode ser aplicada.

Figura 7 - Algoritmo de *Stemming* para Língua Portuguesa

Fonte: Orenge e Huyck (2001)

3.1.4. Mineração de Dados (Data Mining)

A quinta etapa consiste na mineração dos dados em si, utilizando um ou mais métodos particulares de mineração de dados. É nesta etapa que são trabalhados os dados realizando uma compactação, classificação, regressão, agrupamento e outros métodos buscando padrões que possam tornar os dados até o momento insignificativos em informações úteis ao objetivo.

3.1.5. Avaliação

A sexta etapa contempla uma análise, ou seja, uma avaliação mais exploratória dos dados gerados na mineração de dados realizada na etapa anterior. Nesta etapa, considerada a etapa de avaliação do conhecimento, é realizada a escolha do algoritmo *datamining* e o (s) método (s) a serem utilizado para a busca de padrões de dados dentro dos resultados, que atendam a organização. A análise em geral requer a presença de um especialista na área de domínio do problema, para resultados mais significativos.

3.1.6. Consolidação do Conhecimento

Nesta etapa, os dados e padrões são interpretados, podendo ser necessário retornar as etapas anteriores para alguma nova iteração. Age-se sobre o conhecimento descoberto, contemplando outros sistemas para apresentá-lo para a parte interessada.

3.2. MINERAÇÃO DE DADOS

A mineração de dados é um componente do KDD e do KDT que envolve uma aplicação repetida de determinados métodos. Das metas de KDT, pode-se definir dois tipos de objetivos, a verificação e a descoberta. Na verificação, limita-se a verificar a hipótese do cliente. Na descoberta, o sistema encontra automaticamente novos padrões.

O termo "minerar" significa extrair da mina (Minérios). É comum encontrar a necessidade de minerar uma base de banco de dados, ou seja, extrair dados de um grande conjunto de dados, gerados de diversas formas e meios. Portanto, a mineração é a etapa da KDT que consiste na aplicação de algoritmos específicos que extraem padrões a partir desses dados, segundo Fayyad et al. (1996), o que é diferente de consulta de dados.

Dessa forma, deve-se selecionar algoritmos aptos para o processo de mineração que auxiliam o processo para que o conteúdo gerado seja útil à organização. A Mineração de Dados possui a capacidade de realizar algumas tarefas, que são associação, classificação, regressão, análise de agrupamentos, sumarização e detecção de desvio.

3.2.1. Métodos de Mineração

Os métodos de Mineração de dados, segundo CIOS et al. (2007) podem ser divididos entre aprendizado supervisionado (preditivo) e não-supervisionado (descritivo). Apesar do limite dessa divisão ser muito tênue (alguns métodos preditivos podem ser descritivos e vice-versa), ela ainda é interessante para fins didáticos (FAYYAD et al. 1996). A diferença entre os dois métodos está no fato de que métodos não-supervisionados não precisam de uma pré-categorização, ou seja, não é necessário um atributo alvo. Tais métodos geralmente usam alguma medida de similaridade entre os atributos, segundo Mccue (2007). As tarefas de agrupamento e associação são consideradas como não-supervisionadas.

No aprendizado supervisionado, segundo Mccue (2007), os métodos possuem um conjunto de dados com uma variável alvo pré-definida e os registros são categorizados em relação a ela. As tarefas mais comuns de aprendizado supervisionado são as de classificação (que também pode ser não-supervisionado) e de regressão.

3.2.2. Associação (*association*)

A associação é a tarefa que tem como objetivo buscar e encontrar padrões dentro de um conjunto de dados ou seja, busca por dados que ocorrem de forma simultânea, como por exemplo: em uma loja de produtos naturais, pode-se identificar que dois produtos geralmente são vendidos juntos, como iogurte e cereal.

Os termos simultâneos auxiliam a remoção de ambiguidade e caracterização de contextos. Essas palavras frequentes e simultâneas podem identificar pontos fortes e fracos de um produto ou serviço por exemplo, auxiliando as empresas na tomada de decisões ligadas ao marketing. Algoritmos, tais como Apriori, GSP, DHP, entre outros, são exemplos de algoritmos que implementam a tarefa de descoberta de associações, segundo Goldshmidt et al. (2015).

3.2.2.1 Algoritmo de Associação - Apriori

O algoritmo Apriori, segundo Vasconcelos e Carvalho (2004) é um dos algoritmos mais conhecidos para mineração por regras de associação. O Apriori busca em profundidade e gera conjuntos de itens padrões. Os padrões não frequentes são excluídos. O Apriori rastreia toda a base de dados e os conjuntos frequentes são obtidos dos conjuntos de itens candidatos. A Figura 8 apresenta o algoritmo Apriori.

Figura 8 - Algoritmo Apriori

Algoritmo 1: Algoritmo Apriori

```

 $F_1 \leftarrow \{\text{Conjuntos de itens freqüentes de tamanho 1}\}$  /* Na
    primeira passagem  $k = 1$  */
1 para  $k = 2$ ;  $F_{k-1} \neq \text{vazio}$ ;  $k++$  faça
    /* Na segunda passagem  $k = 2$  */
2    $C_k \leftarrow \text{apriori-gen}(F_{k-1})$  /* Novos candidatos */
3   para todo transação  $t \in T$  faça
4      $C_t \leftarrow \text{subconjunto}(C_k, t)$  /* Candidatos contidos
        em  $t$  */
5     para todo candidato  $c \in C_t$  faça
6        $c.\text{contagem}++$ 
7     fim
8      $F_k \leftarrow \{c \in C_k | c.\text{contagem} \geq \text{MinSup}\}$ 
9   fim
10 fim
11 Resposta  $F \leftarrow \text{Reunião de todos os } F_k$ 

```

Fonte: Vasconcelos e Carvalho (2009).

O algoritmo gera um conjunto de regras no formato $A \Rightarrow B$ [**suporte, confiança**], o conjunto A é chamado de antecedente da regra e o conjunto B é chamado de consequente. Cada regra gerada deve ter seu suporte e sua confiança maior ou igual ao suporte e confiança mínimo passado para o algoritmo, respectivamente. (PICHIALIANI, 2008).

Para entender melhor as medidas de suporte e confiança, considerando que X e Y são itens de compra, apresenta-se abaixo um exemplo de como é o cálculo do algoritmo:

Suporte ($X \rightarrow Y$): número de registros contendo X e Y / Total de registros.

Confiança ($X \rightarrow Y$): Número de registros contendo X e Y / Número de registros contendo X.

3.2.3. Análise de Agrupamento (*clustering*)

A análise de agrupamento é uma tarefa comum, analisa-se e procura-se identificar um conjunto finito de categorias ou identificam uma classe, conforme Amo (2004). Esta tarefa contempla a busca da similaridade dos dados. A descoberta de grupos homogêneos de clientes, como exemplo, pode ser usada por uma empresa nas estratégias de marketing, conforme FAYYAD et al. (1996).

Na implementação da análise de agrupamento podem ser utilizados algoritmos tais como: K-Means, K-Modes, K-Prototypes, K-Medoids, Kohonen, dentre outros, conforme citado por Goldshmidt et al. (2006).

3.2.3.1 Algoritmo de Agrupamento - K-Means

Segundo Linden (2009), o K-Means é uma heurística de agrupamento não hierárquico que aplica sua técnica para minimizar a distância dos elementos a um conjunto de k centros. O K-Means se inicia com a escolha dos k elementos que formam as sementes iniciais. Após essas sementes são separadas e calculadas a distância de cada elemento, agrupamento esses elementos ao grupo que possuir uma melhor distância, recalculando o centroide. A Figura 9 apresenta o algoritmo K-Means.

Figura 9 - Algoritmo K-Means

Algoritmo 1: O algoritmo k-means**Entrada:** $X = \{x_1, x_2, \dots, x_n\}$: conjunto de documentos k : número de grupos**Saída:** $P = \{G_1, G_2, \dots, G_k\}$: partição com k grupos

```

1 selecionar aleatoriamente  $k$  documentos como centróides
  iniciais;
2 repita
3   para cada documento  $x \in X$  faça
4     computar a (dis)similaridade de  $x$  para cada
      centroide  $C$  ;
5     atribuir  $x$  ao centroide mais próximo ;
6   fim
7   recomputar o centroide de cada grupo;
8 até atingir um critério de parada;

```

Fonte: Rezende et al. (2011).

3.3. FERRAMENTAS PARA MINERAÇÃO DE DADOS

Existem no mercado diversas ferramentas disponíveis para a realização da mineração de dados. Devido ao crescimento da necessidade de minerar os dados, o mercado de ferramentas se tornou atraente. A Tabela 4 apresenta algumas ferramentas que pode-se trabalhar na mineração e análise de dados, bem como suas características.

Tabela 4 - Ferramentas de Mineração de Dados

Ferramenta	Informações
<i>Clementine</i>	Desenvolvida pela SPSS, suporta CRISP-DM e possui outras facilidades. É uma ferramenta madura de mineração de dados. (PRYKE, 20018).
WEKA	É uma ferramenta livre e considerada uma das melhores do mercado. Disponibiliza diversos algoritmos para a realização das tarefas de mineração, com diversas formas de visualização. (WEKA)
MDR	Ferramenta em Java, de software livre, para detecção de interações entre atributos utilizando o método da <i>multifactor dimensionality reduction</i> (MDR). (MOORE, 2017)

Fonte: Autor (2016).

4. TRABALHOS RELACIONADOS

Neste capítulo serão apresentados dois trabalhos relacionados à análise de dados nas redes sociais e ao marketing digital. Um deles apresenta um estudo realizado na rede social *Facebook* para identificar uma estratégia de marketing para o banco Bradesco S.A., e outro um estudo de perfil de usuário de redes sociais (*Facebook*) para mineração de dados.

4.1. A UTILIZAÇÃO DAS REDES SOCIAIS: UM ESTUDO DE CASO SOBRE O *FACEBOOK* COMO ESTRATÉGIA DE MARKETING NO BANCO BRADESCO S.A

O trabalho proposto por Ferreira e Neves (2013) teve como objetivo analisar os benefícios do Marketing *One-to-One* realizado por meio de relacionamentos e atendimentos realizados pelas redes sociais. Neste trabalho, eles focaram no *Facebook*.

Neste trabalho, a rede social do Bradesco S/A foi analisada durante 3 meses e foram desenvolvidas duas pesquisas, realizadas com dois perfis diferentes. Primeiramente com o gerente de redes sociais do banco e após com alguns usuários, para que se fizesse um levantamento de dados necessários para o estudo.

Os responsáveis decidiram efetuar a coleta de dados por meio de pesquisa quantitativa e a página foi avaliada em um período pré-definido para se levantar os pontos fortes e pontos fracos das postagens e dos atendimentos.

Fonseca relata que “[...] a pesquisa quantitativa se centra na objetividade. [...] recorre à linguagem matemática para descrever causas de um fenômeno, as relações entre variáveis, etc.” (2002, p. 20).

Entre os problemas encontrados durante o estudo, o maior deles foi a falta de um profissional de marketing que cuidasse das redes sociais do banco. A falta de divulgação de informações e novidades na página faz com que a maioria dos entrevistados durante a pesquisa não acessassem a rede do banco e não interagissem com a instituição, assim

não disponibilizando ao banco conquistar mais clientes por intermédio da rede social e nem mesmo efetuar melhorias, se necessário.

Para aplicar o marketing *One-to-One*, foi então analisado o perfil de cada usuário, efetuando uma coleta de dados básica, como idade, gênero, município, se é ou não correntista e o que levou a curtir a página. Com esses dados, será possível identificar qual o perfil do cliente que utiliza a rede social. Após, foi realizada uma pesquisa de satisfação com os clientes com o intuito de avaliar o contentamento dos internautas, com perguntas como, se foram atendidos no prazo, se encontraram todas informações necessárias, sugestões de melhorias. Com esses dados tabulados foi possível identificar os diferentes tipos de clientes que curtem a página do Bradesco.

Diferenciar os clientes torna possível ao Bradesco identificar que tipo de marketing *One-to-One* deverá ser aplicado à página. Não será possível usar o mesmo critério sempre, pois os clientes possuem necessidades diferentes de utilização de serviços.

Com base no estudo feito, pode-se identificar que a página do *Facebook* do banco Bradesco tem grande poder de influência na imagem da instituição, tanto positivamente, quanto negativamente.

Após este resultado, ficou claro para a instituição que o marketing *One-to-One* é necessário para auxiliar a empresa na implantação de um novo canal de comunicação com o cliente, a Rede Social. Foi provado também que esse trabalho é extremamente importante para o Bradesco e ficou claro que o marketing *One-to-One*, ou seja, o marketing feito diretamente com o cliente, conhecendo seu perfil, traz enormes ganhos à empresa, ainda mais quando este é positivo.

4.2. ESTUDO DE PERFIL DE USUÁRIOS DE REDES SOCIAIS ATRAVÉS DA MINERAÇÃO DE DADOS

Este projeto foi desenvolvido na Universidade de Caxias do Sul (UCS), por Puhl (2011), orientada pela Professora Doutora Helena G. Ribeiro.

O foco deste trabalho foi a utilização das práticas de mineração de dados na página do *Facebook* da empresa Promob com o intuito de conhecer o cliente que frequenta as mídias sociais da empresa para auxiliar o marketing das campanhas de relacionamento com o cliente.

Foi realizada a extração de dados por meio da API do *Twitter*, chamado *API Graph*. Para mineração dos dados, foi utilizada a ferramenta WEKA, após, foi realizado um pré-processamento dos dados e foi realizada uma seleção e limpeza dos dados.

Puhl (2011) utilizou em seu trabalho os algoritmos J48 para classificação dos dados, o que apresenta os mesmos em forma de árvore de decisão. A exatidão deste classificador é uma característica de Kappa, que apresenta os dados como fraco, regular, moderado, bom e excelente. A Tabela 5 apresenta um exemplo dos dados gerados e apresentados em forma de tabela.

Tabela 5 - Probabilidade de Kappa

Probabilidade de Kappa	Exatidão do Classificador
Menor que 0.20	Fraco
0.21 – 0.40	Regular
0.41 – 0.60	Moderado
0.61 – 0.80	Bom
Maior que 0.81	Excelente

Fonte: Adaptado de Puhl (2011).

Na análise dos resultados gerados no trabalho foi identificado que, pelos dados serem originados de redes sociais, muitos atributos necessários para as análises estão incompletos ou preenchidos incorretamente, além de que, devido ao grande número de informações vazias, ou seja, não informadas, os resultados podem estar comprometidos.

Puhl (2011), também ressalta que, “a cada dia entram novos perfis, o que ocasionaria um trabalho de limpeza de dados e transformação a cada vez que tenha necessidade de extrair os dados. ”

Para trabalhos futuros, foi então sugerido a criação de um aplicativo de forma de pesquisa com os usuários, com dados pré-definidos para a criação de uma base de dados mais consolidada.

5. ESTUDO DE CASO

Este estudo de caso tem como proposta analisar os dados gerados em redes sociais com o intuito de auxiliar a empresa ALLPS Alimentos Saudáveis a tomar decisões com relação ao investimento em Marketing Digital. O Marketing Digital da empresa é realizado em sua *fanpage*, no *Facebook*. A ALLPS Alimentos necessita conhecer melhor seu público-alvo para entender como aplicar o marketing *One-to-One* com seus clientes. Por meio da análise destes dados, busca-se entender que tipo de produtos a empresa precisa divulgar em seus meios de comunicação e comercializar no seu produto, para atender melhor seus clientes. Necessita-se também conhecer o mercado para que se possa atingir um público maior de pessoas que buscam por uma alimentação saudável.

Os dados coletados são relacionados a alimentação, termos utilizados em produtos considerados saudáveis e comentados nas redes sociais. A mineração destes dados teve como objetivo identificar os termos mais presentes nos comentários coletados, definindo restrições e produtos muito procurados atualmente, e que são utilizados de forma combinada e significativa para a empresa.

5.1. ALLPS ALIMENTOS SAUDÁVEIS

A ALLPS Alimentos Saudáveis é uma empresa inovadora, que trabalha com um portal de assinaturas, voltado a um público que possui alguma restrição alimentar, como intolerância à lactose, glúten ou simplesmente a quem deseja ter uma alimentação saudável.

Impulsionada pelas sócias Camila Colombo e Camila Remonato que possuíam dificuldades em encontrar produtos específicos para restrição a lactose e ao mesmo tempo saborosos, a ALLPS foi fundada em 2015, juntando sabor, comodidade e saúde em um mesmo kit.

A ALLPS realiza uma pesquisa criteriosa sobre os produtos que são enviados em seus kits. Todos os assinantes recebem de 8 a 12 produtos mensalmente, em suas casas, conforme dieta selecionada no momento da assinatura. Todos os produtos

disponibilizados são avaliados por uma nutricionista, para que seus clientes tenham acesso a produtos de qualidade e seguros para sua saúde.

A ALLPS Alimentos tem como visão, ser a melhor opção em alimentos saudáveis, proporcionando bem-estar, saúde e comodidade para todos. Sua missão é dar as pessoas a oportunidade de conhecer produtos de qualidade, específicos para sua saúde, proporcionando prazer em comer bem. A ALLPS tem como seus valores o respeito em entender que todas as pessoas possuem diferenças e devem ser tratadas com integridade. A empresa quer oferecer a seus clientes qualidade de vida, buscando a melhoria contínua dos seus produtos, superando suas expectativas.

Esse estudo de caso pode auxiliar a ALLPS com informações significativas que podem auxiliar na implantação de um Marketing Digital mais efetivo. A ALLPS tem focado seus trabalhos no Marketing Digital, por se tratar de uma empresa *web*, ou seja, todo seu trabalho de divulgação, venda e entrega de produto é feita de forma *on-line*. Devido a isso, a ALLPS possui como base o marketing *One-to-One*, para se trabalhar um marketing direto e focado nos seus clientes e suas necessidades. Além disso, a empresa pretende melhorar sua divulgação, aplicando a técnica de *pull*, para que o público que efetua pesquisas buscando por alimentação saudável veja a ALLPS como um produto que atenda suas necessidades.

A ALLPS possui uma atividade ativa nas redes sociais, como *Twitter*, *Instagram* e *Facebook*, possuindo em torno de 6692mil seguidores no *Instagram* onde realiza portagens diariamente e possuem uma página no *Facebook* com 4321mil curtidas. Apesar de não possuir uma atividade significativa no *Twitter*, as marcas dos principais produtos vendidos pela empresa tuitam frequentemente, gerando assim dados sobre os produtos que poderão ser utilizados neste estudo de caso.

5.2. PLANO DE TRABALHO E IMPLANTAÇÃO

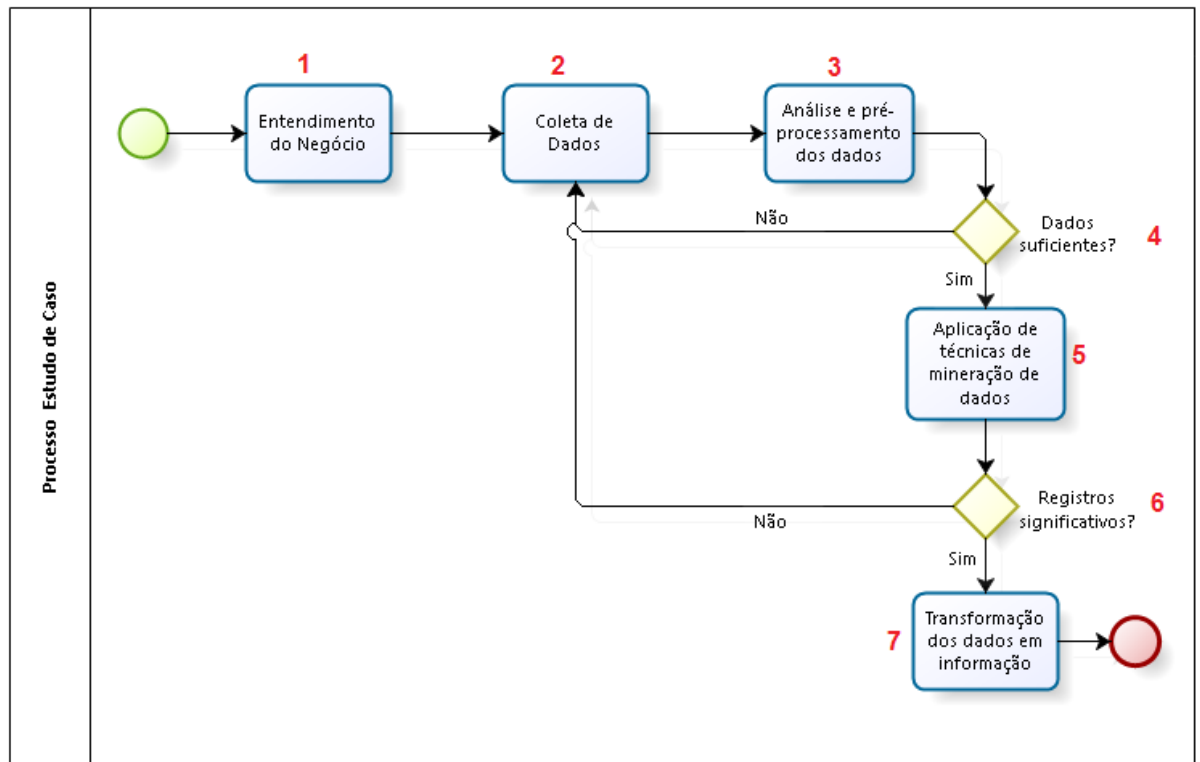
Para a implantação deste estudo de caso foi desenvolvido um plano de trabalho, conforme apresentado na Figura 10. Na etapa de Entendimento do Negócio (1), realizou-se uma pesquisa juntamente com a empresa ALLPS Alimentos para identificar as palavras-chave que foram utilizadas nas etapas de coleta e limpeza dos dados.

Na etapa de Coleta de Dados (2), utilizou-se ferramentas que efetuam a coleta de dados do *Twitter* por meio de API's próprias da rede social. Na etapa de Análise e Pré-Processamento dos Dados (3), realizou-se uma análise e limpeza dos dados coletados. Se necessário efetuar nova coleta, devido aos dados serem insuficientes, conforme apresentado pelo losango "Dados Suficientes?" (4), volta-se a etapa de coleta de dados do processo.

Após, a Aplicação de Técnicas de Mineração de Dados (5) se fazem necessárias. Para essa etapa, utilizou-se a ferramenta WEKA, no qual aplicou-se métodos de mineração com as regras de associação e agrupamento, aplicando algoritmos de mineração de dados, como Apriori e K-Means.

A aplicação destas regras de mineração de dados possui o intuito de trabalhar os dados coletados transformando-os em informações. Após, é possível analisar se os dados minerados são significativos, conforme o losango "Registros Significativos?" (6), caso negativo, uma nova coleta de dados se faz necessária. Após essa etapa, os resultados gerados pela aplicação dos algoritmos são transformados em informações (7), podendo ser apresentados a empresa por meio de grafos, gráficos, tabelas, sendo essa, a última etapa do processo.

Figura 10 - Plano de Trabalho



Fonte: Autor (2016).

5.3. ENTENDIMENTO DO NEGÓCIO

Para o desenvolvimento deste estudo de caso, foi realizada uma pesquisa juntamente com a empresa ALLPS, presente no Anexo C, para efetuar o levantamento das principais palavras-chave que foram buscadas na coleta de dados realizada.

Além da pesquisa, foi realizado um trabalho de construção da Análise SWOT. Conforme Bastos (2015), o termo SWOT¹ significa *Strengths, Weakness, Opportunities and Threats*, que em português significam respectivamente: Forças, Fraquezas, Oportunidades e Ameaças. A construção da análise foi realizada juntamente com a

¹ Ferramenta estrutural da administração, que possui como principal finalidade avaliar os ambientes internos e externos, formulando estratégias de negócios para a empresa com a finalidade de otimizar seu desempenho no mercado.

empresa, para conhecer o planejamento estratégico da empresa e comparar com os resultados deste trabalho.

A Análise SWOT, segundo Medeiros (2016), consiste em construir um sistema de quadrantes como a Figura 11, identificando no Ambiente Interno quais são as Forças e Fraquezas da empresa e no Ambiente Externo as Oportunidades e Ameaças.

Figura 11 - Análise SWOT



Fonte: Medeiros (2015).

A ALLPS ainda não possuía um planejamento estratégico completo. Foram realizadas algumas perguntas para auxiliá-los no entendimento da análise para preenchimento dos quadrantes. Na Tabela 6 apresenta-se as perguntas realizadas.

Tabela 6 - Perguntas para análise SWOT

Forças	Fraquezas
Quais vantagens sua organização tem?	No que sua organização pode melhorar?
Em que é melhor que a concorrência?	O que seus clientes não apreciam em sua organização?
O que é visto como força no mercado?	O que já o fez perder clientes?
Seus clientes apreciam sua organização?	
Quais produtos e serviços da sua organização faz a diferença?	
Oportunidades	Ameaças
Quais as tendências do mercado?	Há novos produtos ainda não utilizados pela empresa?
Há previsões de mudanças no ramo da empresa?	Há novos concorrentes?
	Há mudanças de tecnologia que possam afetar o processo da empresa?
	Há regras do governo que podem afetar a empresa?

Fonte: Autor (2016).

Na Figura 12 apresenta-se o resultado deste trabalho, mediante a Análise SWOT, que foi construído e analisado para auxiliar na análise e conclusão deste estudo de caso.

Figura 12 - Análise SWOT - ALLPS

	Ajuda	Atrapalha
Interna	Forças: • Mapeamento frequente da concorrência • Agilidade no atendimento ao consumidor • Pró atividade da equipe • Produtos de Qualidade • Blog com informações e receitas • Identidade visual atrativa • Mais opções de pagamento (boleto) • Controle de qualidade com nutricionista • Foco em produtos sem Glúten e Lactose	Fraquezas: • Não ter conhecimento total do mercado de atuação • Poucos recursos financeiros para investimento • Baixa prospecção presencial • Indisponibilidade de horário • Processo de entrega (Alto custo e prazo longo) • Dificuldade de contato com fornecedores • Site não responsivo
Externa	Oportunidades: • Possibilidade de abrir um e-commerce • Crescimento do mercado de alimentação saudável • E-commerce já é um espaço de compras consolidado • Número alto de pessoas com restrições alimentares	Ameaças: • Concorrência forte • Carga Tributária elevada para RS • Custo do Frete • Crise

Fonte: Autor (2016).

PARAGRAFO DE CONEXÃO PARA TROCA DE ASSUNTO

Na Tabela 7, apresenta-se as palavras-chave levantadas na pesquisa e que foram capturadas da base de dados do *Twitter* e *Instagram*. Os dados do *Twitter* foram capturados dos campos *text* (Texto do *Tweet*) e *entities* (diversas informações, como por exemplo, *hashtags*, *urls*, etc.).

Tabela 7 - Processamento de Língua Natural

Lactose	Gluten	Semlactose	Semgluten
Zerolactose	Zerogluten	Diet	Light
Semleite	Semovos	Semacucar	Diabete
Celiaco	IntoleranciaaLactose	Intoleranciaaogluten	AlimentcaoSaudavel
Diabetico			

Fonte: Autor (2016).

Além do levantamento das palavras acima, foi identificado juntamente com a ALLPS quais os tipos de doenças/restrições mais ocorrem de forma conjunta, no conhecimento da empresa. Além das combinações, a empresa informou que Glúten e Lactose são as restrições mais trabalhadas pela empresa, e que seus clientes mais procuram. Essas combinações são apresentadas na Tabela 8.

Tabela 8 - Combinações mais conhecidas pela ALLPS

Glúten e Lactose	Glúten e Diabetes
------------------	-------------------

Fonte: Autor (2016).

Alguns alimentos, conforme Tabela 9, foram comentados pela empresa como os mais solicitados e ou procurados pelos seus clientes. Com estes dados, foi possível identificar se a ALLPS está realizando um marketing efetivo, que atende as necessidades dos clientes, ou se necessita reajustar seus produtos e, em consequência, seu Marketing Digital.

Tabela 9 - Alimentos ALLPS

Doce de Leite	Bolos	Massas
Barras de Cereais	Biscoitos	Bebidas

Fonte: Autor (2016).

Após o levantamento dos dados para entendimento do negócio, foi dado início da coleta dos dados nas redes sociais.

5.4. COLETA DE DADOS

Para a coleta de dados foram utilizadas algumas ferramentas, algumas delas já conhecidas no mercado. Primeiramente foi utilizado um programa na linguagem Java, desenvolvido pelo aluno da Universidade de Caxias do Sul, Hoffman (2016), após foram utilizado também os programas Nodexl e Netlytic.

Nos próximos capítulos, apresenta-se como foram realizadas a coleta de dados em cada ferramenta citada acima.

5.4.1. Aplicação em JAVA

O programa desenvolvido por Hoffman (2016), realiza uma conexão HTTP à *timeline* do *Twitter*. Este programa utiliza a página de busca do próprio site do *Twitter* para coleta, que também utiliza as API's para a busca.

O programa original possui alguns parâmetros utilizados para captura dos *tweets*, conforme a Tabela 10.

Tabela 10 - Campos Twitter

Parâmetro	Descrição
<i>username</i>	Nome do usuário do <i>Twitter</i> .
<i>since</i>	Data a partir da qual serão buscadas as mensagens.
<i>until</i>	Data até a qual serão buscadas as mensagens.
<i>near</i>	Localidade do usuário.
<i>within</i>	Expand a coleta para X milhas próximas à localidade escolhida no parâmetro <i>near</i> .
<i>querySearch</i>	Palavras que devem constar nas mensagens a serem coletadas.

Fonte: Hoffman (2016).

Para este estudo de caso, foram alteradas algumas configurações do programa para obter um retorno mais efetivo dos dados necessários. Na tela inicial, foi incluído um campo *Max. Tweet*, no qual é possível preencher a quantidade de *tweets* desejados para retorno, tendo um limite de 10.000 *tweets* para não perder a performance do programa conforme Figura 13.

Figura 13 - Aplicação Java – Max. Tweets

Coletor de Tweets

Usuario: Inicio: ... ate: ... Buscar

Palavras:

Local: Proximo: milhas ☐ Aplicar algoritmo Stemmer Max. Tweets: 10000

Limpar Exportar Sair

Campo de limite de Tweets

Fonte: Aplicação Java (2016).

Foram realizadas coletas dos *tweets* realizados entre as datas de 01/01/2015 até 30/09/2016, contendo as palavras apresentadas na Tabela 6 desde estudo de caso.

Após algumas coletas, foi identificado um grande número de *tweets* em inglês e espanhol, devido a busca de palavras como “*Diet*” e “*Light*”. Para que as buscas não retornassem esse tipo de mensagem, foi alterado no programa a classe *TwitterHttpConnect* que realiza a conexão com o *Twitter*, para que o programa retorne apenas os *tweets* da língua portuguesa. Este é consultado no campo *lang* do *tweet*, e é identificado como pt. O programa possui um parâmetro de geolocalização, identificado no campo “Local”, mas este não foi utilizado neste estudo de caso, pois o programa procura por proximidade do local e não de local informado, que neste caso seria o Brasil.

A aplicação Java também possui a opção de aplicar o algoritmo stemmer antes da busca, o que já retorna os dados com a remoção de sufixos comuns das palavras retornadas. Após alguns testes, foi identificado que a aplicação do algoritmo stemmer neste estudo de caso foi desnecessário. Neste estudo de caso, a coleta foi realizada

apenas com substantivos, ou seja, as palavras-chave trabalhadas não sofriam alterações nos *tweets* pesquisados, para que fosse necessário aplicar o algoritmo para efetuar a busca.

O programa também possui os campos Início/Ate e Palavras, utilizados para a coleta de dados deste trabalho. No campo “Palavras” foram adicionadas as palavras-chave para que essas fossem pesquisadas no campo *text* do *tweet*. Na Figura 14, é possível visualizar a forma como os dados são retornados após clicar no botão “Buscar”.

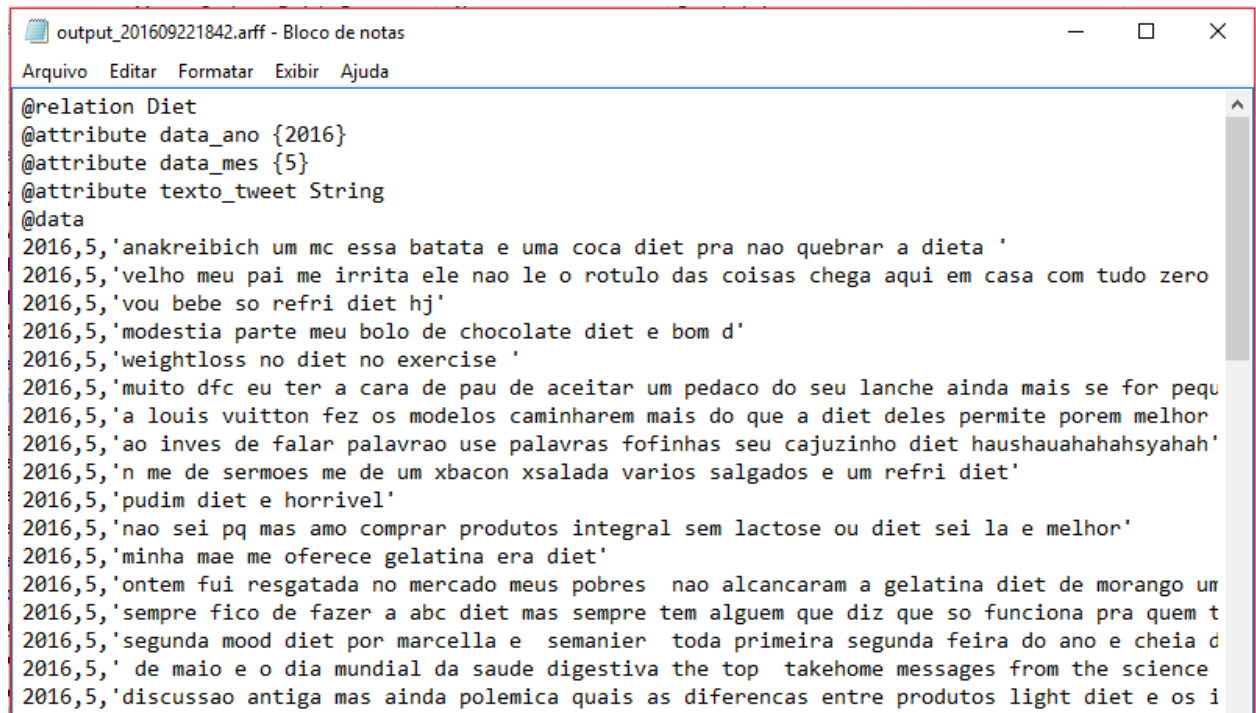
Figura 14 - Tela Inicial - Aplicação Java

Fonte: Aplicação Java (2016).

Depois da busca, é possível clicar no botão “Exportar”. Este gera um arquivo de extensão ARFF com os dados coletados. O arquivo ARFF é o único formato de arquivo reconhecido no programa WEKA, que foi utilizado para análise dos dados.

O arquivo é gerado automaticamente com os seguintes atributos: `data_ano` (ano do *tweet*), `data_mes` (mês do *tweet*), `texto_tweet` (mensagem do *tweet*). Na Figura 15 apresenta-se um exemplo do arquivo gerado inicialmente.

Figura 15 - Arquivo ARFF



```

output_201609221842.arff - Bloco de notas
Arquivo  Editar  Formatar  Exibir  Ajuda

@relation Diet
@attribute data_ano {2016}
@attribute data_mes {5}
@attribute texto_tweet String
@data
2016,5,'anakreibich um mc essa batata e uma coca diet pra nao quebrar a dieta '
2016,5,'velho meu pai me irrita ele nao le o rotulo das coisas chega aqui em casa com tudo zero
2016,5,'vou bebe so refri diet hj'
2016,5,'modestia parte meu bolo de chocolate diet e bom d'
2016,5,'weightloss no diet no exercise '
2016,5,'muito dfc eu ter a cara de pau de aceitar um pedaco do seu lanche ainda mais se for pequ
2016,5,'a louis vuitton fez os modelos caminharem mais do que a diet deles permite porem melhor
2016,5,'ao inves de falar palavrão use palavras fofinhas seu cajuzinho diet haushauahahahsyahah'
2016,5,'n me de sermoes me de um xbacon xsalada varios salgados e um refri diet'
2016,5,'pudim diet e horrivel'
2016,5,'nao sei pq mas amo comprar produtos integral sem lactose ou diet sei la e melhor'
2016,5,'minha mae me oferece gelatina era diet'
2016,5,'ontem fui resgatada no mercado meus pobres  nao alcançaram a gelatina diet de morango um
2016,5,'sempre fico de fazer a abc diet mas sempre tem alguem que diz que so funciona pra quem t
2016,5,'segunda mood diet por marcella e semanier toda primeira segunda feira do ano e cheia d
2016,5,' de maio e o dia mundial da saude digestiva the top takehome messages from the science
2016,5,'discussao antiga mas ainda polemica quais as diferencas entre produtos light diet e os i
  
```

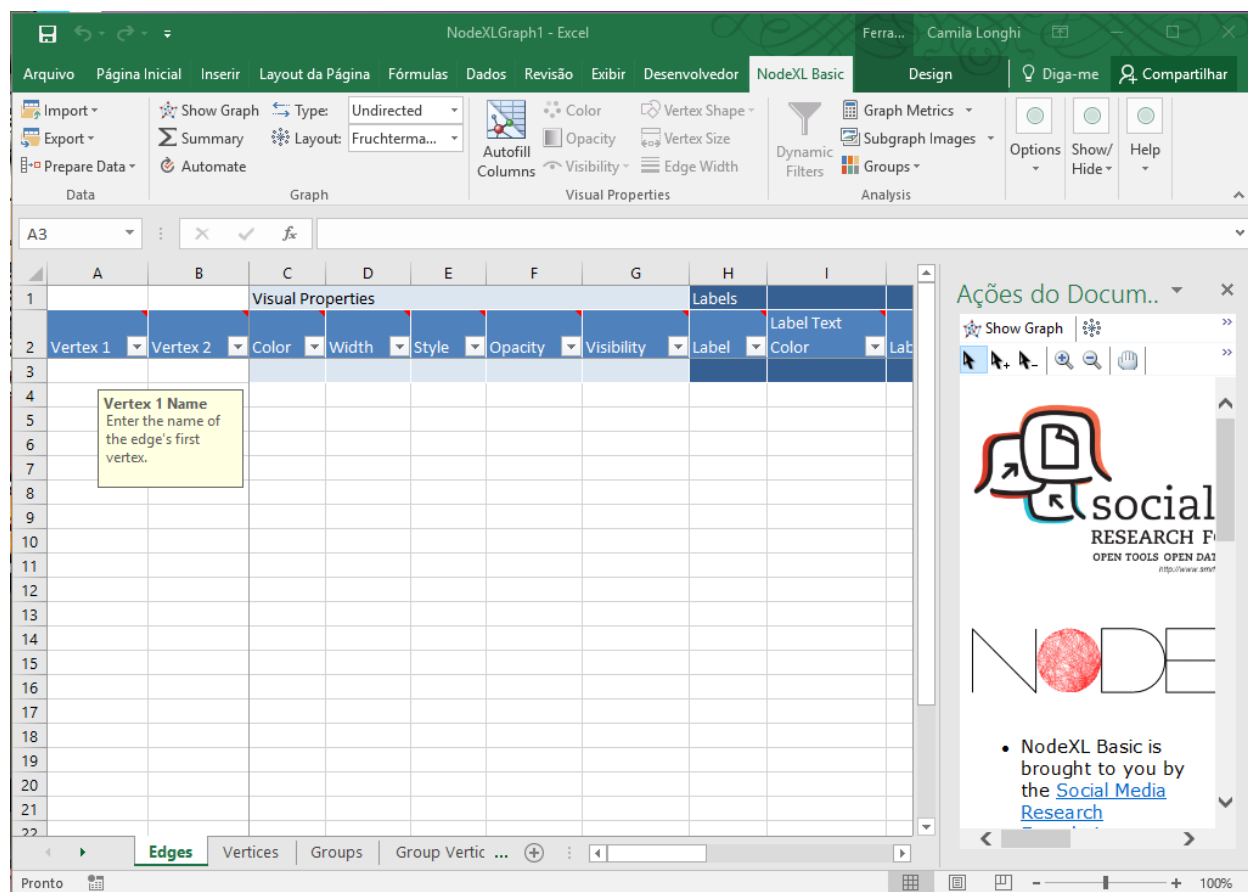
Fonte: Autor (2016).

5.4.2. NodeXL

O programa NodeXL é um aplicativo do Excel, conforme Figura 16, que permite importar dados de diversas redes sociais, como *Facebook*, *Flicker*, *Twitter* e *YouTube*. Para este estudo de caso foram coletados dados do *Twitter*. Na Figura 16 e 17 apresenta-se a interface do NodeXL e os passos necessários para efetuar a captura de dados. Os *tweets* coletados nessa ferramenta foram buscados por meio das palavras “*Diet*”, “*Light*”, “*Gluten*” e “*Lactose*”. As palavras Glúten e Lactose foram escolhidas pela ALLPS como as mais importantes para o seu negócio. As palavras *Diet* e *Light* foram

consideradas pela ALLPS como de grande importância para uma alimentação saudável, mas a empresa desconhece a abrangência deste tipo de restrição com relação ao mercado.

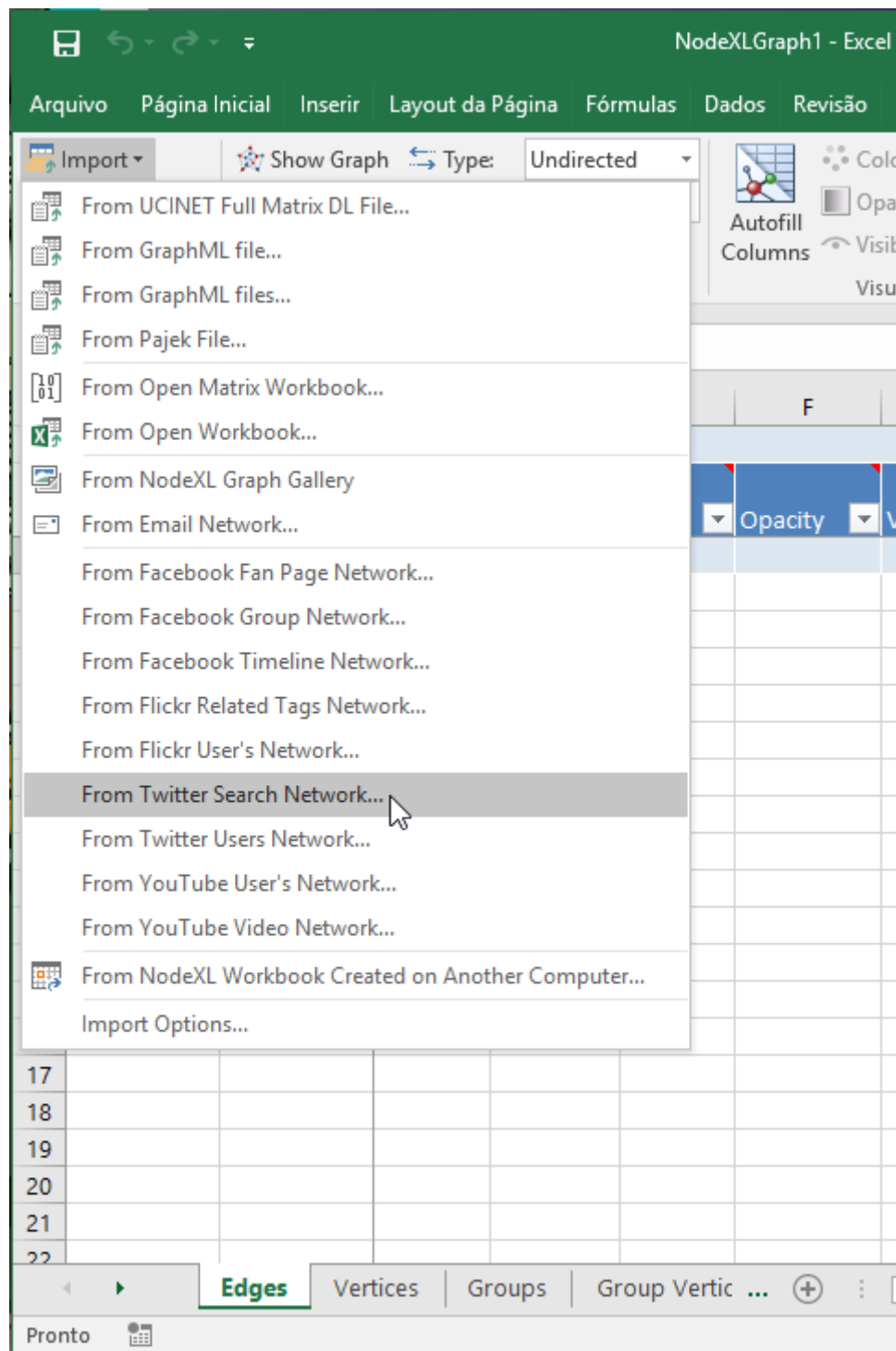
Figura 16 - Tela Inicial NodeXL



Fonte: NodeXL (2016).

Para efetuar a busca de *tweets* é necessário ir em *Import > From Twitter Search Network.*, conforme Figura 17.

Figura 17 - Captura de Tweets



Fonte: NodeXL (2016).

Na janela de importação é possível incluir as palavras desejadas para a busca, podendo ser utilizados comandos como “OR” ou “AND”, para a busca de mais palavras-

chave. Neste estudo de caso, foi utilizado o comando “*Lang:pt*” para que o retorno de dados seja apenas de *tweets* em português.

O limite de *tweets* retornados no NodeXL é de 2.000, na versão *free* do programa. É necessário também possuir uma conta no *Twitter* e autorização a busca dos dados, conforme Figuras 18, 19, 20 e 21.

Figura 18 - Importação de Dados

Import from Twitter Search Network

[Upgrade to NodeXL Pro to expand the number of tweets you can download from 2,000 to the maximum allowed by Twitter, 18,000 \(or 7-8 days, whichever comes first.\)](#)

[This might take a long time: Twitter rate limiting](#)

Search for tweets that match this query:

Lactose lang: pt

[How to use advanced search operators](#)

What to import

- ☒ Basic network
Show who was replied to or mentioned in recent tweets
[More about this option](#)
- ☐ Basic network plus friends (very slow!)
Add some of the users' friends
[More about this option](#)

Your Twitter account

- ☐ I have a Twitter account, but I have not yet authorized NodeXL to use my account to import Twitter networks. Take me to Twitter's authorization Web page.
- ☒ I have a Twitter account, and I have authorized NodeXL to use my account to import Twitter networks.

Limit to 2,000 tweets

☒ Expand URLs in tweets (slower)

OK Cancel

The diagram illustrates a network structure where nodes represent users and edges represent relationships. Labels include 'Replies To' and 'Mentions', showing how tweets are linked through these interactions.

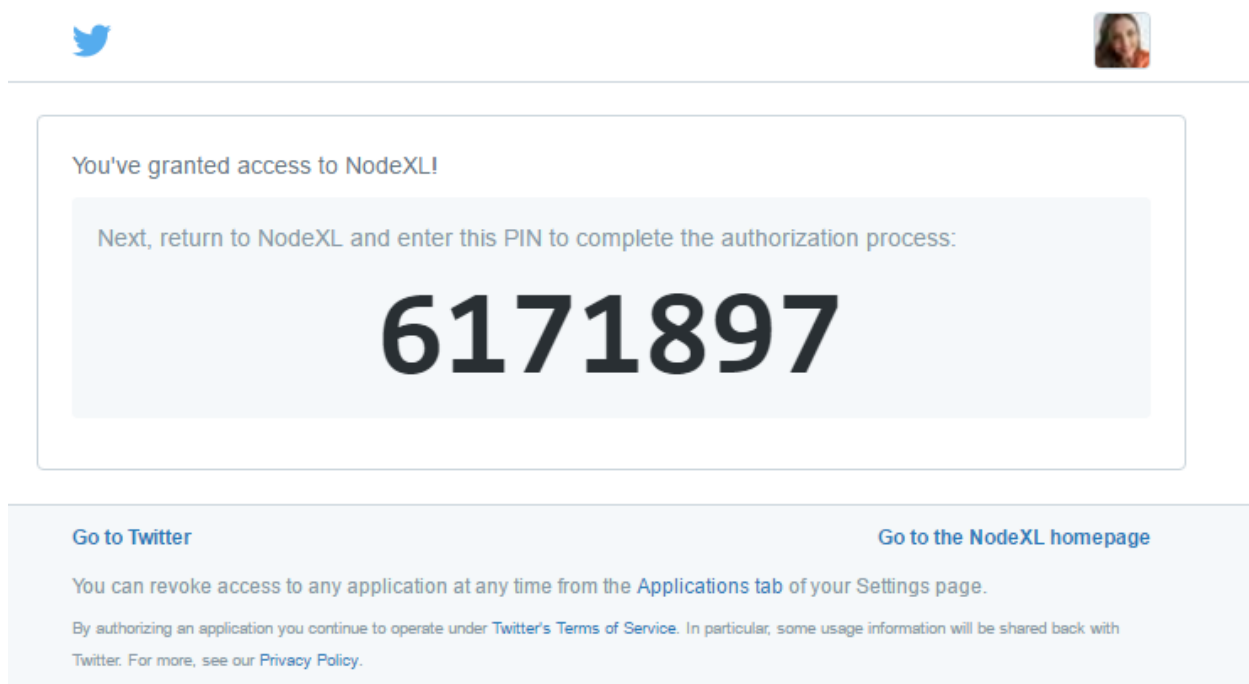
Fonte: NodeXL (2016).

Figura 19 - Autorização Twitter



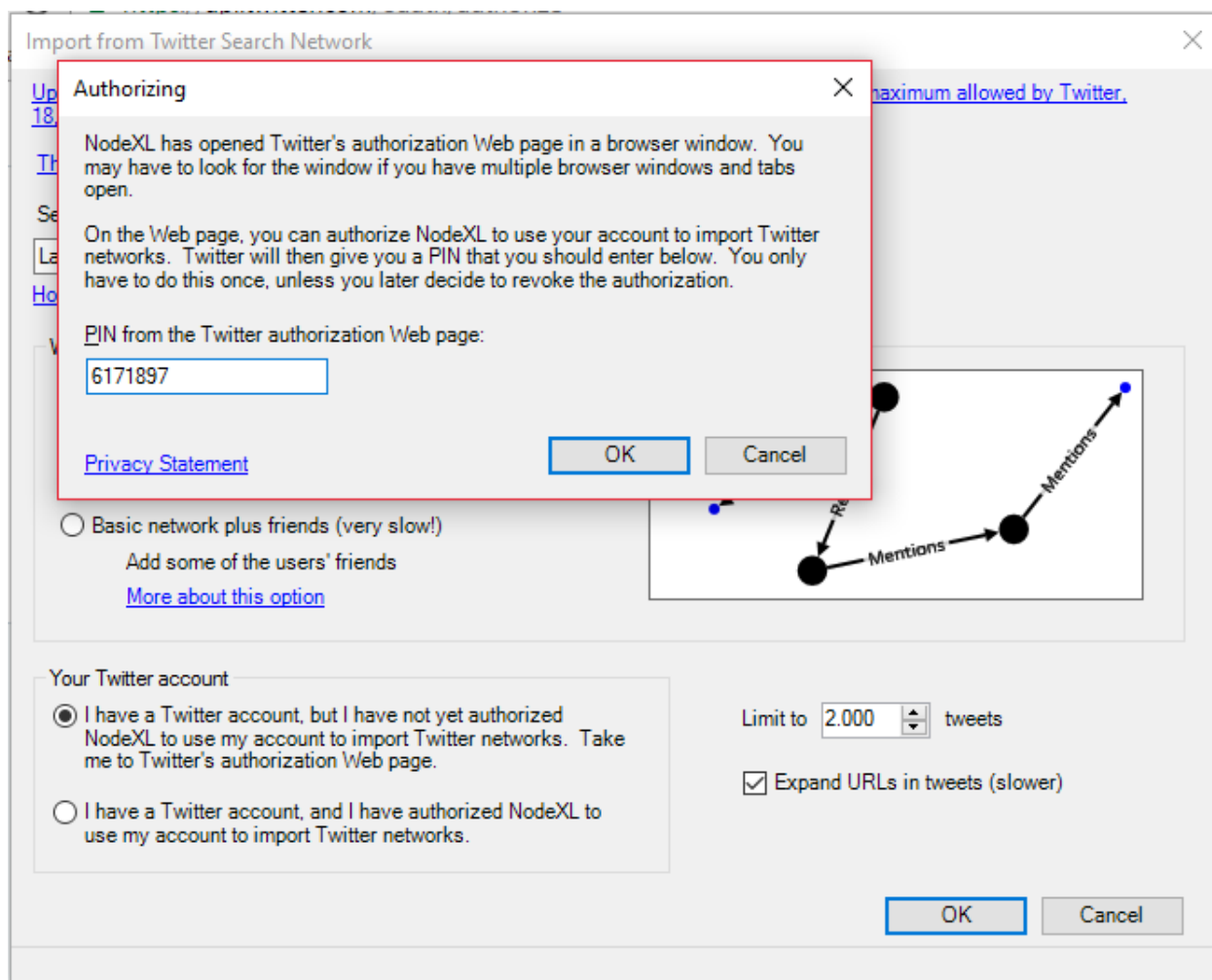
Fonte: NodeXL (2016).

Figura 20 - Código de Autorização



Fonte: NodeXL (2016).

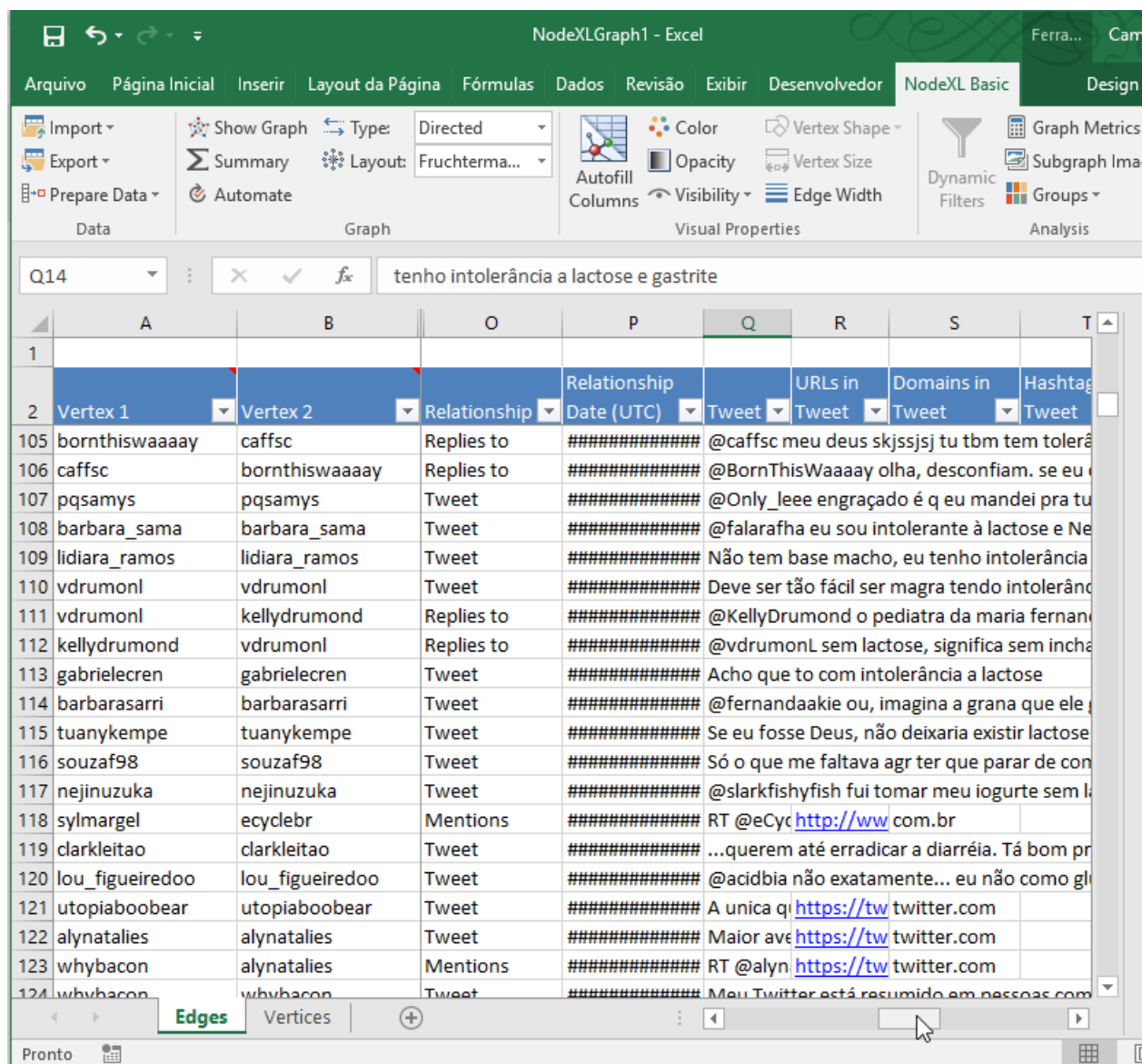
Figura 21 - Autorização NodeXL



Fonte: NodeXL (2016).

Os dados retornados são apresentados em duas planilhas: uma chamada de *Edges* (Nós) e outra de *Vertices*. Na planilha de *Edges* são apresentados todos os *tweets* retornados, bem como os relacionamentos deste *tweet*, se é um *tweet* simples, um *retweet* ou respostas ao *tweet*. Nas colunas *Vertex 1* e *Vertex 2* é apresentado o usuário que *tweetou* e o usuário que *retweetou* ou respondeu, respectivamente, conforme Figura 22.

Figura 22 – Dados retornados Nós.



	A	B	O	P	Q	R	S	T
	Vertex 1	Vertex 2	Relationship	Date (UTC)	Tweet	URLs in	Domains in	Hashtag
105	bornthiswaaaay	caffsc	Replies to	#####	@caffsc meu deus skjssjsj tu tbm tem tolerã			
106	caffsc	bornthiswaaaay	Replies to	#####	@BornThisWaaaay olha, desconfiam. se eu c			
107	pqsamys	pqsamys	Tweet	#####	@Only_leea engraçado é q eu mandei pra tu			
108	barbara_sama	barbara_sama	Tweet	#####	@falarafha eu sou intolerante à lactose e Ne			
109	lidiara_ramos	lidiara_ramos	Tweet	#####	Não tem base macho, eu tenho intolerância			
110	vdrumonl	vdrumonl	Tweet	#####	Deve ser tão fácil ser magra tendo intolerânc			
111	vdrumonl	kellydrumond	Replies to	#####	@KellyDrumond o pediatra da maria fernan			
112	kellydrumond	vdrumonl	Replies to	#####	@vdrumonL sem lactose, significa sem incha			
113	gabrielecren	gabrielecren	Tweet	#####	Acho que to com intolerância a lactose			
114	barbarasarri	barbarasarri	Tweet	#####	@fernandaakie ou, imagina a grana que ele			
115	tuanykempe	tuanykempe	Tweet	#####	Se eu fosse Deus, não deixaria existir lactose			
116	souzaf98	souzaf98	Tweet	#####	Só o que me faltava agr ter que parar de con			
117	nejinuzuka	nejinuzuka	Tweet	#####	@slarkfishyfish fui tomar meu iogurte sem l			
118	sylmargel	ecyclebr	Mentions	#####	RT @eCyc http://ww.com.br			
119	clarkleitao	clarkleitao	Tweet	#####	...querem até erradicar a diarreia. Tá bom pr			
120	lou_figueiredoo	lou_figueiredoo	Tweet	#####	@acidbia não exatamente... eu não como gl			
121	utopiaboobear	utopiaboobear	Tweet	#####	A unica q https://tw.twitter.com			
122	alynatalies	alynatalies	Tweet	#####	Maior ave https://tw.twitter.com			
123	whybacon	alynatalies	Mentions	#####	RT @alyn https://tw.twitter.com			
124	whybacon	whybacon	Tweet	#####	Meu Twitter está resumido em pessoas com			

Fonte: Autor (2016).

Na planilha *Vertices* são apresentados dados de cada *tweet* capturado, como o usuário responsável pelo *tweet*, *retweet* ou resposta, número de seguidores, número de *tweets* já realizados por este usuário, sua descrição no perfil do *Twitter* e link para a foto do usuário, entre outros, conforme Figura 23.

Figura 23 - Dados retornados Vértices

	A	F	AC	AD	AE	AF	AG	AH
1			Other Columns					
2	Vertex	Image File	Add Your Own Columns Here	Followed	Followers	Tweets	Favorites	Time Zone UTC Offset (Seconds)
9	mimirodrigu	http://pbs.twimg.com/profile_ima		558	153	13748	1020	-25200
10	oliverssauro	http://pbs.twimg.com/profile_ima		325	287	59556	4184	-10800
11	sohluscious	http://pbs.twimg.com/profile_ima		43	1831	839	194	
12	fabyx_raque	http://pbs.twimg.com/profile_ima		351	384	13833	1165	-7200
13	anvisa_ofici	http://pbs.twimg.com/profile_ima		498	38733	4302	39	-10800
14	electrabru	http://pbs.twimg.com/profile_ima		1484	3337	90080	33	-10800
15	madalenat2	http://pbs.twimg.com/profile_ima		211	262	24366	31952	
16	pmrepreser	http://pbs.twimg.com/profile_ima		360	58	138	73	
17	whoisllara	http://pbs.twimg.com/profile_ima		111	484	9526	1982	
18	marianaafo	http://pbs.twimg.com/profile_ima		262	653	50277	16826	3600
19	vanessames	http://pbs.twimg.com/profile_ima		138	40	4785	395	-10800
20	laboreirova	http://pbs.twimg.com/profile_ima		40	22	36297	28677	
21	mudarcc	http://pbs.twimg.com/profile_ima		190	70	28341	1073	
22	olinda3	http://pbs.twimg.com/profile_ima		695	510	13315	4411	-10800
23	czarfalido	http://pbs.twimg.com/profile_ima		261	451	31736	12011	-14400
24	maryannea	http://pbs.twimg.com/profile_ima		305	115	4150	23	-10800
25	robsonhcos	http://pbs.twimg.com/profile_ima		787	225	29178	91	-10800
26	bmoreirads	http://pbs.twimg.com/profile_ima		167	65	403	348	
27	take_me_lu	http://pbs.twimg.com/profile_ima		62	359	1077	1059	
28	paquadant	http://pbs.twimg.com/profile_ima		195	237	2920	2310	-25200

Fonte: Autor (2016).

Após as coletas realizadas é possível salvar um arquivo com extensão .xlsx utilizando o Excel. O aplicativo do NodeXL permite a geração de alguns grafos com diversos formatos de visualização.

5.4.3. Netlytic

O Netlytic é uma ferramenta WEB. Para utilizar suas opções de coleta de dados é necessário preencher um breve cadastro e efetuar *login*, conforme Figura 24.

Figura 24 - Tela Inicial Netlytic

The screenshot shows the Netlytic web interface. At the top, the heading "Making sense of online conversations" is displayed. Below it, a descriptive paragraph states: "Netlytic is a cloud-based text and social networks analyzer that can automatically summarize and discover social networks from online conversations on social media sites." To the right of this text is a "Learn More..." button. The main content area is divided into two columns. The left column, titled "SIGN-IN OR REGISTER WITH OPENID ...", contains two buttons: "Sign in with Google" (with a Google+ icon) and "Sign In with a Yahoo ID". The right column, titled "EMAIL LOGIN (NON-OPENID)", contains fields for "Email:" (with the value "camila.longhi@outlook.com") and "Password:" (with masked characters "*****"). Below these fields are links for "Forgot password?" and buttons for "Log In" and "Register (Free)". At the bottom of the page, a dark footer bar contains the following text: "Copyright © 2006-2016 :: Privacy Policy :: System Status" and "Contact Developer Dr. Anatoliy Gruzd at netlytic@gmail.com".

Fonte: Netlytic (2016).

Depois de efetuar o *login* é possível criar *datasets* de algumas redes sociais, ou efetuar o estudo dos dados em conteúdos presentes em arquivos e RSS conforme Figura 25.

Figura 25 - Datasets Netlytic

The screenshot shows the Netlytic web interface. At the top is a navigation bar with links: Home, About, Resources For..., Help, My Datasets, New Dataset, My Account, and Log Out. Below the navigation bar is a green banner that reads: "Need help with data collection? Review our Youtube tutorials!". The main content area is titled "New Dataset" and features a tabbed interface with tabs for Twitter, Facebook, Instagram, YouTube, Cloud Storage, Text File, and RSS. The "Twitter" tab is selected, showing a section titled "Twitter API information and limitations". Below this, it states "Twitter account linked with Netlytic: Mila_Longhi". There are two input fields: "Dataset Name" and "Twitter Search Terms". Below these fields, a note says: "Search terms may be keywords with operators, hashtags, or @usernames. [See all options.](#)". There is a checkbox labeled "Enable data collection from this twitter search every 15 minutes for the next 1 day(s)". At the bottom of the form, it says "You are using 3/3 of your permitted datasets - [Get More](#)". A "Cancel" button is located at the bottom left of the form.

Fonte: Netlytic (2016).

Para este estudo de caso foram criados *datasets* das redes sociais *Instagram* e *Facebook*. No *dataset* do *Instagram* foram efetuadas buscas por *posts* com as *hashtags* “*#Diet*”, “*#Light*”, “*#Gluten*” e “*#Lactose*”. A busca dessas *hashtags* é realizada de forma simples, mas é necessário estar logado em uma conta do *Instagram*, conforme Figura 26.

Figura 26 - Configuração Dataset Instagram

Twitter Facebook Instagram YouTube Cloud Storage Text File RSS

Instagram API information and limitations

Instagram account linked with Netlytic: **camilalonghi**

Diet

Instagram allows searching by a single hashtag **or** a location (latitude/longitude coordinates). Recent media tagged with this term, or taken in a 5km radius of this location, will have their captions and replies collected.

Choose a search type:

☒ Hashtag
☐ Location ([See Instructions](#))

#Diet

Import 10000 records each time.

☒ Enable data collection from this Instagram search **every hour** for the next 1 day(s).

Fonte: Netlytic (2016).

Para a coleta dos dados do *Facebook*, também é necessário efetuar *login* em uma conta da rede social. Neste caso, foram coletados *posts* realizados pelas duas maiores concorrentes da empresa ALLPS, Clube do Zero e Glúten Free Box (conforme informado no levantamento de requisitos realizado junto a ALLPS). Para essa coleta, é necessário que a empresa possua uma *fanpage* no *Facebook*, conforme Figura 27.

Figura 27 - Configuração Dataset Facebook

Twitter Facebook Instagram YouTube Cloud Storage Text File RSS

Facebook API information and limitations

Gluten Free Box
glutenfreebox

Facebook import works for **public** pages. To import from groups, [follow these instructions](#). Paste the code from the Facebook URL, after the forward slash, as shown below.

☒ Enable data collection from this Facebook search **every hour** for the next 1 day(s).

Fonte: Netlytic (2016).

Posteriormente as coletas realizadas, a próxima etapa do plano de trabalho é o pré-processamento dos dados coletados para a realização da mineração e análise final.

5.5. PRÉ-PROCESSAMENTO

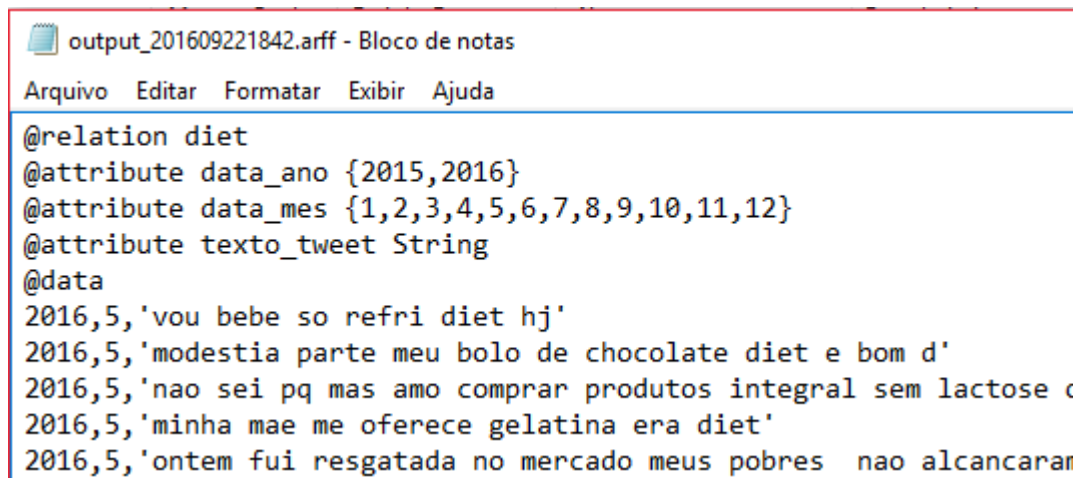
No pré-processamento dos dados coletados, foram necessárias algumas alterações e correções dos arquivos gerados pelos programas utilizados. Em alguns casos, os próprios programas utilizados para a coleta já realizam um pré-processamento dos dados, conforme explicado a seguir.

5.5.1. Arquivos gerados pela Aplicação JAVA

O programa de aplicação Java desenvolvido por Hoffman (2016), efetua automaticamente um pré-processamento dos *tweets* coletados, removendo caracteres especiais, acentuações, vírgulas, pontuações e formata todas as letras encontradas para minúsculo.

O arquivo gerado possui a extensão .ARFF, reconhecido pelo programa de análise de dados WEKA. O programa padrão gera um arquivo que possui uma formatação contendo dados como um nome da relação e atributos (ano, mês, texto), conforme Figura 28.

Figura 28 - Arquivo ARFF



Fonte: Autor (2016).

Para este trabalho, foi necessário realizar algumas alterações no arquivo para que estes fossem aceitos na aplicação dos algoritmos Apriori e K-Means no programa WEKA. Essas alterações serão explicadas a seguir.

Foi necessário transformar o arquivo ARFF em um arquivo CSV. Posteriormente essa transformação, no programa Excel, foi incluído um índice das palavras utilizadas na captura dos *tweets*, conforme Figura 29. Foram incluídas colunas conforme a palavra buscada, dividindo este trabalho em duas etapas. Essas alterações foram necessárias para preparar o arquivo .ARFF para leitura dos algoritmos no programa WEKA.

Figura 29 - Conversão CSV

The screenshot shows a Microsoft Excel spreadsheet with the following columns:

- A**: @attribute_data_ano {2015,2016}
- B**: @attribute_data_mes {1,2,3,4,5,6,7,8,9,10,11,12}
- C**: @attribute_texto_tweet String
- D**: #data

The rows contain tweets related to healthy eating, such as:

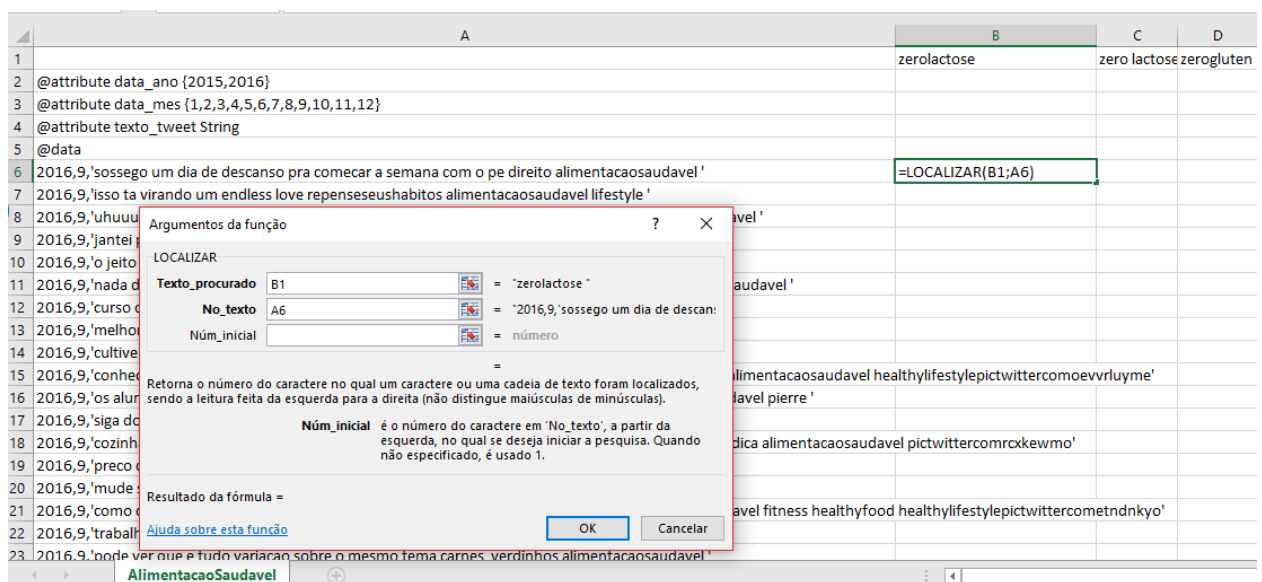
- 2016,9,'sossego um dia de descanso pra começar a semana com o pé direito alimentacaosaudavel'
- 2016,9,'isso tá virando um endless love repense seus hábitos alimentacaosaudavel lifestyle'
- 2016,9,'uhuuu já somos quase bora conhecer nossas dicas de alimentacao saudavel alimentacaosaudavel'
- 2016,9,'jantei pizza ontem almocei lasanha e vou jantar pizza otv alimentacaosaudavel'
- 2016,9,'jeito mais gostoso de curtir a primavera e no saltaoficial salatta alimentacaosaudavel'
- 2016,9,'nada de tentar reaproveitar pão mofado hein pessoal muito cuidado alimentacao alimentacaosaudavel'
- 2016,9,'curso online sobre aproveitamento integral de alimentos alimentacaosaudavel'
- 2016,9,'melhor parte do meu dia foco agora vai alimentacaosaudavel atividade fisica musculacao'
- 2016,9,'cultive uma horta e estimule seu filho a comer melhor'
- 2016,9,'conheça os benefícios da hortela toquedenutricao health saude bemestar healthyfood fitness alimentacaosaudavel healthyfestylepictwittercomoevnrylme'
- 2016,9,'os alunos do ano trabalham natureza com as frutas na nossa quarta saudavel alimentacaosaudavel pierre'
- 2016,9,'siga dolmasdecasa para receitas e dicas sobre reeducacaionalimentar e alimentacaosaudavel'
- 2016,9,'cozinhar a batata antes de descascar facilita o preparo da comida e preserva os nutrientes ficadica alimentacaosaudavel pictwittercomrcxwewm'
- 2016,9,'preço de compra do produto foi reajustado e beneficiará mil produtores do paa'
- 2016,9,'mude sua dieta em passos simples'
- 2016,9,'como desinfetar verduras e legumes toquedenutricao saude bemestar health alimentacaosaudavel fitness healthyfood healthyfestylepictwittercometndrkyo'
- 2016,9,'trabalho alimentacaosaudavel em meio a correria do trabalho e'
- 2016,9,'tudo ver por e tudo utilizar sobre o mesmo tema cores verdeilhos alimentacaosaudavel'

Fonte: Autor (2016).

Depois foi realizado o mesmo trabalho, mas utilizando como índice palavras relacionadas aos alimentos mais procurados por pessoas que possuem algum tipo de restrição alimentar, conforme Figura 29.

A conversão do arquivo para .csv e a inclusão dos índices, foi utilizada a fórmula do Excel “Localizar” em cada cédula entre a coluna com a restrição alimentação e o tweet capturado. Esta fórmula localiza a expressão ou texto procurado no texto selecionado. Utilizou-se esta fórmula para verificar quais outras palavras-chave existiam no tweet coletado especificamente para uma palavra-chave, conforme Figura 30.

Figura 30 - Aplicação fórmula Excel



Fonte: Autor (2016).

Após a aplicação da fórmula, todas as cédulas que obtiverem um valor > (maior) que “0”, encontraram a palavra-chave procurada no texto do tweet, e essas foram convertidas para o número “Sim” no caso de aplicação do algoritmo K-Means ou “y” na aplicação do algoritmo Apriori, para que estas sejam reconhecidas no programa WEKA. Caso a fórmula não encontre a palavra-chave, e a cédula obtiver o valor de #VALOR, as mesmas foram convertidas em “Não” para o algoritmo Apriori e “?” para o algoritmo K-Means, conforme Figuras 31 e 32.

Figura 31 – Preparação dos dados para o Apriori

I16				
	A	B	E	F
1		zerolac	semovos	alimenta
2	@attribute data_ano {2015,2016}			
3	@attribute data_mes {1,2,3,4,5,6,7,8,9,10,11,12}			
4	@attribute texto_tweet String			
5	@data			
6	2016,9,'sossego um dia de descanso pra começar a semana com o pe direito e	Não	Não	Sim
7	2016,9,'isso ta virando um endless love repense seus hábitos alimentares saud	Não	Não	Sim
8	2016,9,'uhuuu já somos quase bora conhecer nossas dicas de alimentação s	Não	Não	Sim
9	2016,9,'jantei pizza ontem almocei lasanha e vou jantar pizza otv alimentac	Não	Não	Sim
10	2016,9,'o jeito mais gostoso de curtir a primavera e no salatta oficial salatta ali	Não	Não	Sim
11	2016,9,'nada de tentar reaproveitar pão mofado hein pessoal muito cuidado	Não	Não	Sim
12	2016,9,'curso online sobre aproveitamento integral de alimentos alimentac	Não	Não	Sim
13	2016,9,'melhor parte do meu dia foco agora vai alimentação saudável atividade	Não	Não	Sim
14	2016,9,'cultive uma horta e estimule seu filho a comer melhor '	Não	Não	Sim
15	2016,9,'conheça os benefícios da horta toques de nutrição saúde bem-est	Não	Não	Sim
16	2016,9,'os alunos do ano trabalham natureza com as frutas na nossa quarta	Não	Não	Sim
17	2016,9,'siga dicas de casa para receitas fit e dicas sobre reeducação alimentar	Não	Não	Sim
18	2016,9,'cozinhar a batata antes de descascar facilita o preparo da comida e pr	Não	Não	Sim
19	2016,9,'preço de compra do produto foi reajustado e beneficiará mil produtos	Não	Não	Sim
20	2016,9,'como desinfetar verduras e legumes toques de nutrição saúde bem-est	Não	Não	Sim
21	2016,9,'trabalho alimentar saudável em meio à correria do trabalho e '	Não	Não	Sim
22	2016,9,'pode ver que é tudo variação sobre o mesmo tema carnes verdinhos	Não	Não	Sim
23	2016,9,'pra começar o prato de hoje primeiro dia sem arroz dietado hoje aliment	Não	Não	Sim

Fonte: Autor (2016).

Figura 32 - Preparação de Dados para o K-Means

	A	B	E	F
1		zerolac	semovos	alimenta
2	@attribute data_ano {2015,2016}			
3	@attribute data_mes {1,2,3,4,5,6,7,8,9,10,11,12}			
4	@attribute texto_tweet String			
5	@data			
6	2016,9,'sossego um dia de descanso pra comecar a semana com o pe direito a y		?	y
7	2016,9,'isso ta virando um endless love repenseseushabitos alimentacaosaud y		y	y
8	2016,9,'uhuuu ja somos quase bora conhecer nossas dicas de alimentacao sa y		y	y
9	2016,9,'jantei pizza ontem almocei lasanha e vou jantar pizza otv alimentacao y		?	y
10	2016,9,'o jeito mais gostoso de curtir a primavera e no salattaoficial salatta ali y		?	y
11	2016,9,'nada de tentar reaproveitar pao mofado hein pessoal muito cuidado y		?	?
12	2016,9,'curso online sobre aproveitamento integral de alimentos alimentacao y		?	?
13	2016,9,'melhor parte do meu dia foco agora vai alimentacaosaudavel atividade y		?	?
14	2016,9,'cultive uma hortinha e estimule seu filho a comer melhor ' y		?	?
15	2016,9,'conheca os beneficios da hortela toquedenutricao health saude beme y		?	?
16	2016,9,'os alunos do ano trabalham natureza com as frutas na nossa quarta : y		?	?
17	2016,9,'siga dolmasdecasa para receitasfit e dicas sobre reeducacaoalimentar y		y	?
18	2016,9,'cozinhar a batata antes de descascar facilita o preparo da comida e pr y		y	?
19	2016,9,'preco de compra do produto foi reajustado e beneficiara mil produtcy		y	?
20	2016,9,'como desinfetar verduras e legumes toquedenutricao saude bemestay		y	?
21	2016,9,'trabalho alimentacaosaudavel em meio a correria do trabalho e ' y		y	?
22	2016,9,'pode ver que e tudo variacao sobre o mesmo tema carnes verdinhos y		y	?

Fonte: Autor (2016).

Os campos Ano e Mês deste arquivo não foram uteis para as análises que foram utilizadas.

Depois das alterações realizadas nos arquivos .csv, os mesmos foram convertidos novamente em .ARRF e os atributos apresentados nas Figuras 33 e 34 foram incluídos, para que o arquivo pudesse ser analisado pelo programa WEKA. Na Figura 33 o arquivo está formatado para aplicação do algoritmo Apriori. Na Figura 34, o arquivo está formatado para aplicação do algoritmo K-Means.

Figura 33 - Arquivo ARFF - Apriori

```

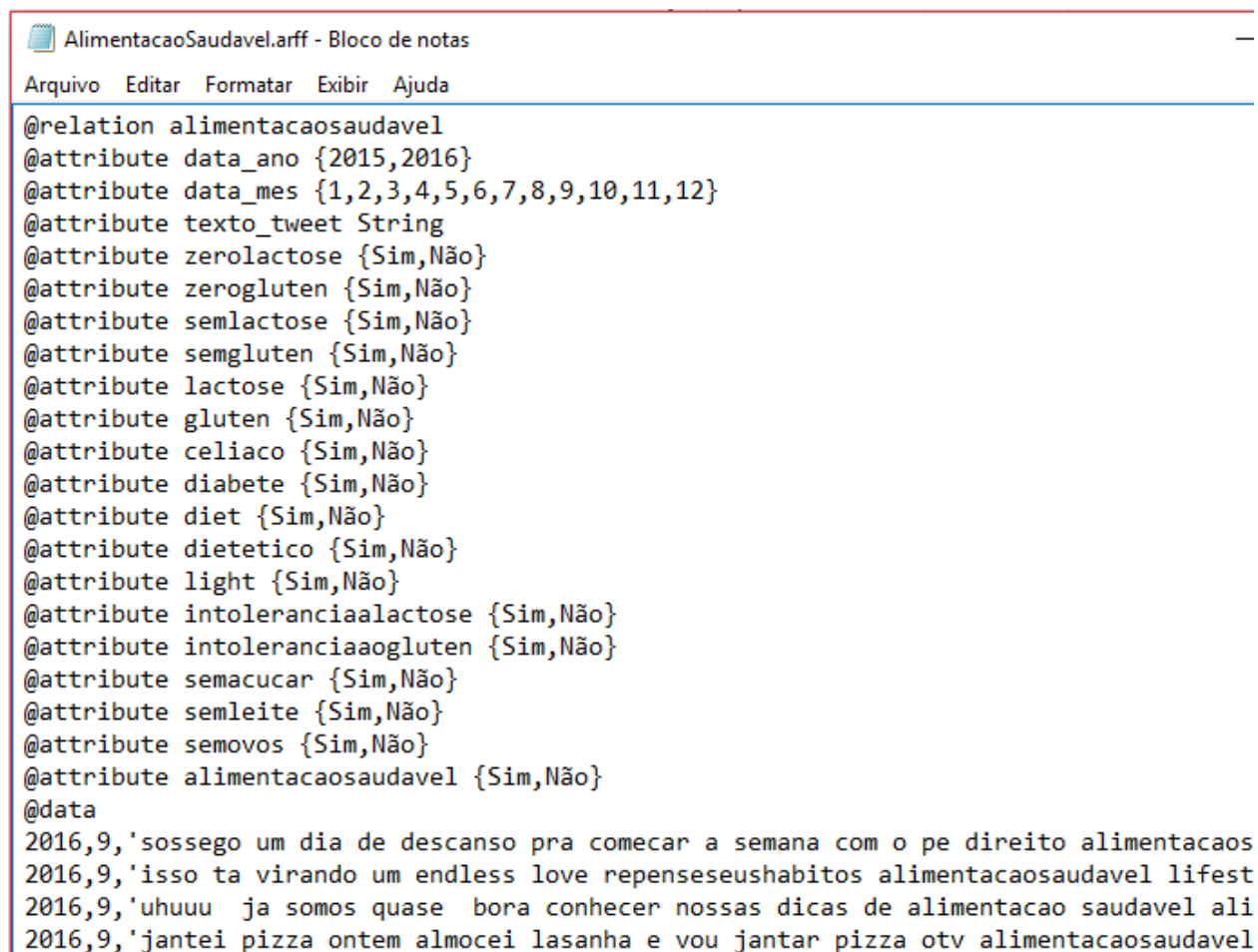
ZeroLactose.arff - Bloco de notas
Arquivo  Editar  Formatar  Exibir  Ajuda

@relation zerolactose
@attribute data_ano {2015,2016}
@attribute data_mes {1,2,3,4,5,6,7,8,9,10,11,12}
@attribute texto_tweet String
@attribute zerolactose {y,?}
@attribute zerogluten {y,?}
@attribute semlactose {y,?}
@attribute semgluten {y,?}
@attribute lactose {y,?}
@attribute gluten {y,?}
@attribute celiaco {y,?}
@attribute diabete {y,?}
@attribute diet {y,?}
@attribute dietetico {y,?}
@attribute light {y,?}
@attribute intoleranciaalactose {y,?}
@attribute intoleranciaaogluten {y,?}
@attribute semacucar {y,?}
@attribute semleite {y,?}
@attribute semovos {y,?}
@attribute alimentacaosaudavel {y,?}
@data
2016,9,'voce merece uma alimentacao saudavel venha ser famillypan alimentacao saudave
2016,9,'vai ter pao semgluten e pudim zerolactose pq sim ',y,?,?,y,y,y,?,?,?,?,,?,,?
2016,9,'bolo laranja zerogluten zerolactose massapura itatiba panificadora massa pur
2016,9,'bolo de cenoura zerogluten zerolactose ',y,y,?,?,y,y,?,?,?,?,,?,,?,,?,,?

```

Fonte: Autor (2016).

Figura 34 - Arquivo ARFF - K-Means



```

AlimentacaoSaudavel.arff - Bloco de notas
Arquivo  Editar  Formatar  Exibir  Ajuda

@relation alimentacaosaudavel
@attribute data_ano {2015,2016}
@attribute data_mes {1,2,3,4,5,6,7,8,9,10,11,12}
@attribute texto_tweet String
@attribute zerolactose {Sim,Não}
@attribute zerogluten {Sim,Não}
@attribute semlactose {Sim,Não}
@attribute semgluten {Sim,Não}
@attribute lactose {Sim,Não}
@attribute gluten {Sim,Não}
@attribute celiaaco {Sim,Não}
@attribute diabete {Sim,Não}
@attribute diet {Sim,Não}
@attribute dietetico {Sim,Não}
@attribute light {Sim,Não}
@attribute intoleranciaalactose {Sim,Não}
@attribute intoleranciaaogluten {Sim,Não}
@attribute semacucar {Sim,Não}
@attribute semleite {Sim,Não}
@attribute semovos {Sim,Não}
@attribute alimentacaosaudavel {Sim,Não}
@data
2016,9,'sossego um dia de descanso pra começar a semana com o pe direito alimentacaos
2016,9,'isso ta virando um endless love repenseseseushabitos alimentacaosaudavel lifest
2016,9,'uhuuu ja somos quase bora conhecer nossas dicas de alimentacao saudavel ali
2016,9,'jantei pizza ontem almocei lasanha e vou jantar pizza otv alimentacaosaudavel
  
```

Fonte: Autor (2016).

5.5.2. Arquivos gerados pelo NodeXL

Foram necessárias diversas coletas, em diferentes dias para obter um número de significativo de dados para análise no NodeXL, devido a limitação de 2.000 *tweets* por dia pela versão gratuita do programa.

O único pré-processamento gerado para estes arquivos foi a união dos dados gerados em dias diferentes para que os mesmos pudessem ser analisados.

Os arquivos foram gerados e analisados de forma separada, por palavra-chave coletada: Glúten, Lactose, *Diet* e *Light* e salvos em extensão .XLSX.

5.5.3. Arquivos gerados pelo Netlytic

As coletas realizadas por meio da ferramenta Netlytic não obtiveram geração de arquivos. O Netlytic se trata de uma ferramenta *WEB*, em que os dados coletados são analisados diretamente na ferramenta, utilizando as opções liberadas para os usuários trabalharem na versão *free* do programa.

5.6. APLICAÇÃO DAS TÉCNICAS DE MINERAÇÃO DE DADOS

Nos dados coletados no *Twitter*, por meio do aplicativo Java, foi aplicado o algoritmo de associação Apriori e o algoritmo de agrupamento K-Means, na ferramenta WEKA. Está descrito a seguir, como foram aplicados estes algoritmos para que estes gerem dados que possam ser analisados conforme a necessidade deste estudo de caso.

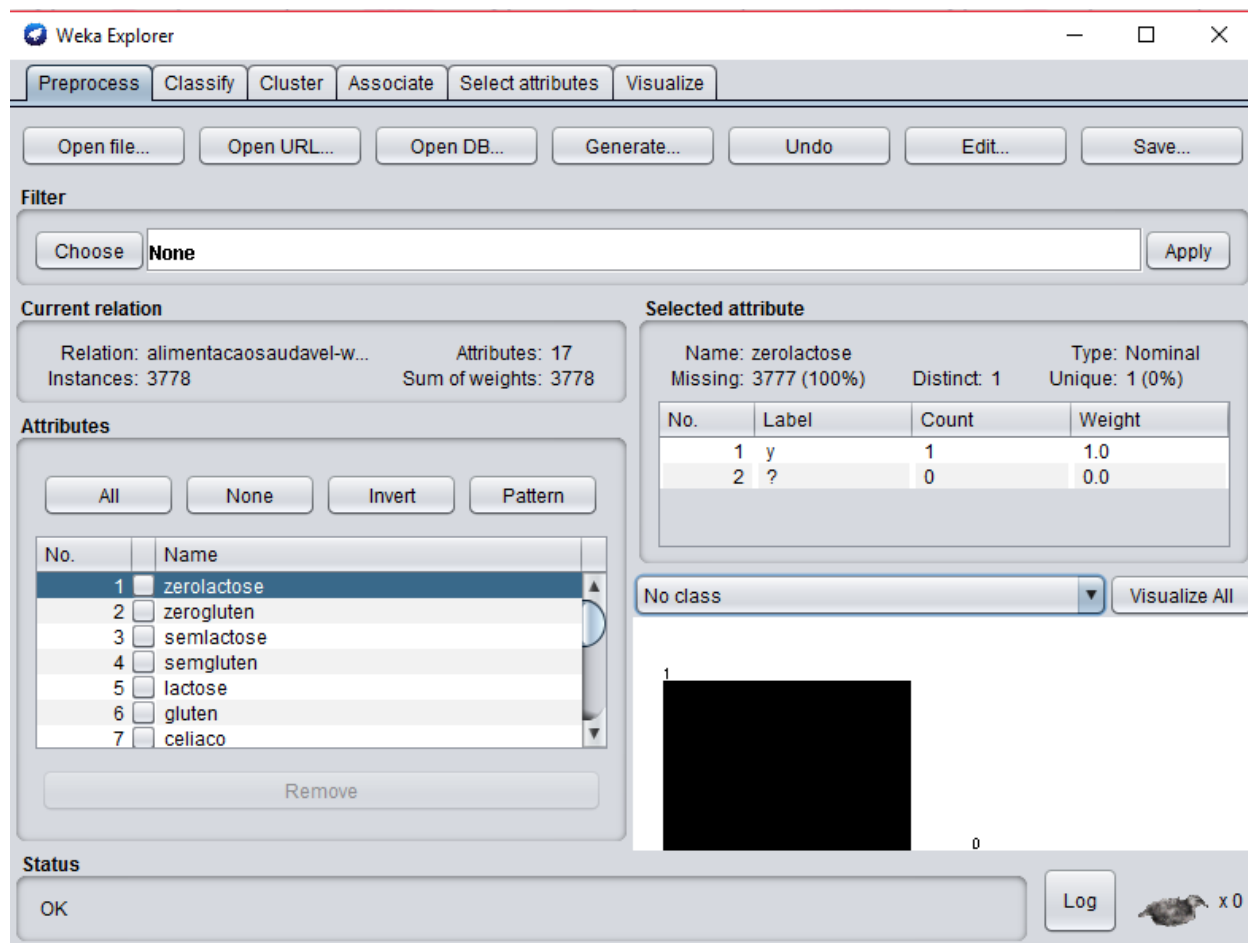
5.6.1. Algoritmo de Associação - Apriori

Na aplicação do algoritmo de associação Apriori nos dados coletados utilizando a aplicação Java, do período de 01/01/2015 até 30/09/2016, foi importado o arquivo gerado .ARFF para a ferramenta WEKA.

Neste estudo de caso, foi criado um arquivo para cada palavra-chave utilizada na etapa de coleta de dados. Estes arquivos foram gerados de forma separada com o intuito de poder realizar-se análises de cada palavra-chave de forma única, e depois, foi gerado um arquivo obtendo todos os *tweets* coletados, para que se pudesse analisar todas as palavras-chave de forma conjunta.

Na Figura 35 apresenta-se a ferramenta WEKA com o arquivo já carregado. O arquivo carregado representa a palavra-chave “Alimentacaosaudavel”. O mesmo processo foi repetido para todas as palavras-chave.

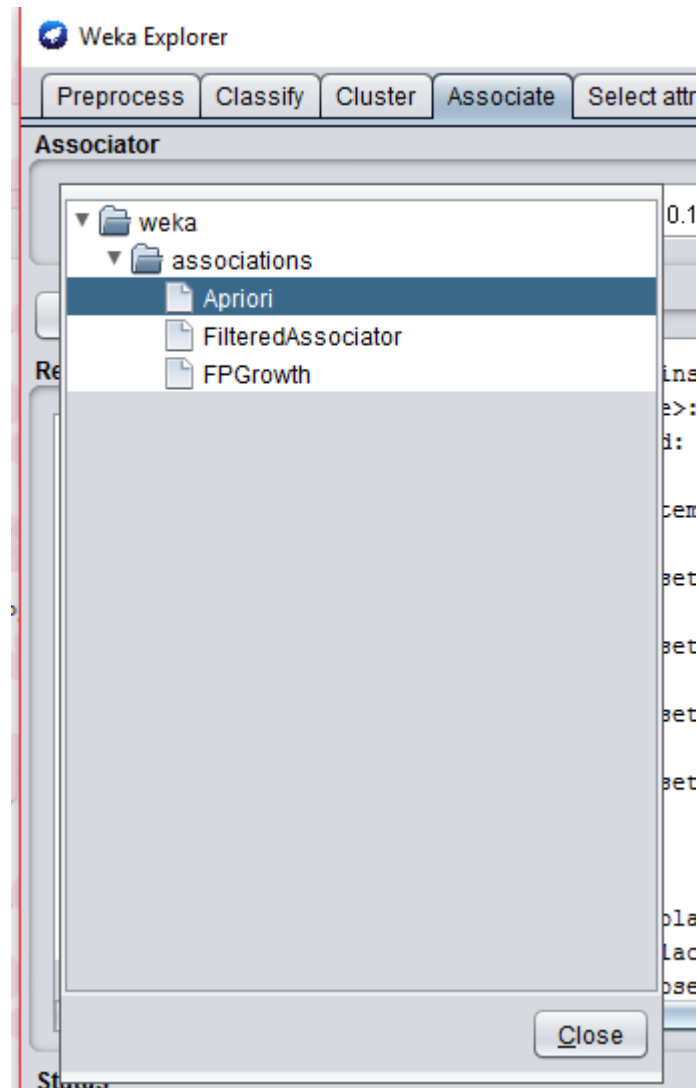
Figura 35 - Tela Inicial WEKA



Fonte: Weka (2016).

Após o carregamento do arquivo, conforme Figura 36, na aba “Associate”, foi selecionado o algoritmo utilizado para essa análise, Apriori.

Figura 36 - Seleção algoritmo Apriori



Fonte: Weka (2016).

Uma análise associativa é útil para se descobrir relacionamentos que se mostrem interessantes em um grande conjunto de dados. Às vezes, essas associações podem ser fracas, apresentando um número insignificante de algum relacionamento específico, como por exemplo: entre 100 (cem) clientes, apenas 1 (um) comprou suco e leite ao mesmo tempo.

Com base em trabalhos relacionados ao WEKA e ao algoritmo Apriori, foi alterado o parâmetro “*metricType*” para a medida de confiança. Neste estudo de caso escolheu-se medir a frequência em que determinado item X aparece, quando existe na transação o item Y.

Para o suporte, devido ao número não tão expressivo de dados, foram alterados os parâmetros “lowerboundMinSupport” para 0.1. Se este campo for preenchido com um número muito alto, itemsets envolvendo itens raros seriam perdidos. Se preenchido com um número muito baixo, as gerações de regras podem se tornar extremamente demorado, caso o conjunto de dados analisado seja de tamanho expressivo. Neste estudo de caso, a seleção de um número baixo não ocasionou demora na geração dos resultados, após aplicar o algoritmo Apriori, devido ao conjunto de dados não ser tão expressivo.

No parâmetro “*numRules*”, que define o número de regras a serem apresentadas em tela, o mesmo foi definido para apresentar 10 regras, para que sejam apresentadas apenas as regras mais significativas, conforme Figura 37.

Figura 37 - Parâmetros Apriori

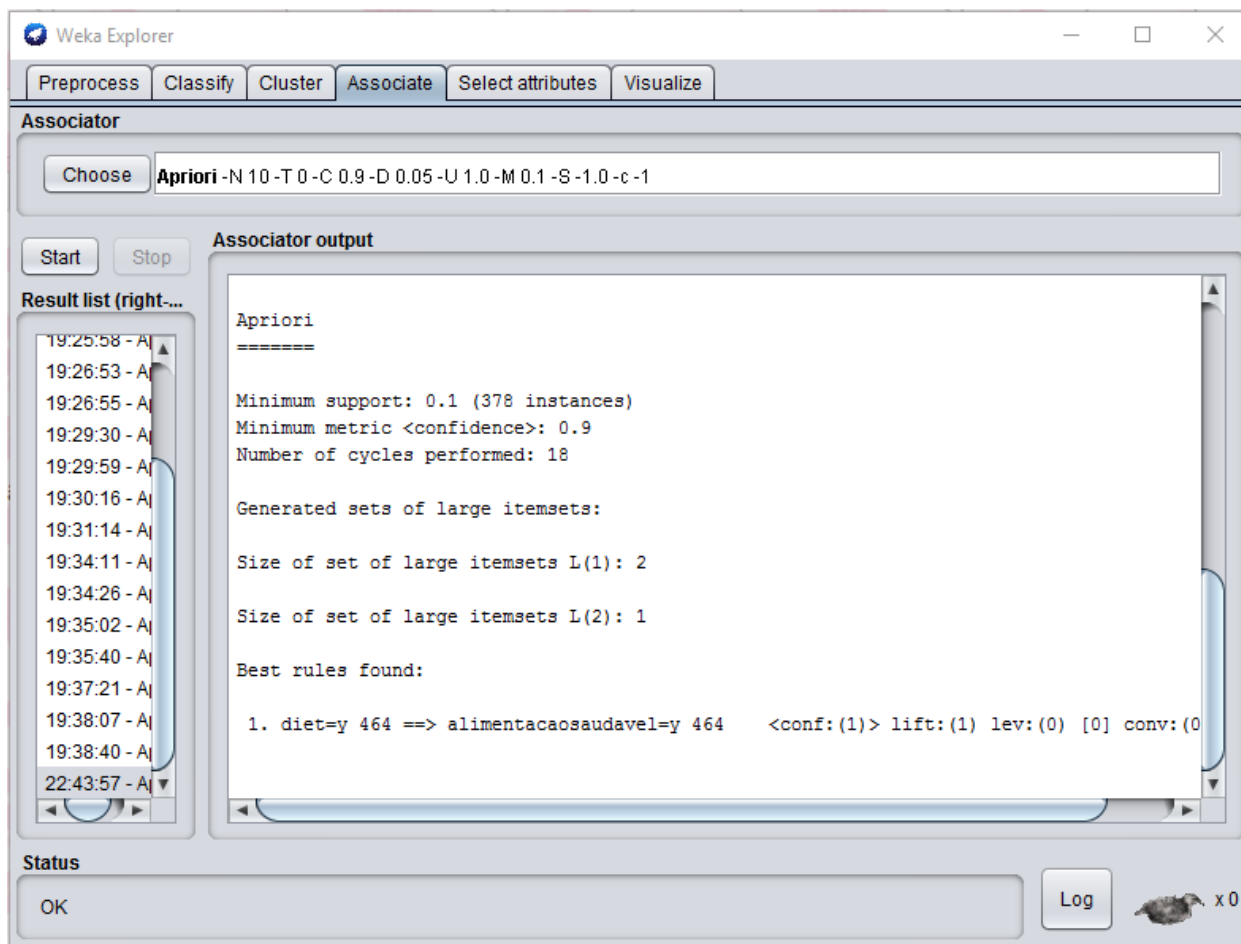
The image shows a screenshot of the Weka GUI's 'GenericObjectEditor' window, specifically for the 'weka.associations.Apriori' class. The window has a title bar with the Weka logo and the text 'weka.gui.GenericObjectEditor'. Below the title bar, the class name 'weka.associations.Apriori' is displayed. The main area is titled 'About' and contains a text box with the description 'Class implementing an Apriori-type algorithm.' and two buttons: 'More' and 'Capabilities'. Below this, there is a list of parameters for the Apriori algorithm, each with a label and a corresponding input field or dropdown menu. The parameters are: 'car' (False), 'classIndex' (-1), 'delta' (0.05), 'doNotCheckCapabilities' (False), 'lowerBoundMinSupport' (0.1), 'metricType' (Confidence), 'minMetric' (0.9), 'numRules' (10), 'outputItemSets' (False), 'removeAllMissingCols' (False), 'significanceLevel' (-1.0), 'treatZeroAsMissing' (False), 'upperBoundMinSupport' (1.0), and 'verbose' (False). At the bottom of the window, there are four buttons: 'Open...', 'Save...', 'OK', and 'Cancel'.

Parameter	Value
car	False
classIndex	-1
delta	0.05
doNotCheckCapabilities	False
lowerBoundMinSupport	0.1
metricType	Confidence
minMetric	0.9
numRules	10
outputItemSets	False
removeAllMissingCols	False
significanceLevel	-1.0
treatZeroAsMissing	False
upperBoundMinSupport	1.0
verbose	False

Fonte: Weka (2016).

Posteriormente a alteração dos algoritmos, clica-se em “Start” para que o WEKA aplique as regras do algoritmo no arquivo selecionado, conforme Figura 38.

Figura 38 - Resultados Apriori



Fonte: Weka (2016).

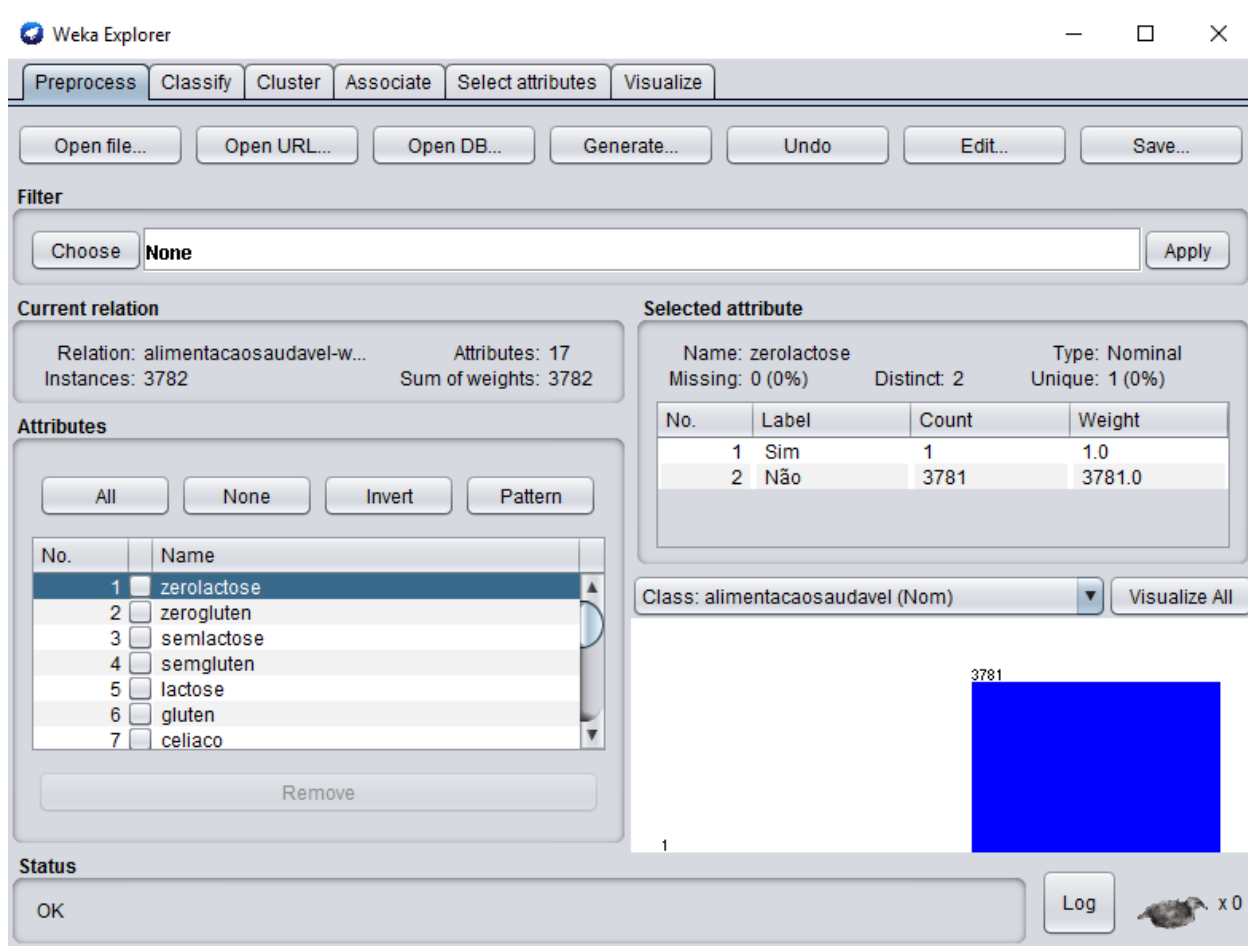
Os resultados gerados após a aplicação do algoritmo Apriori, estão apresentados no Anexo B deste estudo de caso. A análise destes dados e avaliação final dos mesmos estão presentes no capítulo 5.7.

5.6.2. Algoritmo de Agrupamento - K-Means

Para a aplicação do algoritmo de agrupamento K-Means nos dados coletados utilizando a aplicação Java, do período de 01/01/2015 até 30/09/2016, foi importado o arquivo gerado .ARFF para a ferramenta WEKA.

Na Figura 39 apresenta-se a ferramenta WEKA com o arquivo já carregado. Para essa análise, foi criado um arquivo para cada palavra-chave procurada e um arquivo obtendo todos os *tweets* coletados em toda a etapa de coleta de dados.

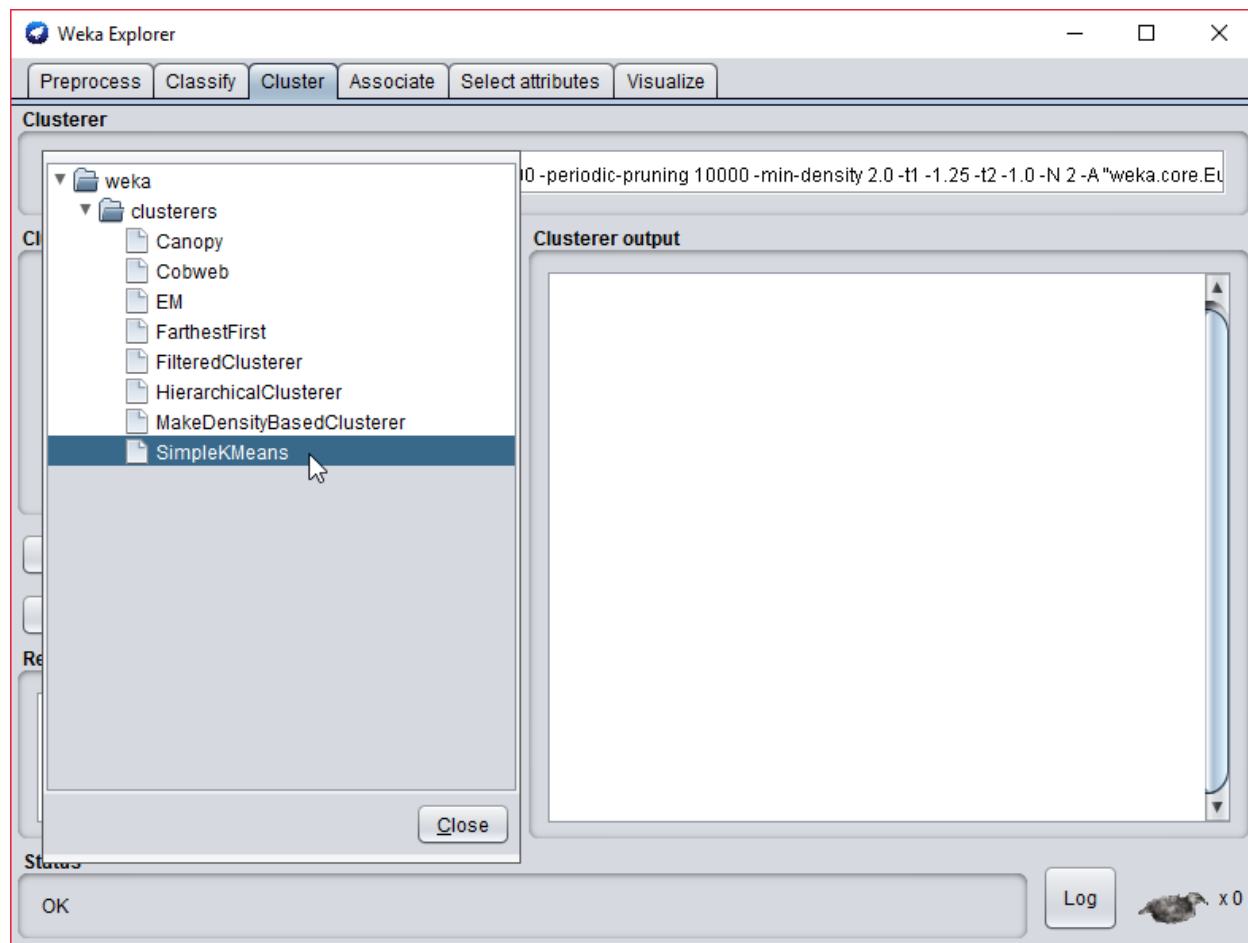
Figura 39 - Tela Inicial WEKA



Fonte: Weka (2016).

Depois do carregamento do arquivo, conforme Figura 40, na aba “Cluster”, foi selecionado o algoritmo utilizado para essa análise, *SimpleKMeans*.

Figura 40 - Seleção algoritmo K-Means

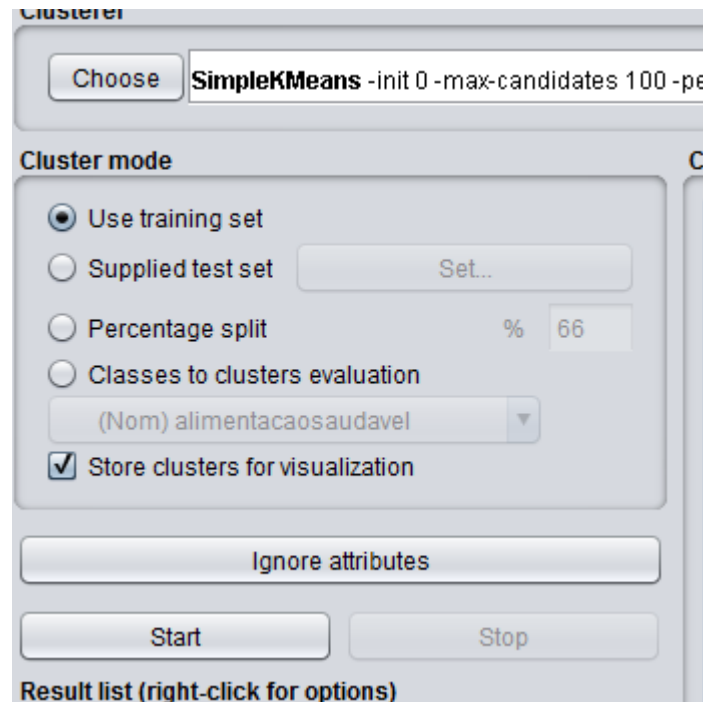


Fonte: Weka (2016).

O algoritmo de agrupamento tem como objetivo a identificação das semelhanças entre dados e dispor estes em grupos de acordo com o seu grau de similaridade.

Com base em trabalhos relacionados ao WEKA e ao algoritmo K-Means, foi selecionada a opção “Use training set”, conforme Figura 41. Optando por essa opção, o modelo gerado pelo algoritmo será avaliado sobre o mesmo conjunto de dados em que foi treinado. Assim, todos os registros são utilizados para construção do resultado.

Figura 41 - Parâmetros iniciais K-Means



Fonte: Weka (2016).

Dos parâmetros com permissão de alteração para este algoritmo, foram deixados alguns dados padrões e os parâmetros *“initializationMethod”*, *“numClusters”*, que foram alterados conforme Figura 42.

O armazenamento em cluster tem suas vantagens quando o conjunto de dados for definido e um padrão geral necessita ser determinado a partir destes dados. É possível criar um número específico de grupos a serem gerados, dependendo das necessidades. O número de clusters foi definido em 10 (dez) para apresentar apenas os grupos de mais relevância, devido volume de conjunto de dados deste estudo de caso não ser volumoso.

Figura 42 - Parâmetros K-Means

weka.gui.GenericObjectEditor

weka.clusterers.SimpleKMeans

About

Cluster data using the k means algorithm. [More](#) [Capabilities](#)

canopyMaxNumCanopiesToHoldInMemory 100

canopyMinimumCanopyDensity 2.0

canopyPeriodicPruningRate 10000

canopyT1 -1.25

canopyT2 -1.0

debug False

displayStdDevs False

distanceFunction [Choose](#) **EuclideanDistance** -R first-l

doNotCheckCapabilities False

dontReplaceMissingValues False

fastDistanceCalc False

initializationMethod **k-means++**

maxIterations 500

numClusters 10

numExecutionSlots 1

preserveInstancesOrder False

reduceNumberOfDistanceCalcsViaCanopies False

Fonte: Weka (2016).

Os resultados gerados após a aplicação do algoritmo K-Means estão apresentados no Anexo A deste estudo de caso. A análise destes dados e avaliação final dos mesmos estão presentes no capítulo 5.7.

5.7. ANÁLISE DOS DADOS

Posteriormente ao pré-processamento dos dados coletados e aplicação dos algoritmos de mineração de dados, os resultados finais foram analisados e apresentados para a empresa ALLPS. A apresentação dos resultados foi realizada por meio de resultados textuais, grafos, gráficos e um questionário foi encaminhado a empresa para obter um retorno da análise realizada para consideração final deste estudo de caso.

O questionário encaminhado a empresa e seu retorno estão presentes no Anexo D e os dados mais significativos desta análise estão apresentados abaixo.

5.7.1. Resultado Algoritmos de Mineração

Os resultados gerados por meio da mineração de dados utilizando os algoritmos K-Means e Apriori foram apresentados a empresa e analisados. Abaixo apresenta-se os resultados analisados com cada técnica de mineração.

5.7.1.1 Algoritmo de Agrupamento K-Means

Por meio da técnica de mineração de dados, com o algoritmo de agrupamento K-Means foi possível identificar algumas combinações de restrições alimentares e alimentos que ocorreram com mais frequência nas coletas geradas.

Posteriormente algumas análises, foi possível identificar também que o algoritmo de agrupamento, para este estudo de caso, deveria ter sido tratado de forma diferente, pois seus resultados foram muito semelhantes aos do algoritmo Apriori. Para um

algoritmo de agrupamento, os dados a serem analisados devem possuir mais características, como por exemplo, gênero, idade, quantidade, para que os agrupamentos formem mais grupos de dados, montando diversos perfis de interesse, o que não foi o caso deste trabalho.

Contudo, após as análises, os resultados foram apresentados a empresa da seguinte forma, apresentado apenas os resultados mais significantes, outros resultados estão apresentados no Anexo A deste estudo de caso.

Das restrições alimentares, quando analisadas todas as restrições em conjunto, foi possível analisar que a Lactose e/ou Intolerância a Lactose foi a restrição mais presente nos *tweets* coletados. Para surpresa da empresa ALLPS Alimentos Saudáveis, a Diabete este mais presente que o Glúten, até então o foco principal dos produtos da empresa. Além destes, o Light e o Diet também apareceram de forma significativa.

Quando analisado apenas os *tweets* que continha a expressão “Alimentação Saudável”, identificou-se que os usuários comentavam mais sobre as restrições Diet e Light.

Foi identificado também que o glúten e a lactose estão interligados com alimentos sem açúcar e sem leite, o que mostra a empresa que um marketing digital direcionado com estes tipos de alimentos, (sem leite e sem açúcar) pode interessar também seus clientes que procuram produtos sem glúten e sem lactose.

Dentro das análises geradas com o algoritmo K-Means foi realizado um estudo para identificar quais alimentos eram mais comentados juntamente com as restrições mais frequentes, lactose, glúten, *diet* e *light*.

Apresentado estes dados a empresa, é possível auxiliá-los a conhecer melhor os produtos que os usuários estão buscando, conforme Tabela 11 e que podem chamar mais a atenção em uma propaganda e auxiliá-los a incluir em seus produtos alimentos que realmente atendam às necessidades do seu público alvo.

Tabela 11 - Resultado Alimentos – K-Means

Sem Lactose	Sem Glúten	<i>Diet</i>	<i>Light</i>
Pão, Leite Condensado, Bolos, Doce de Leite, Biscoitos e Creme de Leite.	Bolos, Pão e Biscoitos	Bolo, Pão, Biscoito e Achocolatado	Pão, Bolo, Biscoito, Iogurte, Creme de Leite, Biscoitos e Doce de Leite.

Fonte: Autor (2016).

Com destes dados, é possível que a empresa realize um Marketing Digital *One-to-One*, pois conhece e atende de forma mais direta as necessidades do seu público alvo.

5.7.1.2 Algoritmo de Agrupamento Apriori

Nos resultados gerados pelo algoritmo Apriori, foi possível analisar e apresentar a empresa ALLPS as associações mais significativas.

Quando analisados todos os *tweets* gerados, as associações entre Lactose e Glúten, ou seja, sem lactose, sem glúten, zero lactose, zero glúten foram os mais significativos, apresentando assim que estas restrições são as mais comentadas no *Twitter* e geralmente de forma conjunta. Outros agrupamentos se tornaram significativos, quando comparados a análise SWOT da empresa. Estes foram apresentados conforme a Tabela 12.

Tabela 12 - Resultados Apriori

Alimentação Saudável -> Diet
Diabete -> Sem Açúcar, Lactose, Light, Diet
Diet -> Light
Diabético -> Light
Light -> Diet
Glúten -> Lactose
Sem Açúcar -> Glúten
Sem Leite -> Glúten, Lactose
Sem Ovos -> Glúten

Fonte: Autor (2016).

Após apresentação destes dados a ALLPS, a mesma retornou, mediante a respostas do questionário presente no anexo D deste estudo caso, a conclusão de que a inclusão de produtos Diet, Light poderão chamar a atenção de um público maior, principalmente devido ao resultado da busca pela expressão Alimentação Saudável, que retornou um agrupamento com produtos Diet.

5.7.2. Resultado NodeXL

A coleta de dados geradas pelo programa NodeXL permitiu a criação de grafos no qual é possível visualizar a quantidade de postagens e iterações que ocorrem com *tweets* que possuem as palavras “Glúten”, “Lactose”, “*Diet*” e “*Light*”.

Essas palavras foram escolhidas pela empresa ALLPS, pois o foco atual da empresa é com produtos sem glúten e sem lactose. A empresa necessitava entender se seu foco estava nos produtos corretos e se precisariam ou não mudar seus produtos oferecendo também opções “*Light*” e “*Diet*”.

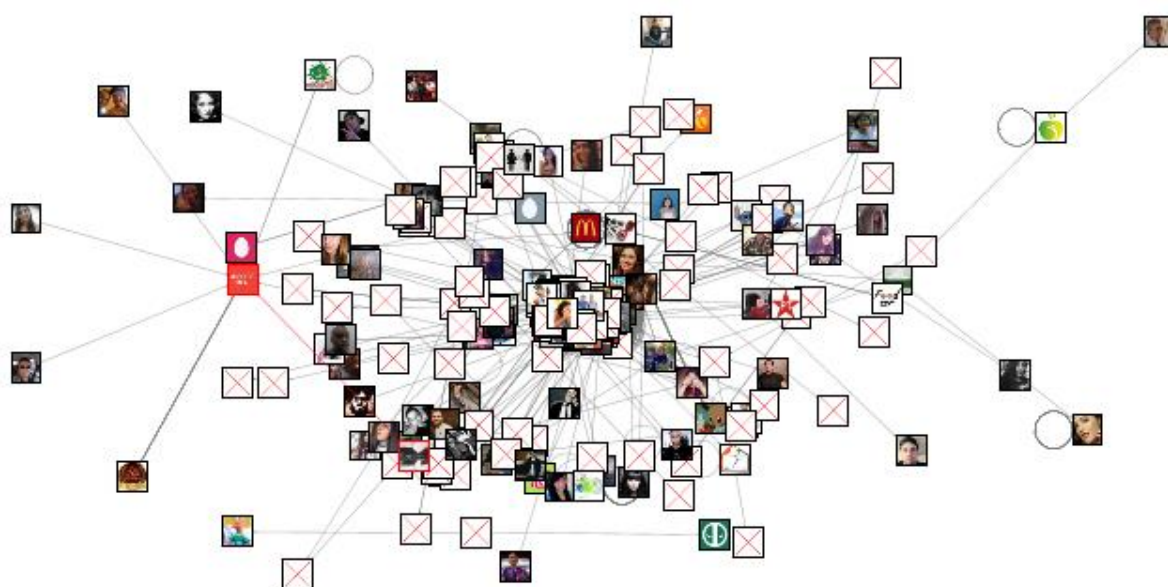
Nos grafos gerados, as figuras representam os usuários que realizaram o *tweet* na rede social. Os círculos representam a resposta de um usuário ao *tweet* publicado e as linhas representam as ligações entre os *tweets* e suas ligações como um *retweet* ou quando um usuário é mencionado a mensagem de outros usuários.

Por meio destes grafos, foi possível analisar que o foco da empresa em produtos sem glúten e sem lactose está correto, pois há um número expressivamente maior de pessoas com interesse nesses tipos de produtos conforme Figuras 45 e 46. Mas os grafos gerados por meio dos *tweets* coletados com ênfase em “*Diet*” e “*Light*”, por se tratarem de produtos rigorosamente utilizados em diversos tipos de dietas, é possível visualizar que existe uma quantidade relativamente significativa de demanda de público com essa necessidade, conforme Figuras 43 e 44.

Depois da apresentação destes grafos e do questionário presente no ANEXO D deste estudo de caso, obteve-se o retorno da empresa de que, após análise destes dados, irá solicitar aos seus fornecedores a apresentação de produtos *Diet* e *Light*, com ou sem lactose/glúten, para inclusão destes em seu produto, as caixas. Além disso, após a escolha destes produtos, a empresa irá investir em posts patrocinados no *Facebook* para a divulgação destes novos produtos em sua página, realizando um marketing digital com foco no público que possui a necessidade de produtos *Diet* e *Light*.

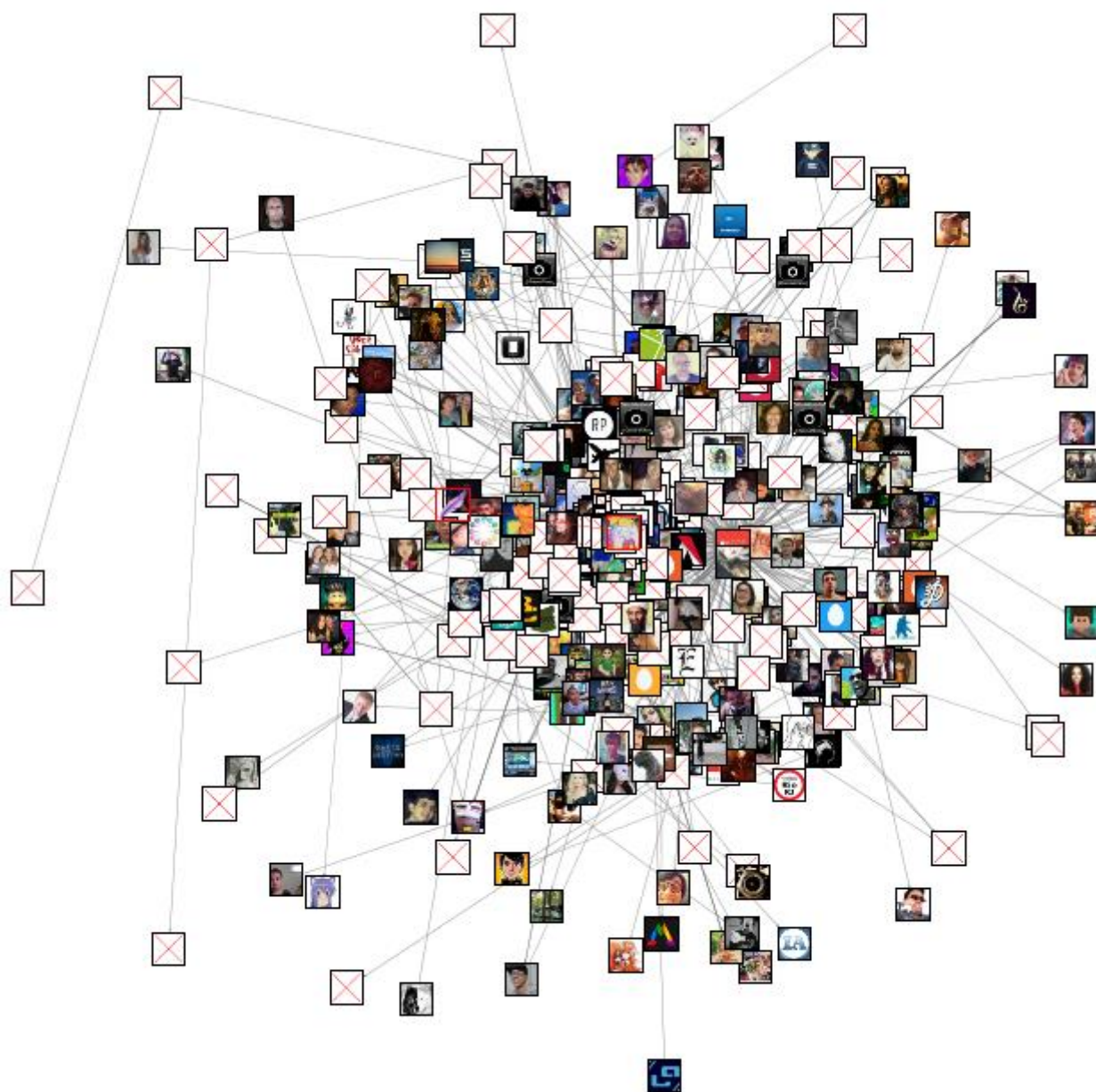
Figura 43 - Resultados NodeXL – Diet

Social media network connections

Created with NodeXL Basic (<http://nodexl.codeplex.com>) from the Social Media Research Foundation (<http://www.smrfoundation.org>)

Fonte: Autor (2016).

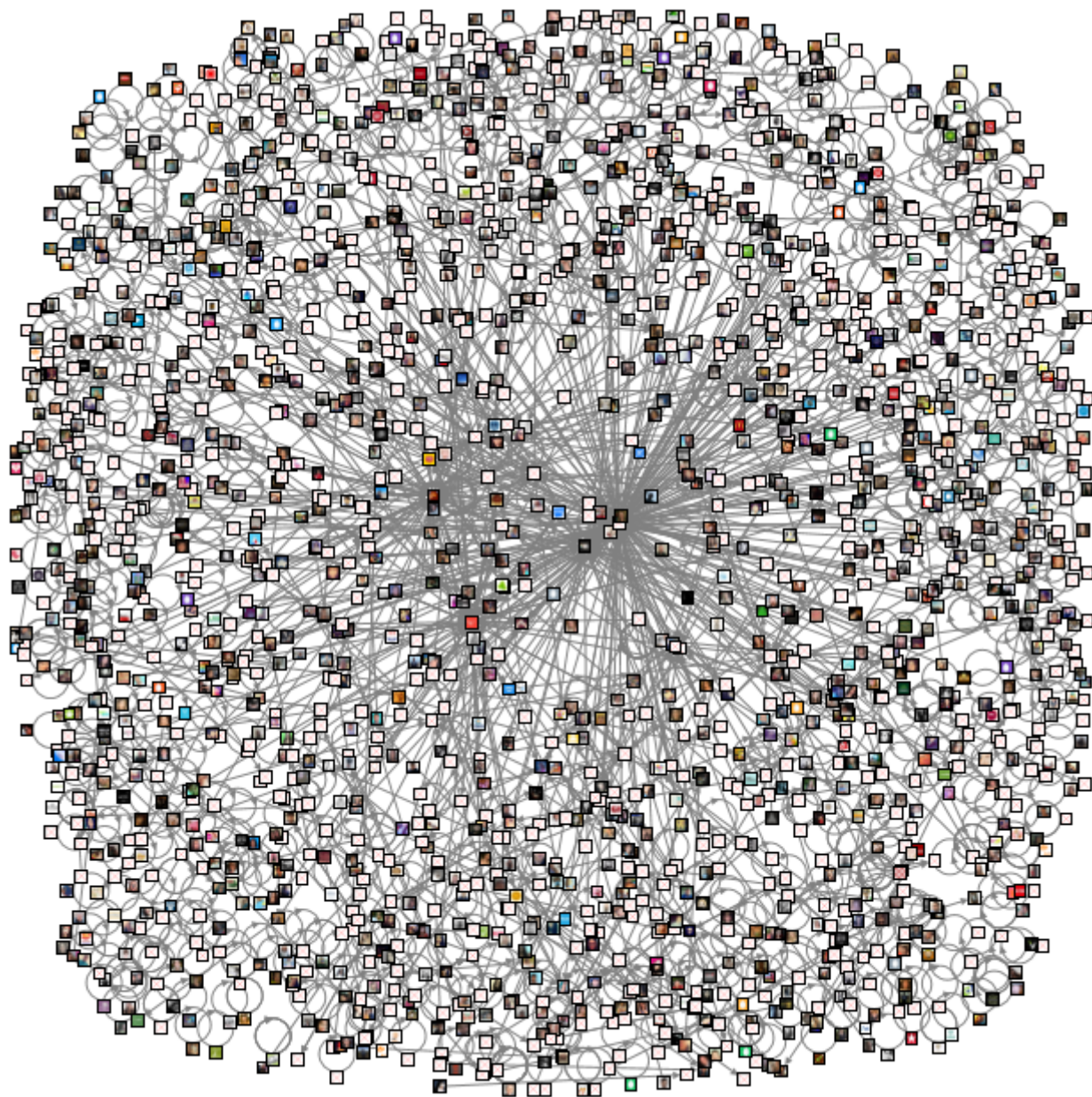
Figura 44 - Resultados NodeXL - Light



Created with NodeXL (<http://nodexl.codeplex.com>)

Fonte: Autor (2016).

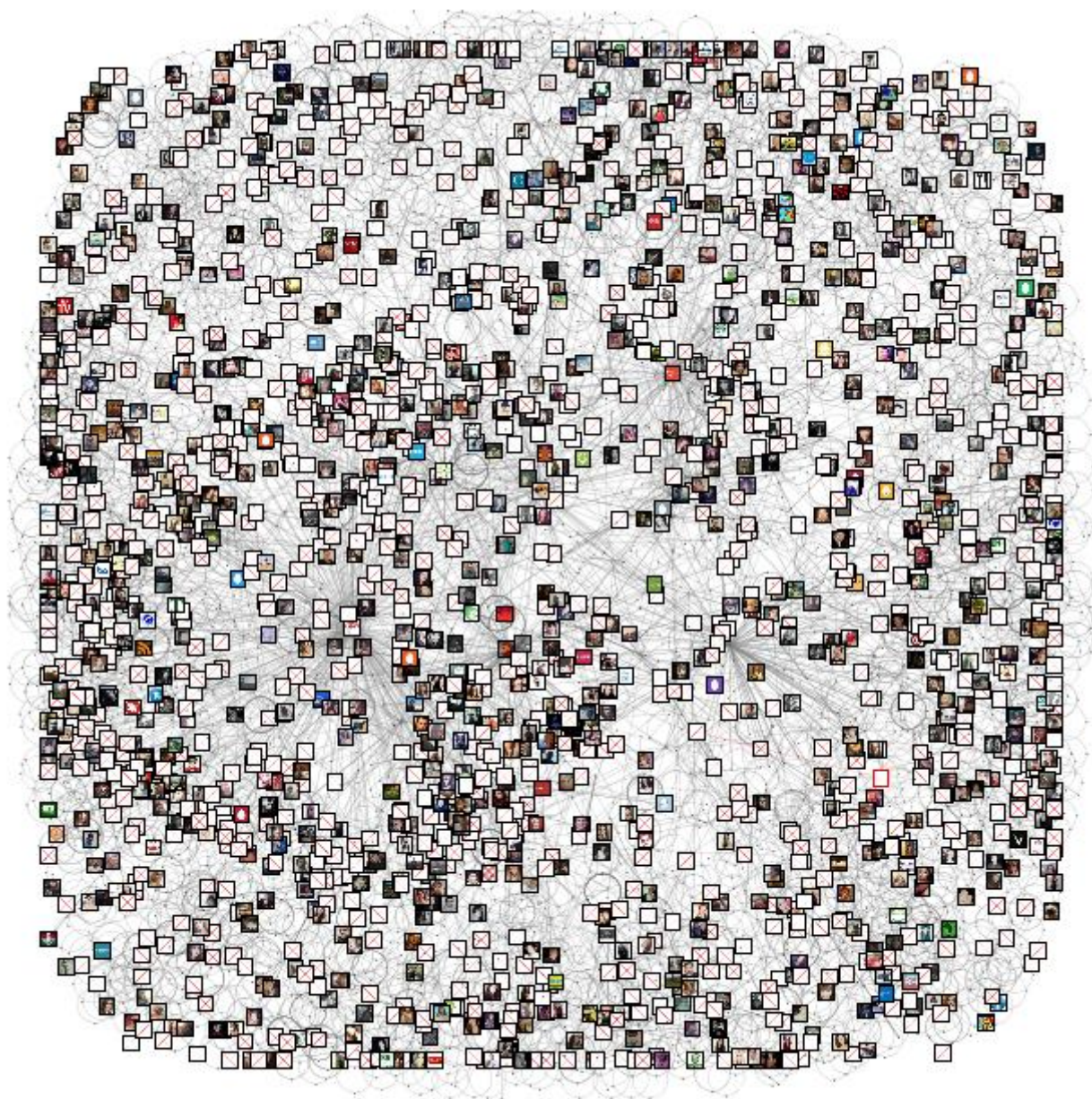
Figura 45 - Resultados NodeXL – Glúten



Created with NodeXL (<http://nodexl.codeplex.com>)

Fonte: Autor (2016).

Figura 46 - Resultados NodeXL - Lactose



Created with NodeXL (<http://nodexl.codeplex.com>)

Fonte: Autor (2016).

Figura 48 - Nuvem de Tags - Clube do Zero



Fonte: Autor (2016).

Na Figura 4, visualiza-se as palavras-chave mais utilizadas pela ALLPS Alimentos Saudáveis.

Figura 49 - Nuvem de Tags - ALLPS Alimentos



Fonte: Autor (2016).

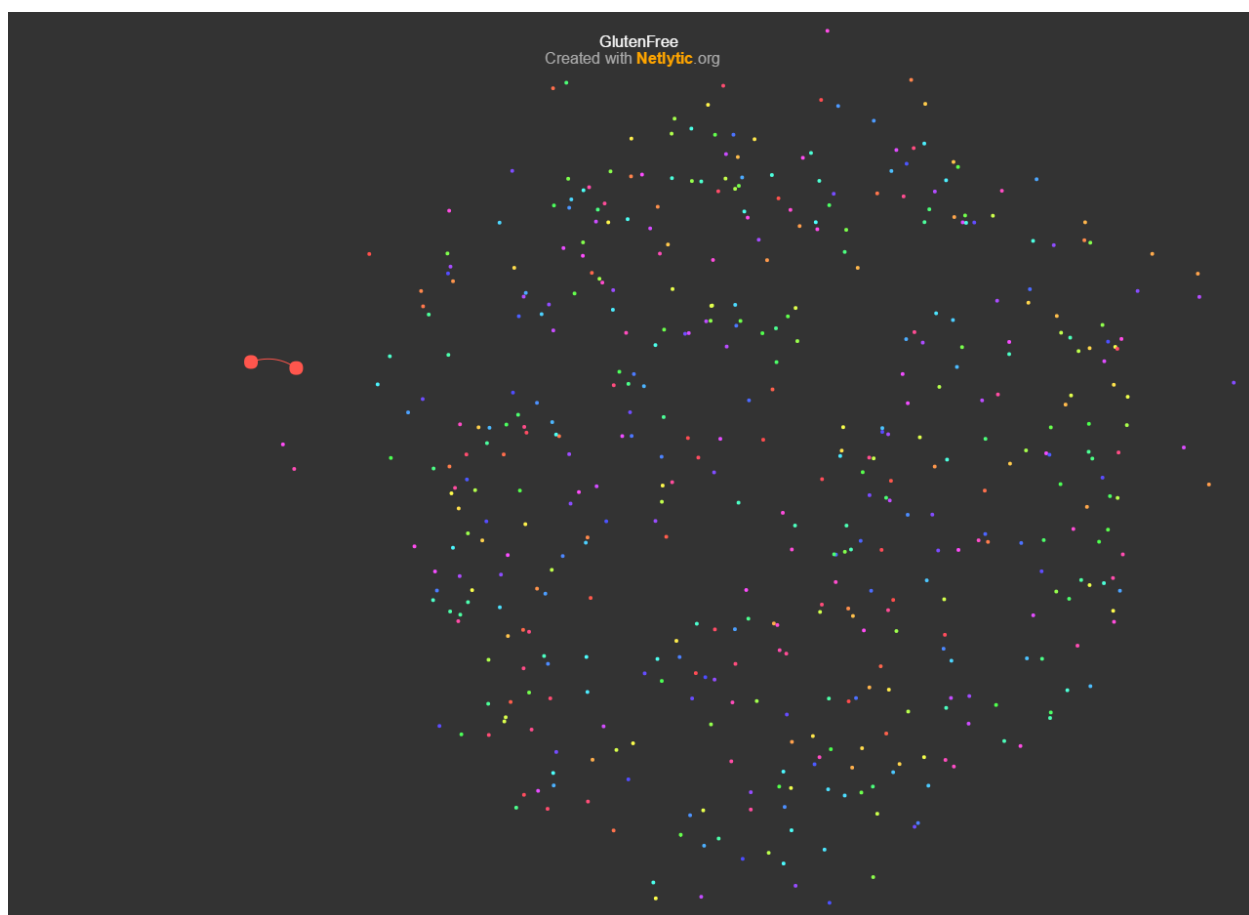
Depois da análise dessas Nuvens de *Tags* geradas por cada empresa, é possível analisar que as três empresas possuem como foco produtos referentes ao Glúten e a Lactose, com maior ênfase no Glúten.

Subsequentemente a apresentação destes dados para a empresa, no questionário do ANEXO D, foi possível identificar que a ALLPS Alimentos possuirá um diferencial, em comparação aos seus concorrentes, incluído em seu produto alimentos com outros tipos de características, como *Diet*, *Light*, Sem Ovos, entre outros.

A coleta realizada no *Facebook* foi transformada também em uma rede de interações, na qual é possível analisar o volume de dados de cada empresa gera na rede social.

Na Figura 50 é possível analisar a rede da empresa Glúten *Free Box*.

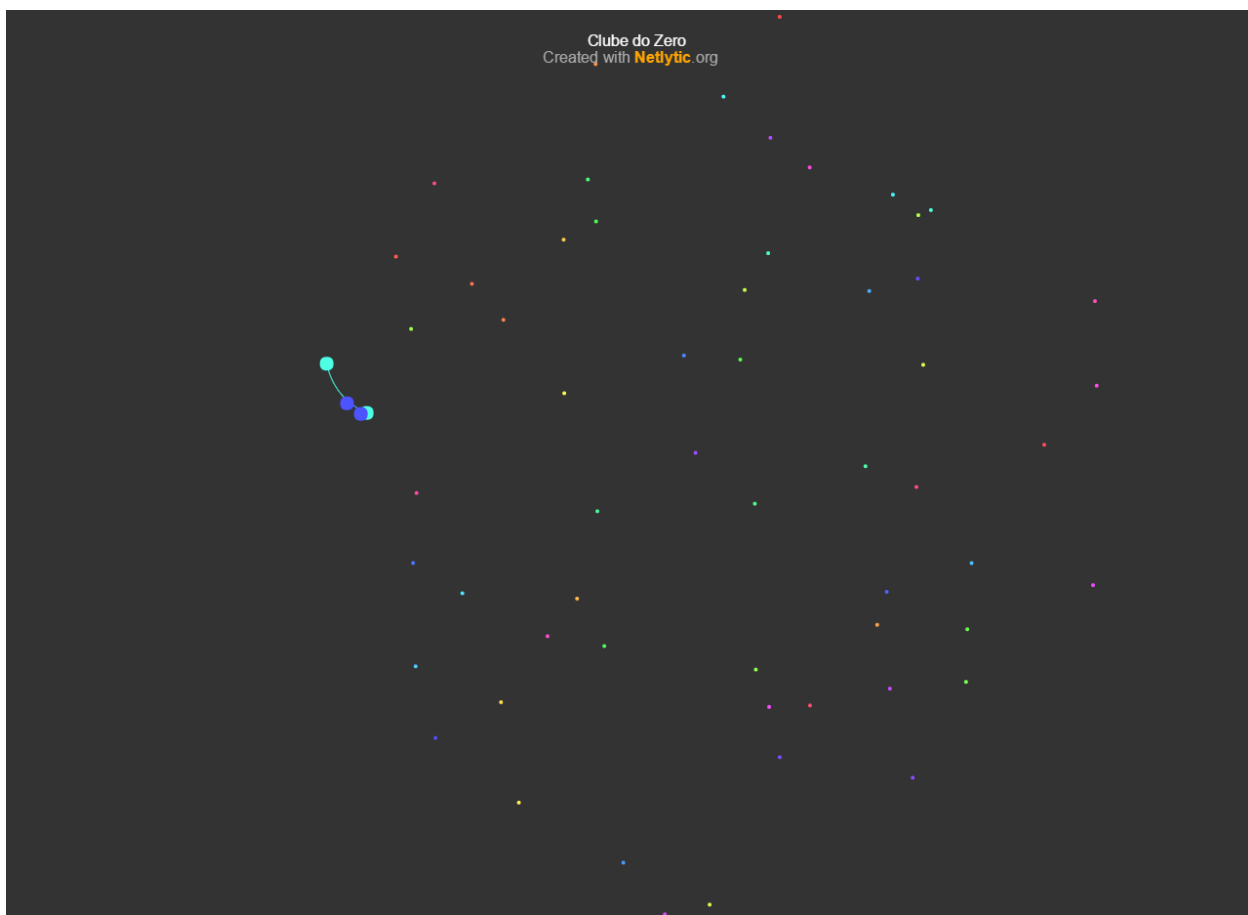
Figura 50 - Network *Facebook* - Glúten Free



Fonte: Autor (2016).

Na Figura 51, é possível visualizar a rede da empresa Cube do Zero.

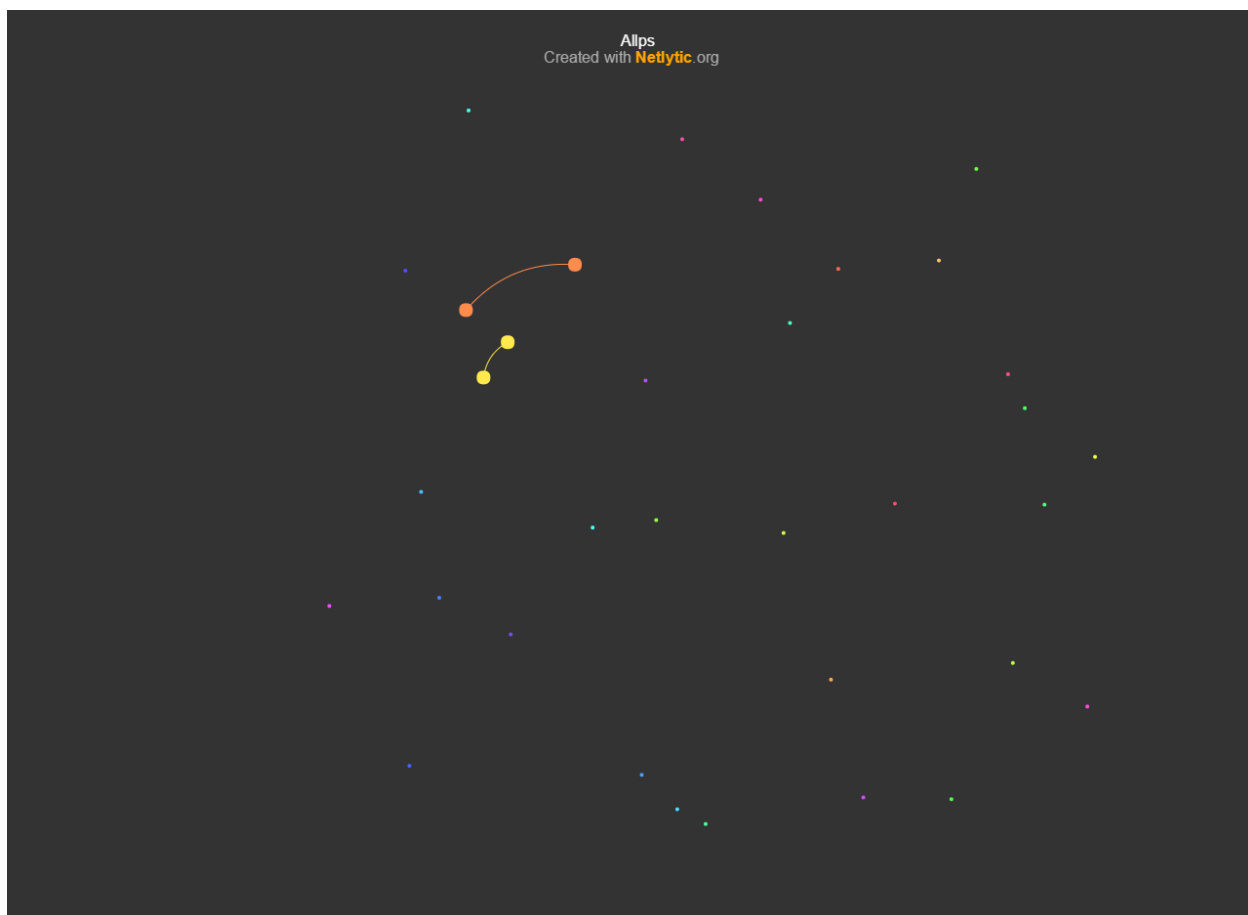
Figura 51 - Network *Facebook* - Clube do Zero



Fonte: Autor (2016).

Na Figura 52, visualiza-se a rede da empresa ALLPS.

Figura 52 - Network *Facebook* - ALLPS Alimentos



Fonte: Autor (2016).

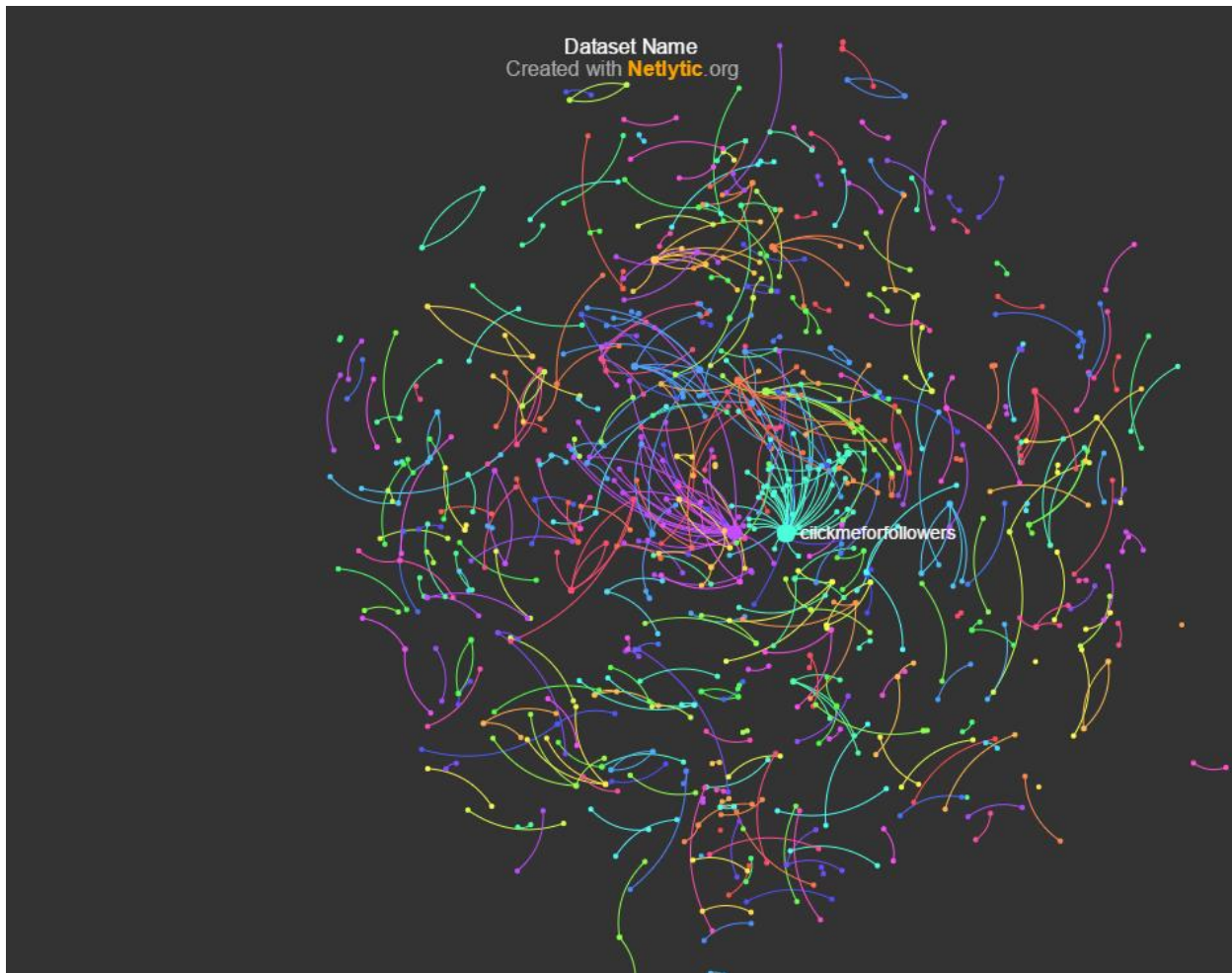
Pelo *Facebook* ser um dos principais meios de contato com o cliente e a principal ferramenta de Marketing Digital da empresa ALLPS, pode-se observar a análise de rede da empresa, comparando com suas concorrentes, que ainda há uma diferença de iterações entre a ALLPS e seus concorrentes. Devido a isso, a empresa a necessidade de investir mais no Marketing Digital para possuir mais acessos e iterações de clientes e possíveis clientes em sua rede social.

5.7.3.2 Análise coleta Instagram

A coleta realizada na rede social *Instagram*, buscando postagens com as #Gluten, #Lactose, #Diet, #Light, foi transformada em um grafo de *network*, no qual é

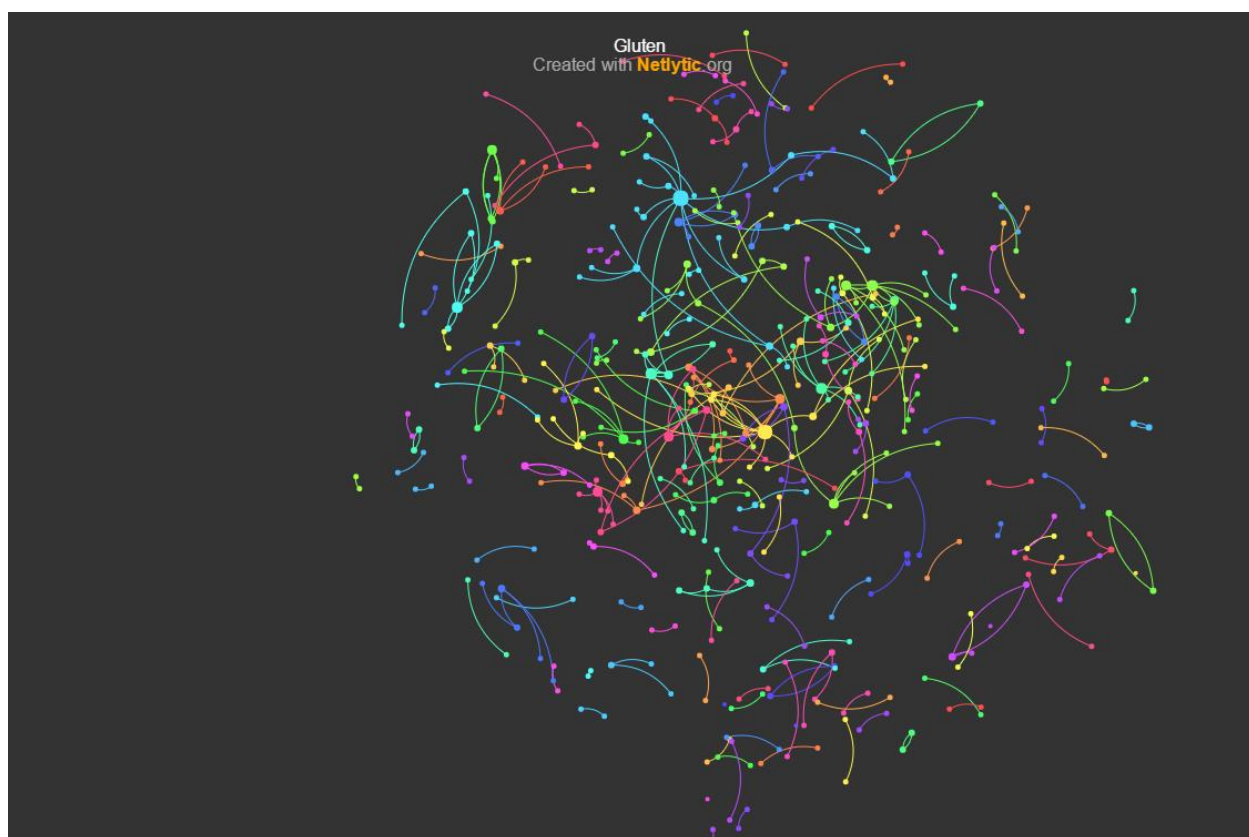
possível analisar o volume de posts realizados com cada palavra-chave, conforme Figuras 53, 54, 55 e 56.

Figura 53 - Resultado Instagram – Diet



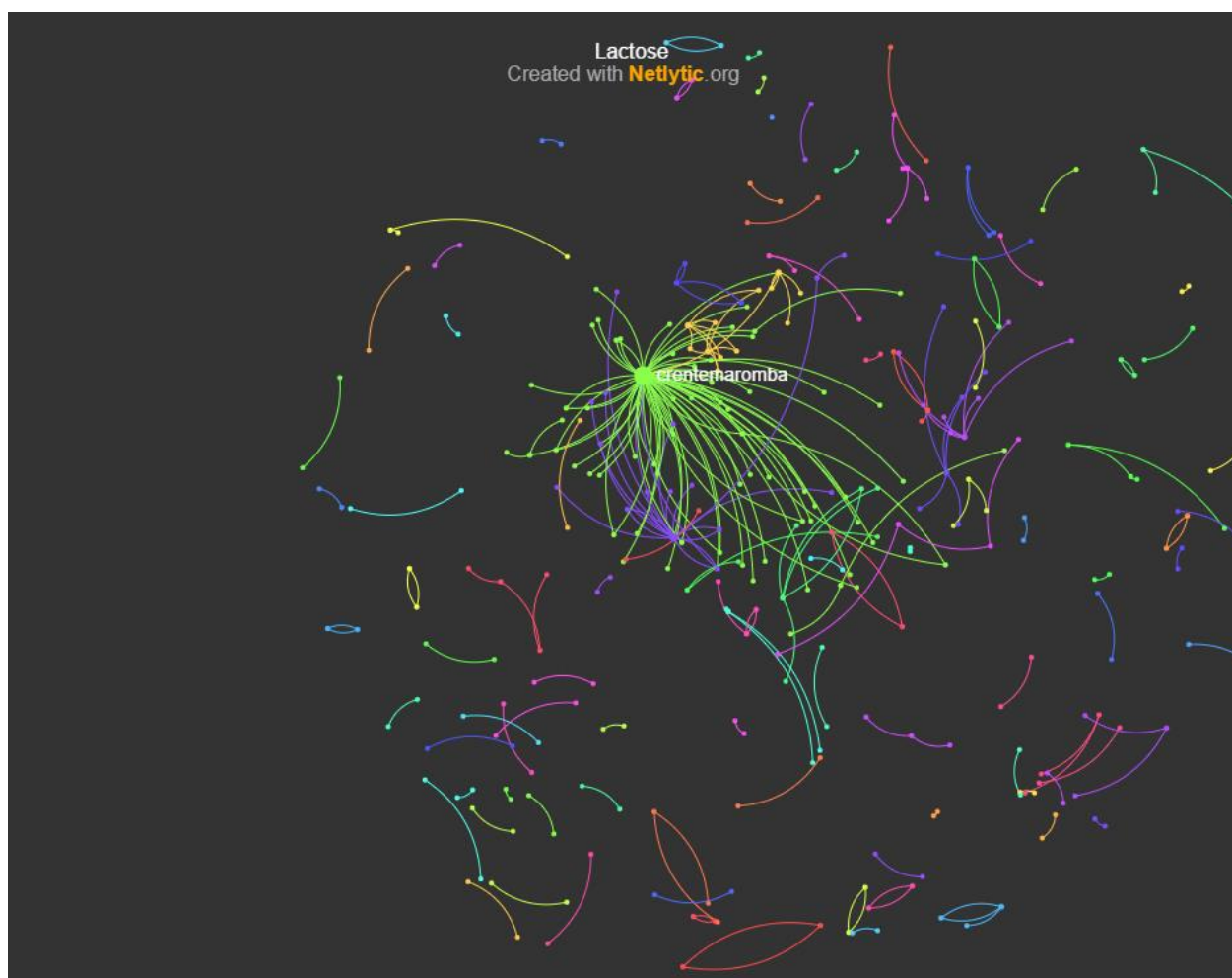
Fonte: Autor (2016).

Figura 54 - Resultado Instagram – Glúten



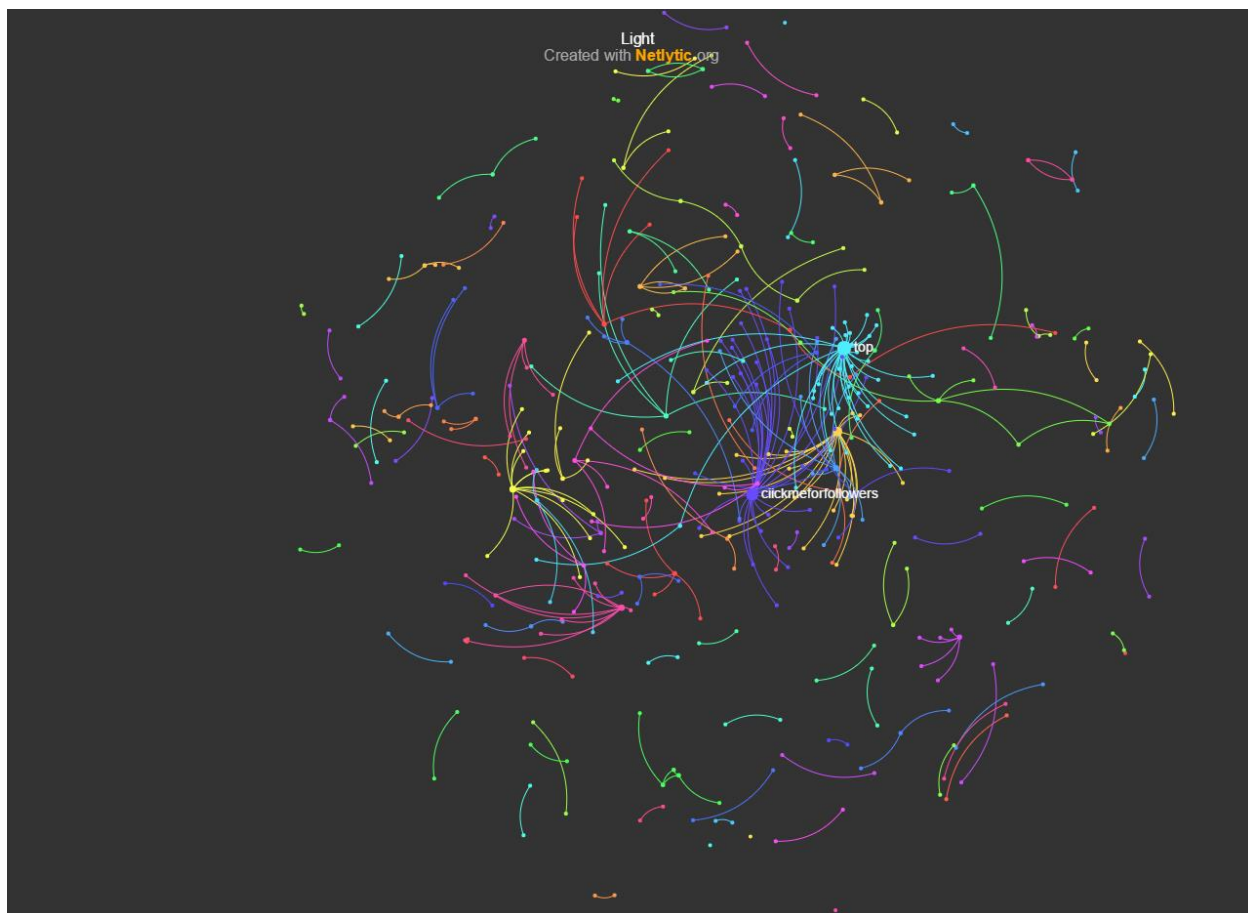
Fonte: Autor (2016).

Figura 55 - Resultado Instagram – Lactose



Fonte: Autor (2016).

Figura 56 - Resultado Instagram - Light



Fonte: Autor (2016).

Por meio destes grafos foi possível apresentar a empresa que na rede social *Instagram*, diferente do apresentado nas coletas do *Twitter*, as restrições *light* e *diet* aparecem em um volume muito próximo ao Glúten e Lactose, mostrando assim que a inclusão destes tipos de produtos e um Marketing Digital focado no *Instagram* pode aproximar um público ainda desconhecido pela empresa.

5.7.4. Nuvem de Tags

Para uma melhor visualização dos termos mais comentados no *Twitter*, referente a alimentação saudável, foi gerada uma Nuvem de Tag no website Tagul. Este site permite a inclusão de um texto no qual o próprio site analisa e apresenta os termos mais encontrados neste arquivo, conforme Figura 57.

Figura 57 - Tela Inicial - Tagul

The screenshot displays the Tagul website's main interface for creating word clouds. At the top, there is a navigation bar with the Tagul logo, links for GALLERY, PRICING, and COMMUNITY, and buttons for CREATE, SIGN UP, and Login. A 'Hide tips' button is also present. Below the navigation bar, a green tip box provides instructions on disabling word repetition. The main interface is divided into two main sections: 'Words' on the left and 'Visualize' on the right. The 'Words' section includes a table for adding words, buttons for 'Import words', 'Add', 'Remove', and 'Clear all', and a list of customization options like 'Shapes', 'Fonts', 'Layout', 'Colors and Animations', and 'Download and Share'. The 'Visualize' section features a large text input area with the prompt 'Input words Click Visualize Customize', a 'Visualize' button, and a 'Have Fun ;-)' message. At the bottom, there are buttons for 'Order a poster, t-shirt, mug and more!', 'Print', 'Save changes', and a close button.

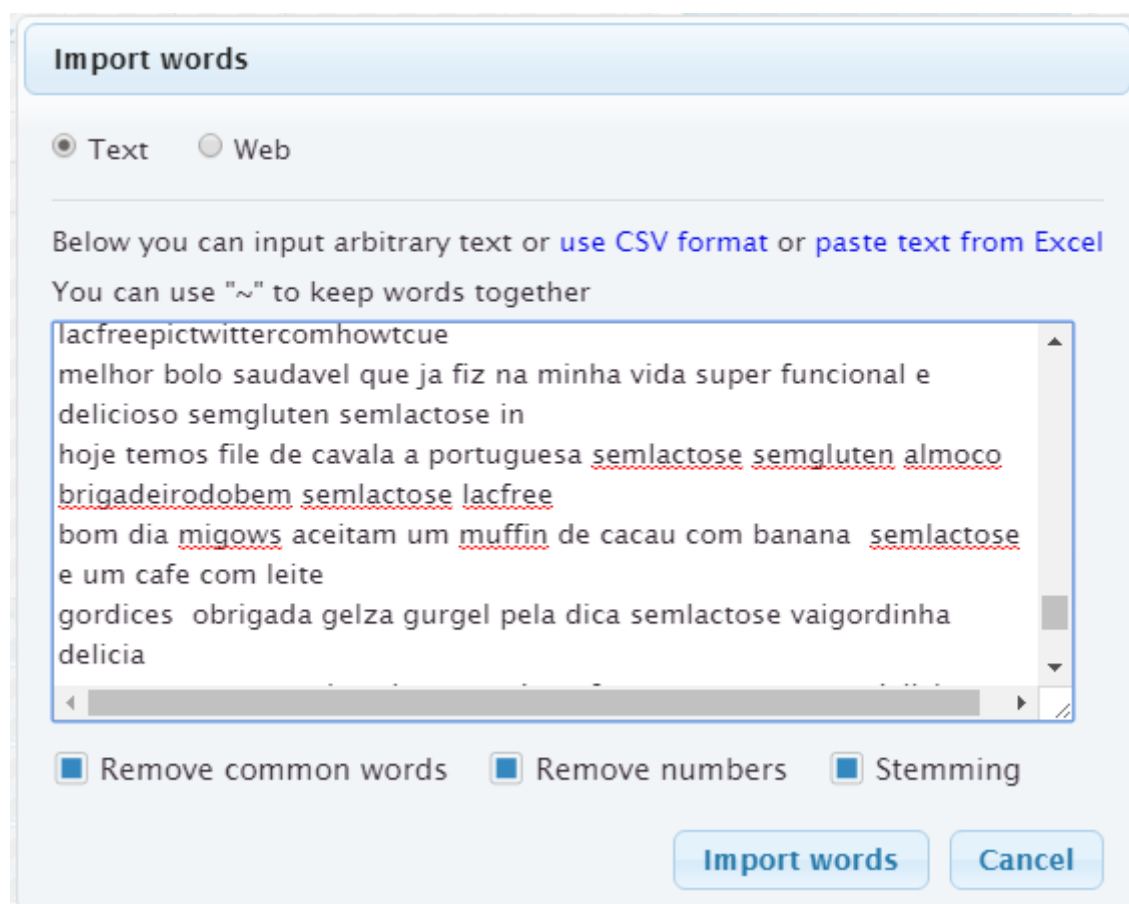
Text	Size	Color	Angle	Font	Repeat?
Word	1	Random	Random	Random	☒
Cloud	1	Random	Random	Random	☒

Fonte: Tagul (2016).

Após clicar em “*Import Words*”, é possível colar o texto na janela que será apresentada.

Para este estudo de caso, utilizou-se o conteúdo do arquivo .txt que continha todos os *tweets* capturados para a análise realizada no WEKA, conforme Figura 58.

Figura 58 - Importação texto



Fonte: Tagul (2016).

Depois de clicar em “*Import Words*”, o sistema possibilita a remoção de palavras que não insignificantes para este estudo. Neste caso foram removidas palavras como “Eu”, “Com”, “Eles”, entre outras, conforme Figura 59.

Figura 59 - Limpeza de Palavras

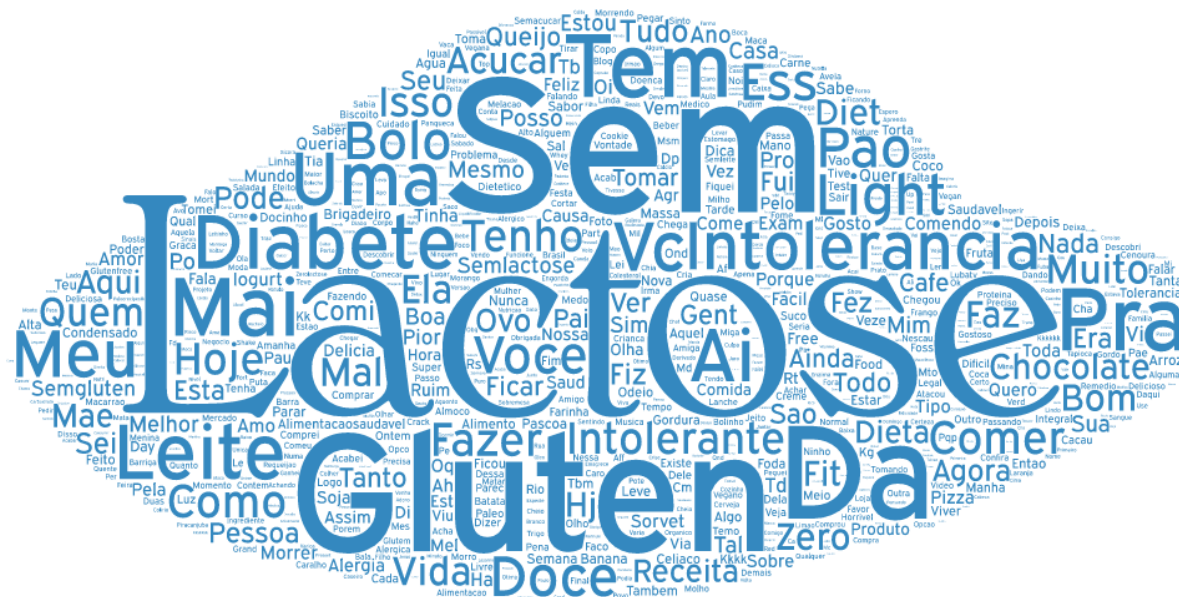
Text	Size	Color	Angle	Font	Repeat?
Lactose	95583	Random	Random	Random	☒
Sem	71791	Random	Random	Random	☒
Gluten	39719	Random	Random	Random	☒
Diabet	36807	Random	Random	Random	☒
Com	34289	Random	Random	Random	☒
Eu	31892	Random	Random	Random	☒
Nao	29302	Random	Random	Random	☒
Da	22232	Random	Random	Random	☒
Intolerancia	21602	Random	Random	Random	☒
Um	19212	Random	Random	Random	☒

↑ UPPER ↓ lower ⇅ Capitalize Links options

Fonte: Tagul (2016).

Após clicar em “Visualize”, o sistema gera uma nuvem de *Tags*, conforme Figura 60, na qual é possível visualizar quais palavras foram mais comentadas em toda coleta realizada no *twitter*, neste estudo de caso.

Figura 60 - Resultado - Nuvem de Tags



Fonte: Autor (2016).

Analisando a Figura 60, após retorno do questionário, apresentado no anexo D, foi possível identificar que a empresa está acertando na venda de produtos sem glúten e sem lactose e devem investir no Marketing Digital destes produtos. Também pode-se identificar um volume de interesse em produtos *Diet* e *Light* e que a ALLPS deve investir e incluir estes tipos de produtos em sua caixinha.

Foi identificado que alguns produtos mais comentados pelos usuários do *twitter*, são bolos, chocolates, pão, queijo, iogurte, entre outros. Com essa informação, a empresa ALLPS irá solicitar aos seus representantes que encaminhem mais destes alimentos para incluírem em seu produto, assim, atendendo de forma mais direta seus clientes, divulgando produtos de interesse conhecido, assim trabalhando com o marketing *One-to-One*, um marketing mais direcionado.

6. CONCLUSÃO

Durante o desenvolvimento deste estudo de caso, na coleta dos dados, foi identificado a necessidade da coleta em um período maior do que o tempo definido inicialmente, que eram de 2 meses. Foram coletados então dados de 21 meses, para que se pudesse acumular um número significativo de dados a serem analisados.

Na aplicação das técnicas de mineração de dados, foi encontrada algumas dificuldades para se obter resultados. Posteriormente a aplicação do algoritmo de agrupamento K-Means, identificou-se que seriam necessárias mais categorias, para que o agrupamento fosse mais trabalhado e trouxesse mais informações, como gênero, idade dos usuários, localidade, entre outros. O uso das regras de associação, por meio do algoritmo Apriori foi mais acertado para capturar os termos que estavam mais juntos.

No decorrer deste estudo de caso, outras ferramentas foram analisadas e utilizadas para complementar as informações obtidas na mineração de dados: NodeXL, Netlytic e Tagul. Essas ferramentas acrescentaram mais informações sobre as informações geradas pela mineração de dados, e puderam dar maior subsídio para a análise da empresa do estudo de caso.

Foi necessário a realização de pré-processamento nos arquivos gerados pela aplicação Java, para após serem aplicados na ferramenta WEKA. Foi utilizado o programa Excel, onde foram aplicadas fórmulas de busca de texto para analisar os tweets coletados.

Em termos gerais, após a apresentação das análises realizadas a empresa ALLPS Alimentos, este trabalho atingiu seus objetivos com sucesso em auxiliar a empresa a conhecer melhor seu público alvo para realizar um Marketing Digital mais efetivo. Conhecendo melhor seu público alvo, a ALLPS poderá realizar um marketing *One-to-One*, atendendo com mais assertividade as necessidades de seus clientes ou futuros clientes. Foi possível também apresentar a ALLPS uma forma de se diferenciar de seus concorrentes e a importância de possuir um marketing mais bem trabalhado em suas redes sociais, aumento assim as interações com seus clientes.

Durante este estudo de caso, a empresa ALLPS foi auxiliada na construção de sua análise SWOT. Posteriormente a apresentação dos resultados, algumas informações foram alteradas da Análise SWOT inicial, conforme apresentado na Figura 61.

Figura 61 - Análise SWOT atualizada - ALLPS

	Ajuda	Atrapalha
Interna	Forças: • Mapeamento frequente da concorrência • Agilidade no atendimento ao consumidor • Pró atividade da equipe • Produtos de Qualidade • Blog com informações e receitas • Identidade visual atrativa • Mais opções de pagamento (boleto) • Controle de qualidade com nutricionista • Foco em produtos sem Glúten, Lactose, Diet, Light. • Conhecimento amis abrangente do mercado de atuação.	Fraquezas: • Poucos recursos financeiros para investimento • Baixa prospecção presencial • Indisponibilidade de horário • Processo de entrega (Alto custo e prazo longo) • Dificuldade de contato com fornecedores • Site não responsivo
Externa	Oportunidades: • Possibilidade de abrir um e-commerce • Crescimento do mercado de alimentação saudável • E-commerce já é um espaço de compras consolidado • Número alto de pessoas com restrições alimentares • Diferencial de produtos perante aos concorrentes.	Ameaças: • Concorrência forte • Carga Tributária elevada para RS • Custo do Frete • Crise

Fonte: Autor (2016).

Como conclusão deste estudo de caso foi possível identificar a importância da Mineração de Dados para os negócios e o quanto foi importante obter informações que podem ser utilizadas diretamente no marketing realizados pelas empresas. Há outras formas de realizar a mineração de dados, até mesmo em outras bases de dados além das redes sociais, mas neste estudo de caso, foi possível identificar que as redes sociais são ricas em conteúdo e que podem auxiliar as empresas em diversas áreas.

Pode-se identificar também que para um bom trabalho de marketing digital é essencial ter um profissional de TI qualificado para auxiliar na exploração dos dados com as ferramentas adequadas trabalhando colaborativamente com um especialista da área do negócio.

Posteriormente as análises apresentadas a empresa ALLPS, a mesma iniciou seus trabalhos no Marketing Digital da empresa, publicando em suas redes sociais posts abrangendo mais restrições alimentares, assim trabalhando com o Marketing *One-to-One*, ou seja, diretamente a necessidade de seu público e o Marketing *Pull*, tornando a empresa mais visível quando algum usuário buscar por algum tipo de restrição.

6.1. TRABALHOS FUTUROS

Para trabalhos futuros, há a necessidade de uma análise de dados de perfis de usuários que consomem produtos direcionados a alguma restrição alimentar, ou então, uma análise de sentimentos com relação a alimentos específicos, direcionados a alguma restrição.

Também seria interessante uma análise de geolocalização, sendo possível assim identificar regiões do país que podem ser melhor exploradas por meio do marketing digital.

Poderiam ser aplicados outros algoritmos de mineração de dados em outros tipos de base de dados, que venham de outras ferramentas que não sejam as redes sociais.

Outra sugestão seria a realização de uma pesquisa on-line, seja em um website ou em redes sociais, obtendo assim dados mais concretos sobre as necessidades dos usuários, bem como, obtendo maiores informações sobre os perfis destes usuários.

REFERÊNCIAS

ADOLPHO, Conrado. **Os 8 P's do Marketing Digital**. O Guia Estratégico de Marketing Digital. 3. ed. São Paulo: Novatec, 2010.

AMO, Sandra de. **Técnicas de Mineração de Dados**. Universidade Federal de Uberlândia. p. 17. 2004. Disponível em: < <http://www.deamo.prof.ufu.br/arquivos/JAI-cap5.pdf>>. Acesso em: 22 mai. 2016.

AGRAWAL, Rakesh; IMIELINSKI, Tomasz; SWAMI, Arun. **Mining Association Rules between Sets of Items in Large Databases**. Acm Sigmod Conference on Management of Data. p. 207 – 216, Washington, DC, USA, 1993.

ARANHA, Cristian Nunes. **Uma Abordagem de Pré-Processamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional**. 2007. Tese de Doutorado (Doutor em Engenharia Elétrica) -PUC – Rio. Rio de Janeiro, 2007.

BARION, E. C. N.; LAGO, D. Mineração de Textos. **Revista de Ciências Exatas e Tecnologia**, Valinhos, SP, p. 123-140, 08 dez. 2008.

BASTOS, Marcelo. Análise Swor (Matriz) – Conceito e Aplicação. **Portal Administração – Tudo sobre administração**. 2015. Disponível em: <<http://www.portal-administracao.com/2014/01/analise-swot-conceito-e-aplicacao.html>>. Acesso em: 01 nov. 2016.

BERGSTRÖM, Thamwika; BÄCKMAN, Lisa. Marketing and PR in Social Media. **How the utilization of Intragram builds and maintains customer relationships**. 2013. Disponível em: <<http://www.diva-portal.org/smash/get/diva2:625012/FULLTEXT01.pdf>>. Acesso em: 12 mai. 2016.

BEVENUTO, Fabrício. Redes sociais online: Técnicas de coleta, abordagens de medição e desafios futuros. In **Tópicos em Sistemas Colaborativos, Interativos, Multímídia, Web e Banco de Dados**, Sociedade Brasileira de Computação, Belo Horizonte, capítulo 2, p. 41–70, 2010.

BOONE, Louis E.; KURTZ, David L. **Marketing Contemporâneo**. 12. Ed. São Paulo: Cengage, 2009.

BRAGA, Luis Paulo Vieira. **Introdução à Mineração de Dados**. 2. Ed. Rio de Janeiro: E-Papers, 2005.

CECI, Flávio. **Um Modelo Semiautomático para a Construção e Manutenção de Ontologias a partir de Bases e Documentos Não Estruturados**. 2010. 131 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Florianópolis, 2010.

CIOŚ, Krzysztof. J; PEDRYCZ, Witold; SWINIARSKI, Roman. W; KURGAN, Lukasz. A. **Data Mining - A Knowledge Discovery Approach**. out, 2007.

DINIZ, Lucas. Como identificar seu público-alvo em redes sociais. **Exame.com**, ago. 2013. Seção PME. Disponível em: <<http://exame.abril.com.br/pme/noticias/como-identificar-seu-publico-alvo-em-redes-sociais>>. Acesso em: 24 mai. 2016.

DRUBSCKY, Luiza. Quais são as redes sociais mais utilizadas no Brasil. **Marketing de Conteúdo**, jun. 2015. Disponível em: <<http://marketingdeconteudo.com/redes-sociais-mais-usadas-no-brasil/>>. Acesso em: 16 jun. 2016.

ELLIOTT, Nate. **Instagram Is The King Of Social Engagement**. **Forrester**. abr. 2014. Disponível em: <http://blogs.forrester.com/nate_elliott/14-04-29-Instagram_is_the_king_of_social_engagement>. Acesso em: 10 mai. 2016.

EMIRBAYER, Musfata, GOODWIN, Jeff. Network analysis, culture and the problem of agency. **American Journal of Sociology**, v. 99, n. 6, p. 1411-1454, 1994.

FAYYAD, Usama M.; PIATETSKY-SHAPIO, Gregory; SMYTH, Padhraic; UTHURUSAMY, Rumasamy. **Advances in knowledge discovery and data mining**. [S.l.]: AAAI/MIT Press, 1996.

FERREIRA, Juliana Xavier; NEVES, Leonardo Holzmman. A utilização das redes sociais: um estudo de caso sobre o *Facebook* como estratégia de marketing no banco Bradesco S.A. **Revista Gestão Premium/Cursos de Administração e Ciências Contábeis – FACOS/CNEC**. Osório, dez, 2013. Disponível em: <http://facos.edu.br/publicacoes/revistas/gestao_premium/dezembro_2013/pdf/a_utilizacao_das_redes_sociais_-_um_estudo_de_caso_sobre_o_Facebook_como_estrategia_de_marketing_no_banco_bradesco_s.a.pdf>. Acesso em: 22 mai. 2016.

FILHO, José Adail Carvalho. Mineração de Textos: Análise de Sentimento utilizando *Tweets* referente à Copa do Mundo 2014. 2014. **Trabalho de Conclusão de Curso**. Universidade Federal do Ceará. Bacharelado em Engenharia de Software, Quixadá, 2014.

FOLHA de São Paulo. **Twitter perde valor e vive em meio a rumores de aquisição**. Disponível em: <<http://www1.folha.uol.com.br/tec/2016/01/1731938-twitter-perde-valor-e-vive-em-meio-a-rumores-de-aquisicao.shtml>> Acesso em: 14 mai. 2016.

FONSECA, João José Saraiva da. **Metodologia da pesquisa científica**. 2002. Disponível em: <<https://books.google.com.br/books?id=oB5x2SChpSEC&pg=PA20&dq=PESQUISA+#+v=onepage&q=PESQUISA&f=false>>. Acesso em: 22 mai. 2016.

GALVÃO, Noemi Dreyer; MARIN, Heimar de Fátima. Técnica de mineração de dados: uma revisão da literatura. **SciELO**. São Paulo, v. 22, n. 5, out. 2009.

GOLDENBERG, Jacob; LIBAI, Barak; MULLER, Eitan. Talk of the network: a complex system look at the underlying process of word-of-mouth. **Marketing Letters**, v. 12, n. 3, p. 211-223, 2001.

GOLDSHMIDT, Ronaldo, PASSOS, Emmanuel; BEZERRA, Eduardo. **Data Mining, Conceitos, técnicas, algoritmos, orientações e aplicações**. 2. Ed. Rio de Janeiro: Elsevier, 2015.

GONÇALVES, Marcus Paulo. Empresas brasileiras que fazem sucesso no *Twitter*. **Enlink**, out. 2012. Seção Social Media. Disponível em: <<http://www.agenciaenlink.com.br/blog/empresas-brasileiras-que-fazem-sucesso-no-twitter/>>. Acesso em: 22 mai. 2016.

GPEA, Grupo de Pesquisa em Engenharia de Algoritmo. **Regra de Associação**. Disponível em:< <http://www.din.uem.br/~gpea/linhas-de-pesquisa/mineracao-de-dados/tarefas-de-mineracao-de-dados/regras-de-associacao/>>. Acesso em: 21 nov. 2016.

GUALDA, Paula. Marketing Digital: Estratégias, Diferenças e Potencialidades. **Slideshare**, mai. 2012. Disponível em: <<http://pt.slideshare.net/paulagualda/marketing-digital-estrategias-diferenas-e-potencialidades>>. Acesso em: 25 abr. 2016.

GUEDES, Fabrícia. A Evolução das Redes Sociais. **Insite**, 2012. Disponível em: <http://www.insite.pro.br/saladeaula/fabricia.pdf>. Acesso em: 10 jun. 2016.

HAYKIN, Simon. **Redes Neurais – Princípios e Práticas**. Porto Alegre: Bookman 2001.

HORTINHA, Joaquim. **E-Marketing, um guia para a nova economia**. Sílabo, p. 259, 2001.

HOFFMAN, Diego. **Mineração dos Dados no Twitter com Foco na Criminalidade da Região de Caxias do Sul**. 2016. Bacharel em Ciência da Computação. Universidade de Caxias do Sul, Caxias do Sul, 2016.

KENDZERSKI, Paulo Roberto. **Web Marketing e Comunicação Digital**. 2. Ed. Produção Independente. 2009.

KOTLER, Philip; KELLER, Kevin Lane. **Administração de Marketing**. 4. ed. São Paulo: Person, 2006.

LINDEN, Ricardo. Técnicas de Agrupamento. **Revista de Sistemas de Informação da FSMA**. n. 4. Pg 18-36. 2009.

LIU, Bing. Sentiment Analysis and Subjectivity. In. **Handbook of Natural Language Processing**. 2. Ed. 2010.

MARROQUÍN, Juan Carlos. Novo Consumidor: mais informado, mais exigente, mais consciente. **Folha de São Paulo**, São Paulo, out. 2014. Disponível em: <<http://www1.folha.uol.com.br/topofmind/2014/10/1528719-novo-consumidor-mais-informado-mais-exigente-mais-consciente.shtml>>. Acesso em: 25 abr. 2016.

MARTELETO, Regina Maria. Análise de redes sociais: aplicação nos estudos de transferência da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 71-81, jan./abr. 2001.

MCCUE, C. **Data Mining and Predictive Analysis - Intelligence Gathering and Crime Analysis**. Elsevier, 2007.

MEDEIROS, Saulo. Análise Swot de uma Instituição de Ensino: Descubra suas Fraquezas e Oportunidades. **Marketing Educacional Experts**. jun. 2015. Disponível em: <http://5seleto.com.br/analise-swot-de-uma-instituicao-de-ensino-descubra-suas-fraquezas-e-oportunidades/>. Acesso em: 15 out. 2016.

MOORE, H. Jason. **MDR - Multifactor Dimensionality Reduction**. jan. 2007. Disponível em: <http://www.multifactor dimensionality reduction. org/>. Acesso em: 21 mai. 2016.

MORAIS, Edilson Andrade Martins; AMBRÓSIO, Ana Paula L. **Mineração de Textos**. Goiânia: UFG. 2007. (Série Texto Técnico, INF_005/07).

MORAIS, Rodrigo. Marketing Tradicional x Marketing Digital – Entenda as diferenças. **Empreendedor X**. jun. 2013. Disponível em: <http://empreendedorx.com.br/marketing/marketing-tradicional-marketing-digital-entenda-as-diferencas>. Acesso em: 28 jul. 2016.

NETLYTIC. **About**. Disponível em: https://netlytic.org/home/?page_id=10834%20Netclit. Acesso em: 13 mai. 2016.

ORENGO, V. M; HUYCK, C. A. Stemming algorithm for the portuguese language. **Eighth Internacional Symposium on String Processing and Information Retrieval**. Pg. 186–193, 2001.

PICHILIANI, Mauro. **Data Mining na Prática: Regras de Associação**. Desenvolvimento, 2008. Disponível em: <http://imasters.com.br/artigo/7853/sql-server/data-mining-na-pratica-regras-de-associacao?trace=1519021197&source=single>. Acesso em: 15 nov. 2016.

PRYKE, Andy. SPSS Clementine. **The Data Mine**. mar. 2008. Disponível em: <http://www.the-data-mine.com/Software/SPSSClementine>. Acesso em: 13 mai. 2016.

PUHL, Elisangela. **Estudo de Perfil de Usuários de Redes Sociais através da Mineração de Dados**. 2011. Trabalho de Conclusão de Curso. Universidade de Caxias do Sul. Bacharelado em Sistemas de Informação, Caxias do Sul, 2011.

RAPOSO, Marcel Antunes. A importância do Data Mining na tomada de decisões. **Administradores.com.br**, jun. de 2010. Seção Marketing. Disponível em: <http://www.administradores.com.br/noticias/marketing/a-importancia-de-data-mining-na-tomada-de-decisoes/35063/>. Acesso em: 24 mai. 2016.

REBELATTO, Daysi; MARIANO, Enzo; CAMIOTO, Flavia; LEITE, Leonardo. **Marketing para Cursos de Engenharia**. São Paulo: Elsevier, 2014.

REZENDE, Solange O. MARCACINI, Ricardo M. MOURA, Maria F. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. **Revista de Sistemas de Informação da FSMA**. N 7. Pg 7-21, 2011.

RIBEIRO, Laura. Quais são as redes sócias mais usadas no Brasil. **Marketing de Conteúdo**. Jun. de 2016. Seção Marketing. Disponível em: <<http://marketingdeconteudo.com/redes-sociais-mais-usadas-no-brasil/>>. Acesso em: 27 jul. 2016.

SALVADOR, Alexandre Borba. **Segmentação e posicionamento: o coração do plano de marketing**. Inovcom. Vol. 6, n. 1. 2014.

SILVA, Fabrício Aguiar. A Avalanche de Dados e a Análise de Sentimentos. **Webholic**. Jan. 2010. Disponível em: <<http://webholic.com.br/analise-de-sentimento/>>. Acesso em: 14 mai. 2016.

SILVA, Fábio Ribeiro; ANGELONI, Maria Terezinha; GONÇALO, Cláudio Reis. As redes sociais digitais em estratégias de marketing: um estudo de caso em uma indústria de alimentos. **Gestão Contemporânea**, Porto Alegre, ano 10, n. 13, p. 97-121, jan/jun. 2013.

SILVA, Thales do Nascimento. **Uma Arquitetura para Descoberta de Conhecimento a Partir de Bases Textuais**. 78p. Monografia (Bacharelado em Tecnologias da Informação e Comunicação), Universidade Federal de Santa Catarina, Araranguá, 2012.

SIQUEIRA, André. Relacionamento em Marketing Digital: Por que é importante e como fazer. **Resultados Digitais**, ago. 2011. Seção Blog de Marketing de Resultados. Disponível em: <<http://resultadosdigitais.com.br/blog/relacionamento-em-marketing-digital-por-que-e-importante-e-como-fazer/>>. Acesso em: 01 jun. 2016.

SMAAL, Beatriz. A História do *Twitter*. **Tecmundo**. Fev. 2010. Seção redes sociais. Disponível em: <<http://www.tecmundo.com.br/rede-social/3667-a-historia-do-twitter.htm>>. Acesso em: 01 mai. 2016.

SOCIALMEDIA. **About Us**. Disponível em:<<http://www.smrfoundation.org/about-us/>>. Acesso em: 13 mai. 2016.

TERRA, Luciano. O Comportamento do Consumidor na Internet | Como Vender Mais. **Consultoria**, 2015. Disponível em: <<http://lucianoterra.net.br/o-comportamento-do-consumidor-na-internet-como-vender-mais/>>. Acesso em: 25 abr. 2016

VASCONCELOS, L. M. R de; CARVALHO, C. L. Aplicação de Regras de Associação para Mineração de Dados na Web. **Instituto de Informática Universidade Federal de Goiás**. Nov. 2004.

VIEIRA, Rafael. SILVA, Tarcízio. **Para entender o Monitoramento de Mídias Sociais**. 2012.

WEKA The University of Waikato. **Weka 3: Data Mining Software in Java**. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 23 mai. 2016.

WIVES, L. Tecnologias de descoberta de conhecimento em textos aplicadas à inteligência competitiva. **Exame de Qualificação EQ-069**, PPGC-UFRGS, 2002.

ANEXO A – APLICAÇÃO DO ALGORITMO K-MEANS

Neste estudo de caso foi aplicado o algoritmo K-Means nas coletas realizadas no *Twitter* pela aplicação Java. A aplicação do algoritmo foi realizada pelo programa WEKA. A análise de dados foi executada por palavra-chave capturada e por todos os dados capturas juntos.

Abaixo é possível visualizar os resultados gerados pela aplicação do algoritmo K-Means.

RESTRIÇÕES ALIMENTARES

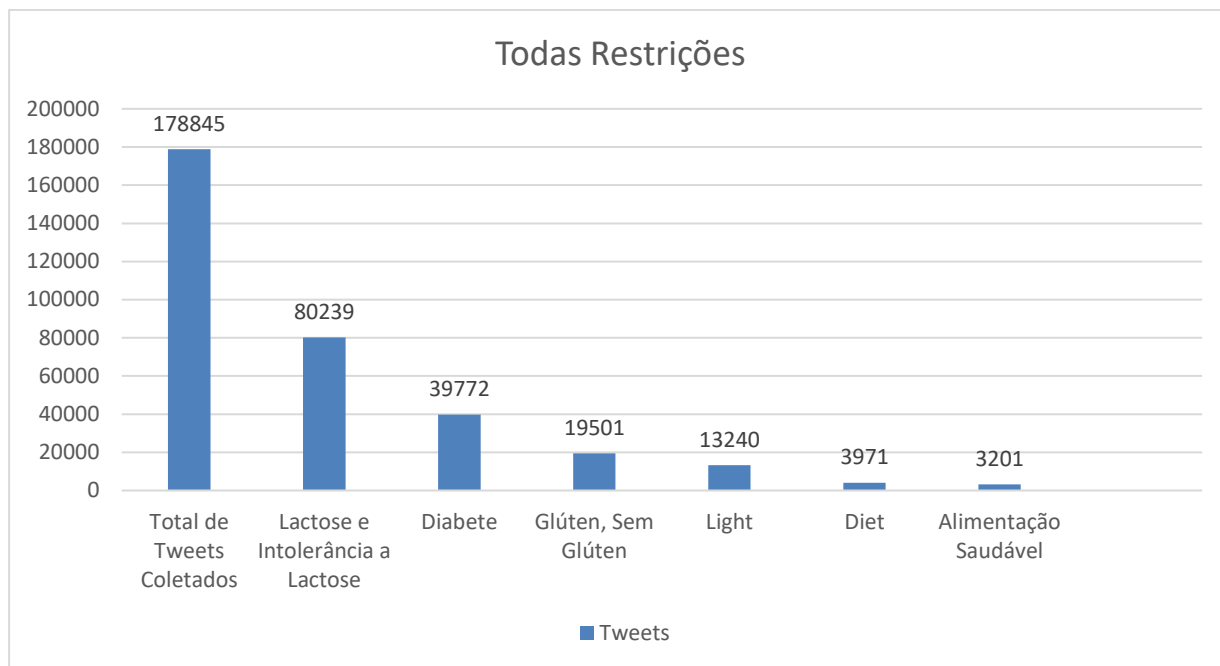
Conjunto de Todos os Dados

Final cluster centroids:

Attribute	Cluster#											
	Full Data (178845.0)	0 (19093.0)	1 (1593.0)	2 (11358.0)	3 (17328.0)	4 (39772.0)	5 (39410.0)	6 (21736.0)	7 (13240.0)	8 (8143.0)	9 (3971.0)	10 (3201.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semilactose	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Não	Não
semgluten	Não	Não	Sim	Sim	Sim	Não	Não	Não	Não	Não	Não	Não
lactose	Sim	Sim	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não	Não
gluten	Não	Não	Sim	Sim	Sim	Não	Não	Não	Não	Sim	Não	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
diet	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Sim	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
intoleranciaalactose	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim

Clustered Instances

0	19093 (11%)
1	1593 (1%)
2	11358 (6%)
3	17328 (10%)
4	39772 (22%)
5	39410 (22%)
6	21736 (12%)
7	13240 (7%)
8	8143 (5%)
9	3971 (2%)
10	3201 (2%)

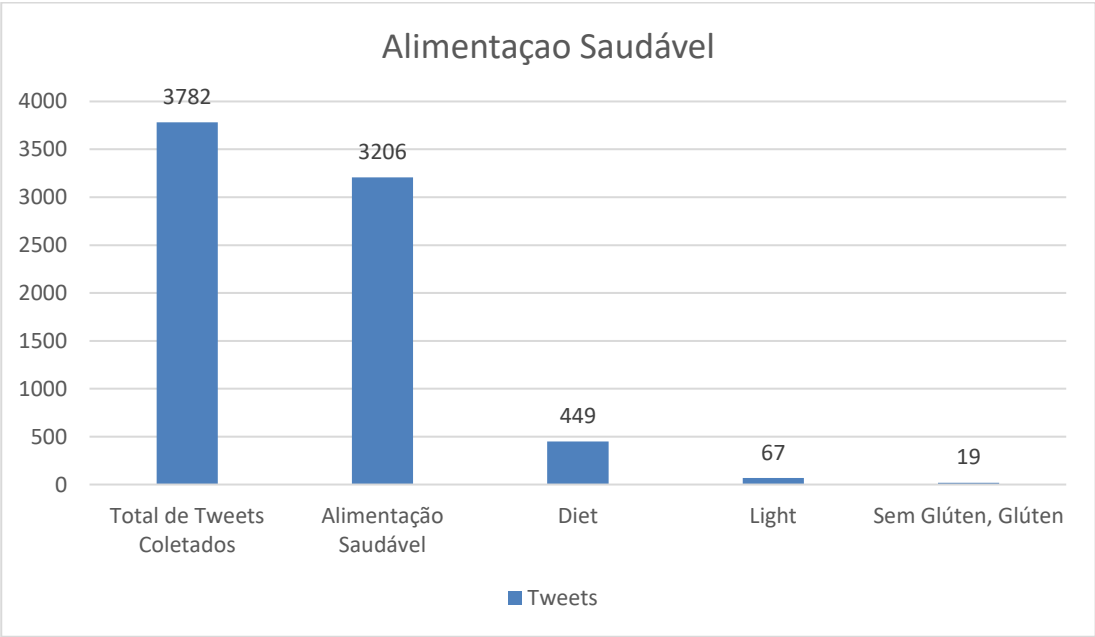


Alimentação Saudável

Final cluster centroids:													
Attribute	Cluster#												
	Full Data	0	1	2	3	4	5	6	7	8	9	10	11
	(3782.0)	(3206.0)	(3.0)	(449.0)	(19.0)	(67.0)	(1.0)	(9.0)	(9.0)	(2.0)	(11.0)	(5.0)	(1.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semilactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não
semgluten	Não	Não	Sim	Não	Sim	Não	Não	Não	Sim	Não	Sim	Não	Sim
lactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não
gluten	Não	Não	Sim	Não	Sim	Não	Sim	Sim	Sim	Não	Sim	Não	Sim
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Sim	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Sim	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim
alimentacaosaudavel	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Clustered Instances

0	3206 (85%)
1	3 (0%)
2	449 (12%)
3	19 (1%)
4	67 (2%)
5	1 (0%)
6	9 (0%)
7	9 (0%)
8	2 (0%)
9	11 (0%)
10	5 (0%)
11	1 (0%)

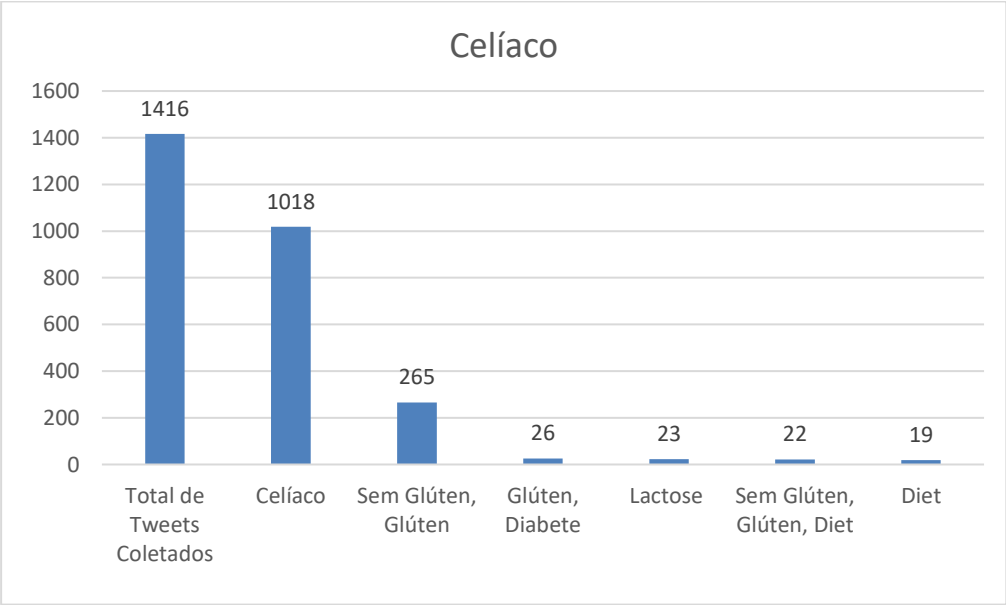


Celíaco

Final cluster centroids:													
Attribute	Full Data (1416.0)	Cluster#											
		0 (1018.0)	1 (168.0)	2 (22.0)	3 (97.0)	4 (13.0)	5 (4.0)	6 (19.0)	7 (9.0)	8 (16.0)	9 (1.0)	10 (26.0)	11 (23.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem lactose	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
sem gluten	Não	Não	Não	Sim	Sim	Não	Sim	Não	Não	Não	Não	Não	Não
lactose	Não	Não	Não	Não	Não	Sim	Sim	Não	Sim	Não	Sim	Não	Sim
gluten	Não	Não	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Sim	Sim	Não
celiaco	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Sim	Sim	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem leite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem ovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacao saudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	1018 (72%)
1	168 (12%)
2	22 (2%)
3	97 (7%)
4	13 (1%)
5	4 (0%)
6	19 (1%)
7	9 (1%)
8	16 (1%)
9	1 (0%)
10	26 (2%)
11	23 (2%)



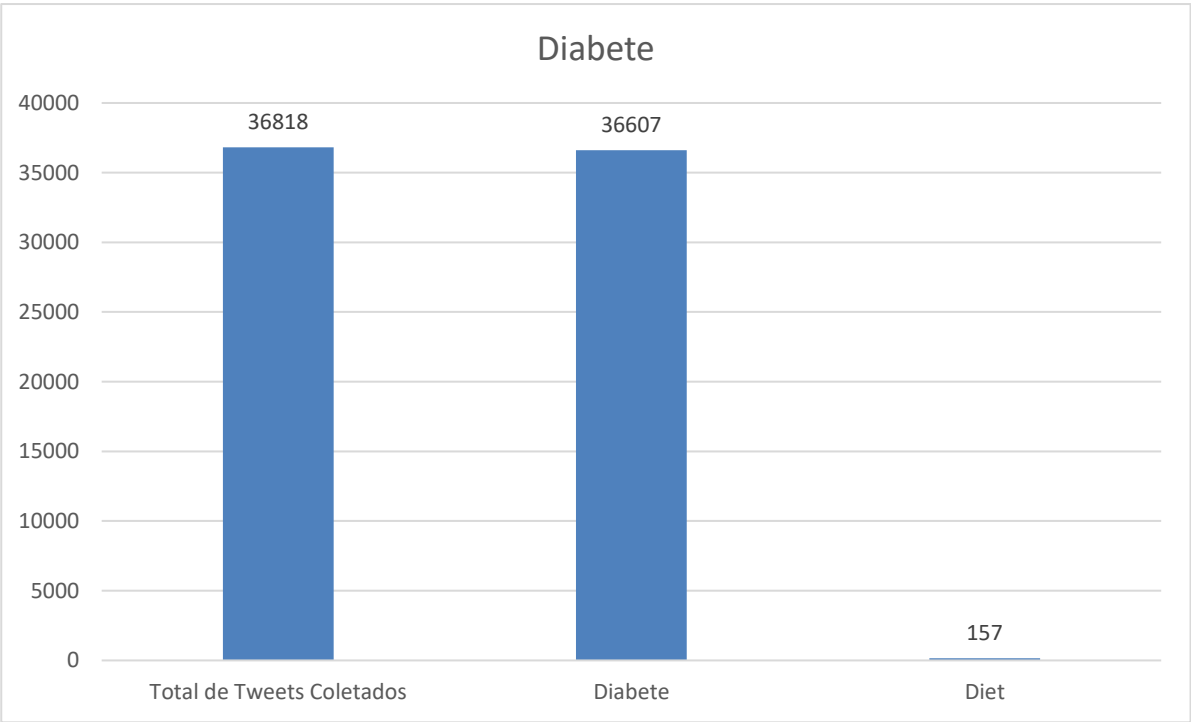
Diabete

Final cluster centroids:

Attribute	Cluster#												
	Full Data (36818.0)	0 (36607.0)	1 (21.0)	2 (157.0)	3 (5.0)	4 (13.0)	5 (4.0)	6 (4.0)	7 (2.0)	8 (1.0)	9 (1.0)	10 (1.0)	11 (2.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem lactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem gluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
lactose	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim	Sim	Não
gluten	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Sim	Sim	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
diet	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Sim	Não	Não	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
semleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	36607 (99%)
1	21 (0%)
2	157 (0%)
3	5 (0%)
4	13 (0%)
5	4 (0%)
6	4 (0%)
7	2 (0%)
8	1 (0%)
9	1 (0%)
10	1 (0%)
11	2 (0%)



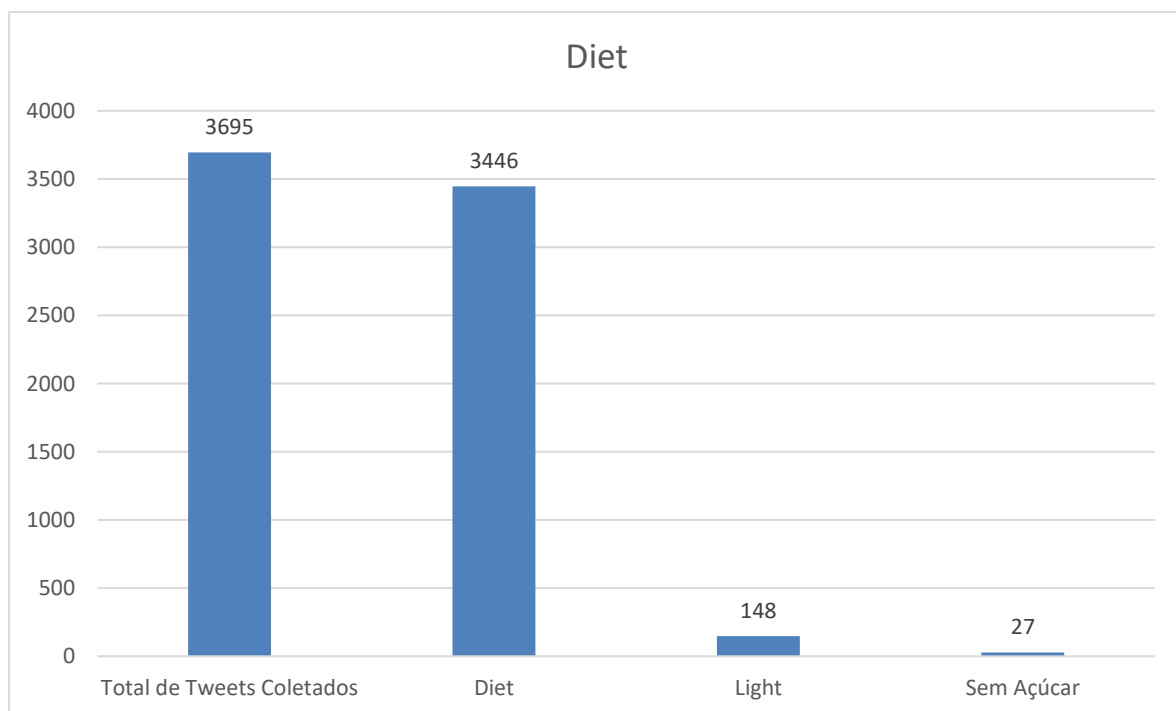
Diet

Final cluster centroids:

Attribute	Cluster#												
	Full Data (3695.0)	0 (3446.0)	1 (148.0)	2 (27.0)	3 (10.0)	4 (6.0)	5 (3.0)	6 (23.0)	7 (8.0)	8 (8.0)	9 (11.0)	10 (1.0)	11 (4.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semilactose	Não	Não	Não	Não	Sim	Sim	Não	Não	Sim	Não	Não	Não	Não
semgluten	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Não
lactose	Não	Não	Não	Não	Sim	Sim	Não	Não	Sim	Não	Sim	Não	Não
gluten	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	Não	Não	Não	Não
diet	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Sim	Sim
semleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	3446 (93%)
1	148 (4%)
2	27 (1%)
3	10 (0%)
4	6 (0%)
5	3 (0%)
6	23 (1%)
7	8 (0%)
8	8 (0%)
9	11 (0%)
10	1 (0%)
11	4 (0%)



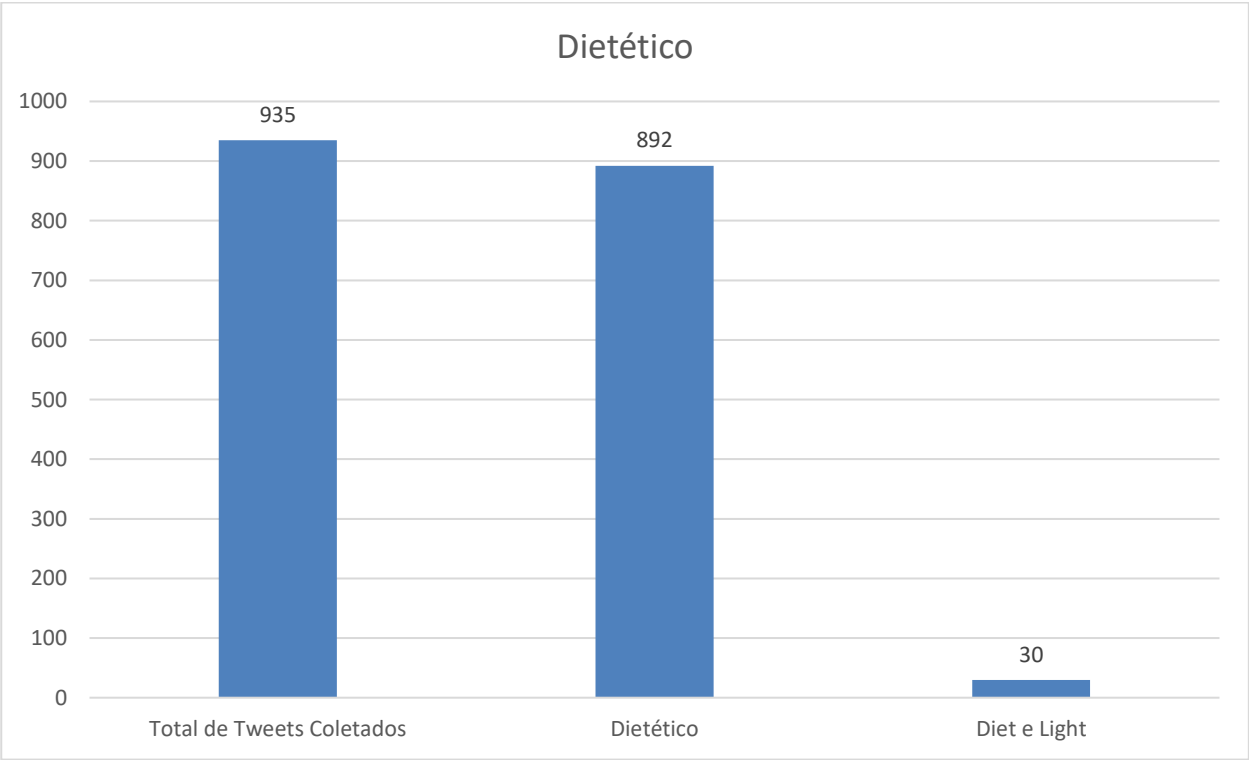
Dietético

Final cluster centroids:

Attribute	Full Data (935.0)	Cluster#									
		0 (892.0)	1 (12.0)	2 (11.0)	3 (7.0)	4 (2.0)	5 (1.0)	6 (2.0)	7 (2.0)	8 (3.0)	9 (3.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem lactose	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
sem gluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim
lactose	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
gluten	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Sim
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Sim	Não	Não	Sim	Não	Não	Não	Sim	Não
dietetico	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim
light	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
sem leite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem ovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	892 (95%)
1	12 (1%)
2	11 (1%)
3	7 (1%)
4	2 (0%)
5	1 (0%)
6	2 (0%)
7	2 (0%)
8	3 (0%)
9	3 (0%)



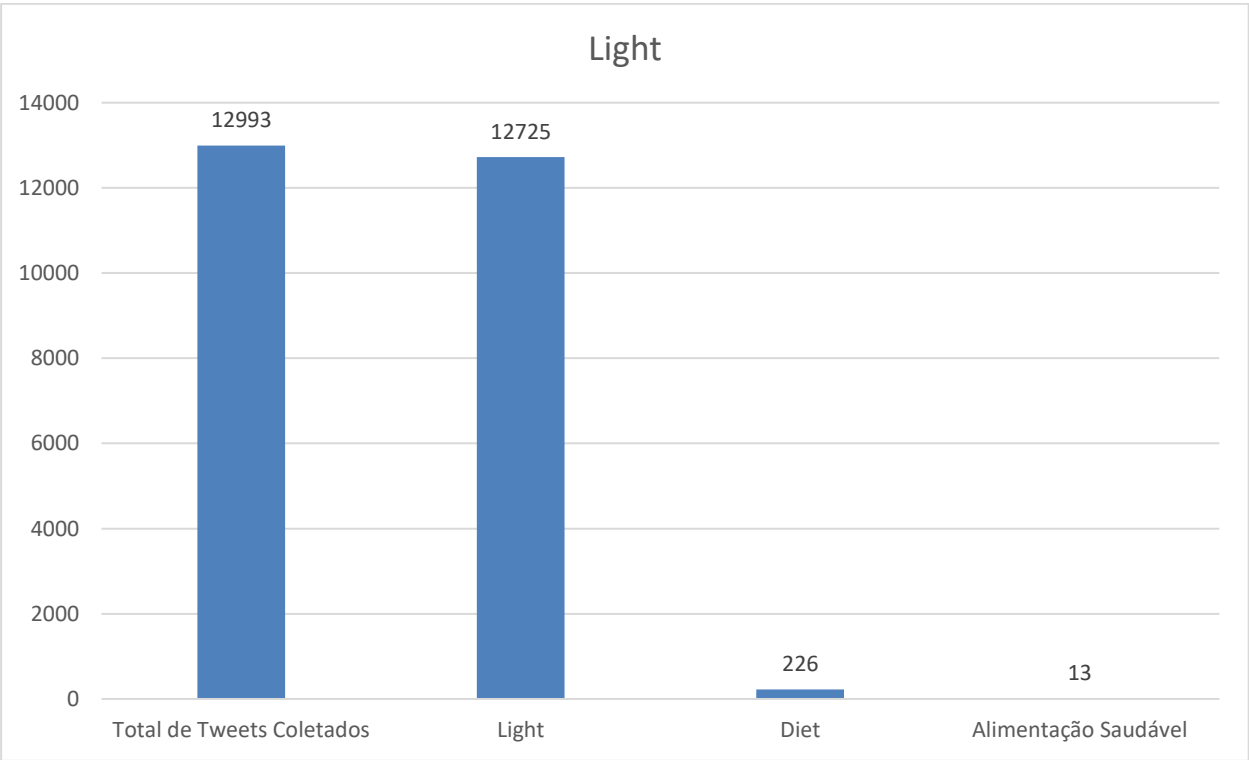
Light

Final cluster centroids:

Attribute	Cluster#											
	Full Data (12993.0)	0 (12725.0)	1 (226.0)	2 (2.0)	3 (13.0)	4 (7.0)	5 (1.0)	6 (2.0)	7 (1.0)	8 (10.0)	9 (4.0)	10 (2.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semilactose	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não
semgluten	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim
lactose	Não	Não	Não	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não
gluten	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
diet	Não	Não	Sim	Não	Não	Não	Sim	Não	Sim	Não	Não	Sim
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não
semleite	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	12725 (98%)
1	226 (2%)
2	2 (0%)
3	13 (0%)
4	7 (0%)
5	1 (0%)
6	2 (0%)
7	1 (0%)
8	10 (0%)
9	4 (0%)
10	2 (0%)



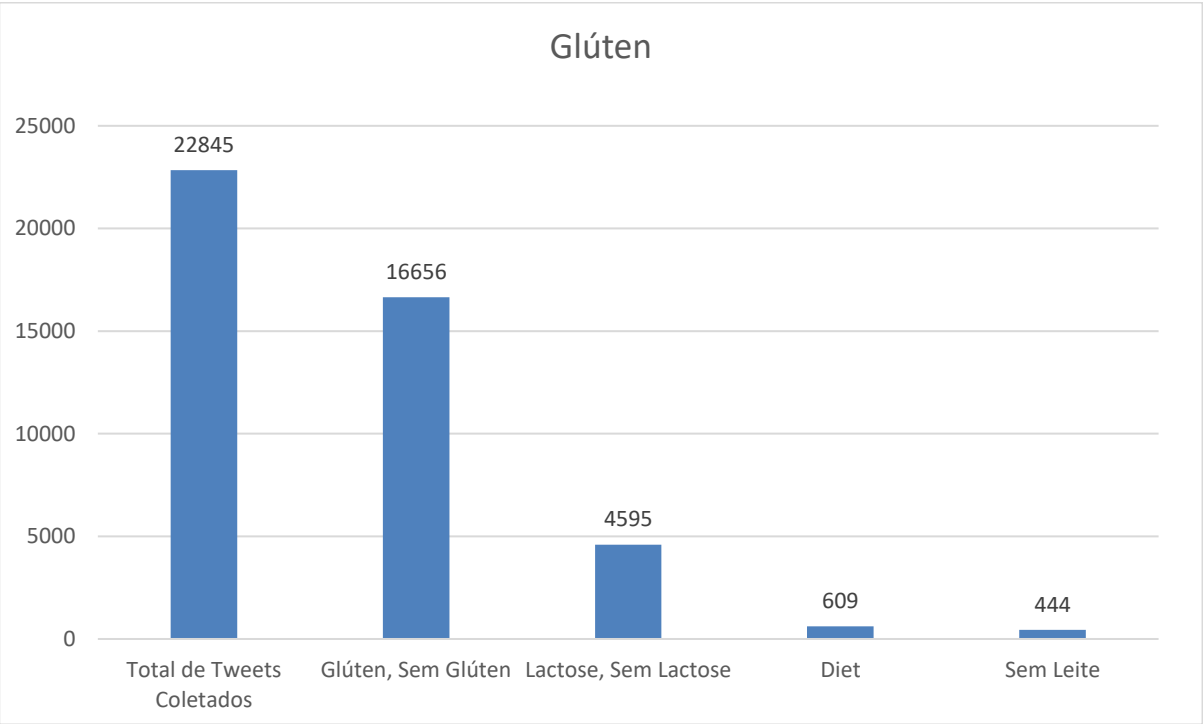
Glúten

Final cluster centroids:

Attribute	Cluster#												
	Full Data (22845.0)	0 (8686.0)	1 (7667.0)	2 (180.0)	3 (3061.0)	4 (1534.0)	5 (231.0)	6 (444.0)	7 (1.0)	8 (108.0)	9 (208.0)	10 (106.0)	11 (619.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Sim	Sim	Não
sem lactose	Não	Não	Não	Sim	Sim	Não	Não	Não	Não	Não	Não	Não	Não
sem gluten	Sim	Sim	Não	Sim	Sim	Não	Não	Sim	Sim	Não	Não	Não	Sim
lactose	Não	Não	Não	Sim	Sim	Sim	Não	Não	Sim	Não	Não	Sim	Não
gluten	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Sim
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaa lactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaog gluten	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem leite	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não

Clustered Instances

0	8686 (38%)
1	7667 (34%)
2	180 (1%)
3	3061 (13%)
4	1534 (7%)
5	231 (1%)
6	444 (2%)
7	1 (0%)
8	108 (0%)
9	208 (1%)
10	106 (0%)
11	619 (3%)



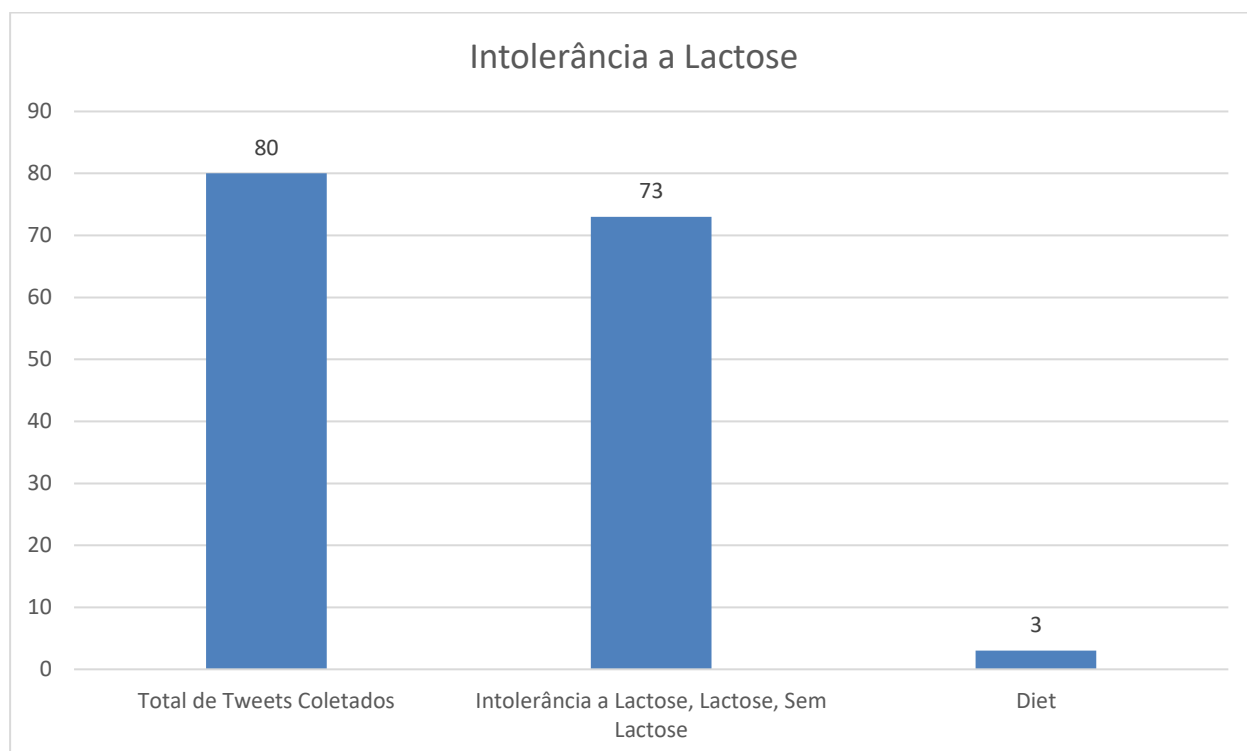
Intolerância a Lactose

Final cluster centroids:

Attribute	Full Data (80.0)	Cluster#						
		0 (68.0)	1 (3.0)	2 (1.0)	3 (1.0)	4 (5.0)	5 (1.0)	6 (1.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não
sem lactose	Não	Não	Não	Não	Não	Sim	Sim	Não
sem gluten	Não	Não	Não	Não	Não	Não	Não	Não
lactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
gluten	Não	Não	Não	Não	Sim	Não	Não	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Sim	Não	Não	Não	Não	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaa lactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
intoleranciaa gluten	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não
sem leite	Não	Não	Não	Sim	Não	Não	Sim	Não
sem ovos	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	68 (85%)
1	3 (4%)
2	1 (1%)
3	1 (1%)
4	5 (6%)
5	1 (1%)
6	1 (1%)



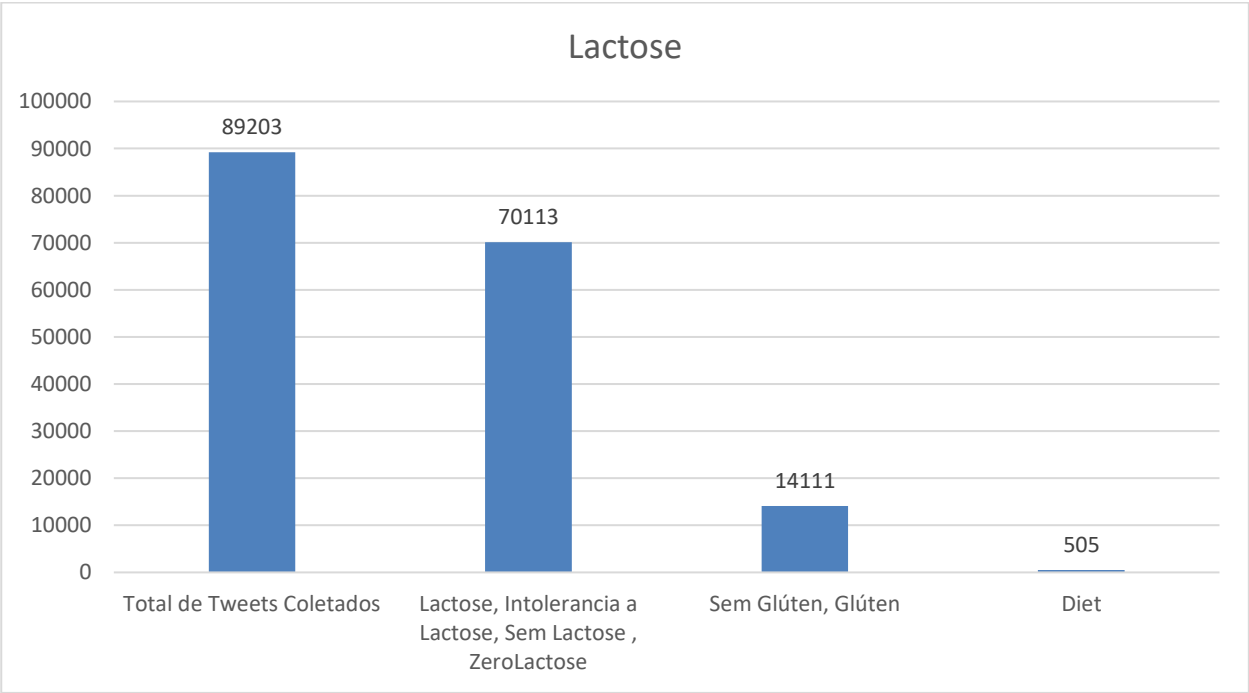
Lactose

Final cluster centroids:

Attribute	Cluster#												
	Full Data (89203.0)	0 (31258.0)	1 (18569.0)	2 (20286.0)	3 (10532.0)	4 (462.0)	5 (2757.0)	6 (3579.0)	7 (205.0)	8 (285.0)	9 (492.0)	10 (505.0)	11 (273.0)
zerolactose	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semilactose	Não	Não	Não	Sim	Sim	Sim	Não	Não	Sim	Não	Sim	Sim	Não
semgluten	Não	Não	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Sim	Não
lactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
gluten	Não	Não	Não	Não	Sim	Sim	Não	Sim	Sim	Sim	Não	Sim	Sim
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Sim
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	31258 (35%)
1	18569 (21%)
2	20286 (23%)
3	10532 (12%)
4	462 (1%)
5	2757 (3%)
6	3579 (4%)
7	205 (0%)
8	285 (0%)
9	492 (1%)
10	505 (1%)
11	273 (0%)



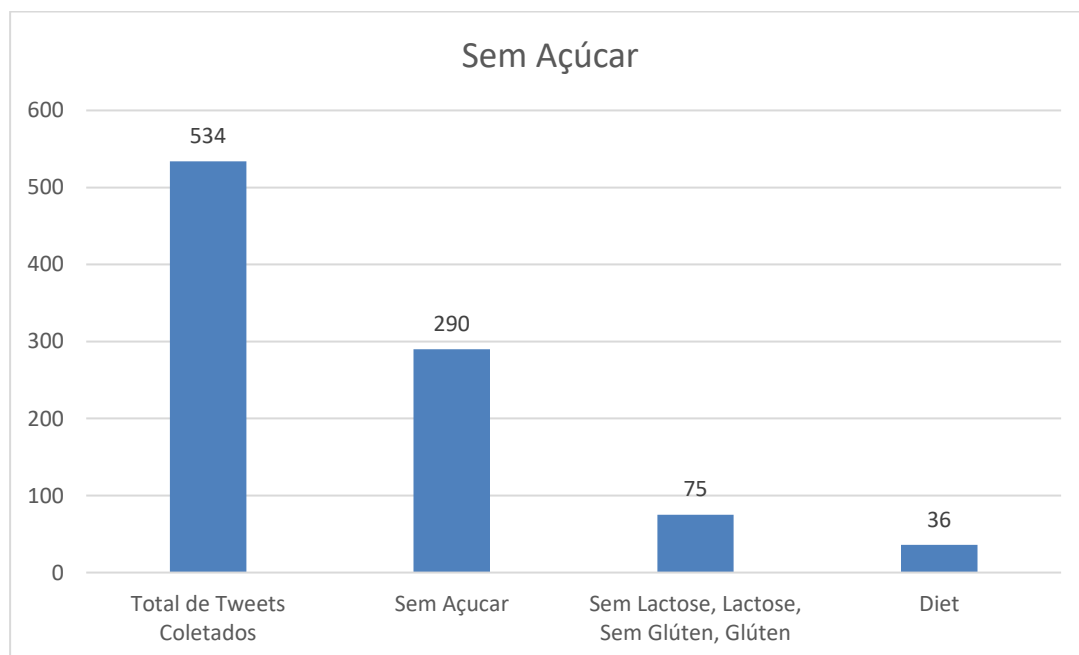
Sem Açúcar

Final cluster centroids:

Attribute	Cluster#												
	Full Data (534.0)	0 (18.0)	1 (4.0)	2 (62.0)	3 (290.0)	4 (75.0)	5 (36.0)	6 (9.0)	7 (29.0)	8 (1.0)	9 (7.0)	10 (2.0)	11 (1.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem lactose	Não	Sim	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim	Sim
sem gluten	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Não	Não	Não	Não
lactose	Não	Sim	Não	Não	Não	Sim	Não	Não	Sim	Sim	Não	Sim	Sim
gluten	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Não	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
semleite	Não	Sim	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	18 (3%)
1	4 (1%)
2	62 (12%)
3	290 (54%)
4	75 (14%)
5	36 (7%)
6	9 (2%)
7	29 (5%)
8	1 (0%)
9	7 (1%)
10	2 (0%)
11	1 (0%)

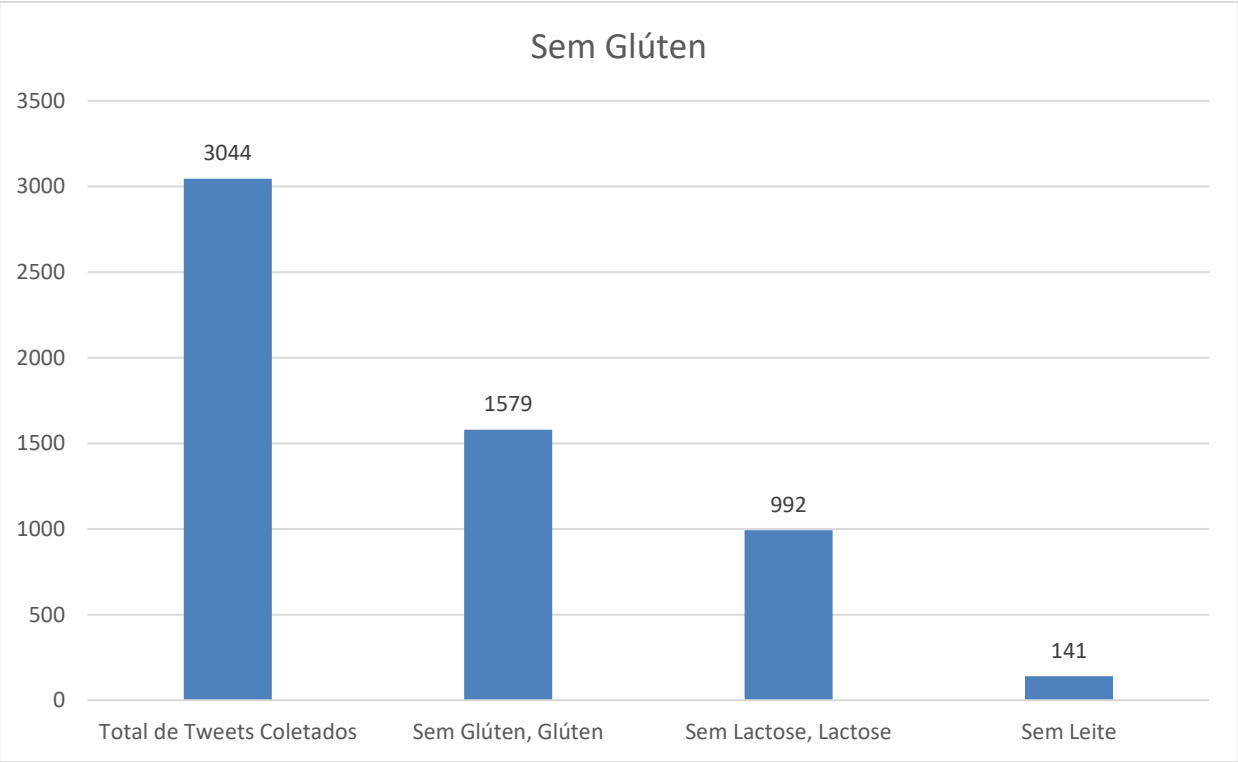


Sem Glúten

Final cluster centroids:														
Attribute	Cluster#													
	Full Data (3044.0)	0 (57.0)	1 (992.0)	2 (1579.0)	3 (58.0)	4 (5.0)	5 (24.0)	6 (13.0)	7 (66.0)	8 (84.0)	9 (141.0)	10 (12.0)	11 (13.0)	
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
sem lactose	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim	Sim	
sem gluten	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	
lactose	Não	Não	Sim	Não	Sim	Sim	Não	Não	Sim	Não	Não	Sim	Sim	
gluten	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	
celiaco	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
diet	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
light	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
semacucar	Não	Sim	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim	Não	
semleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	

Clustered Instances

0	57 (2%)
1	992 (33%)
2	1579 (52%)
3	58 (2%)
4	5 (0%)
5	24 (1%)
6	13 (0%)
7	66 (2%)
8	84 (3%)
9	141 (5%)
10	12 (0%)
11	13 (0%)



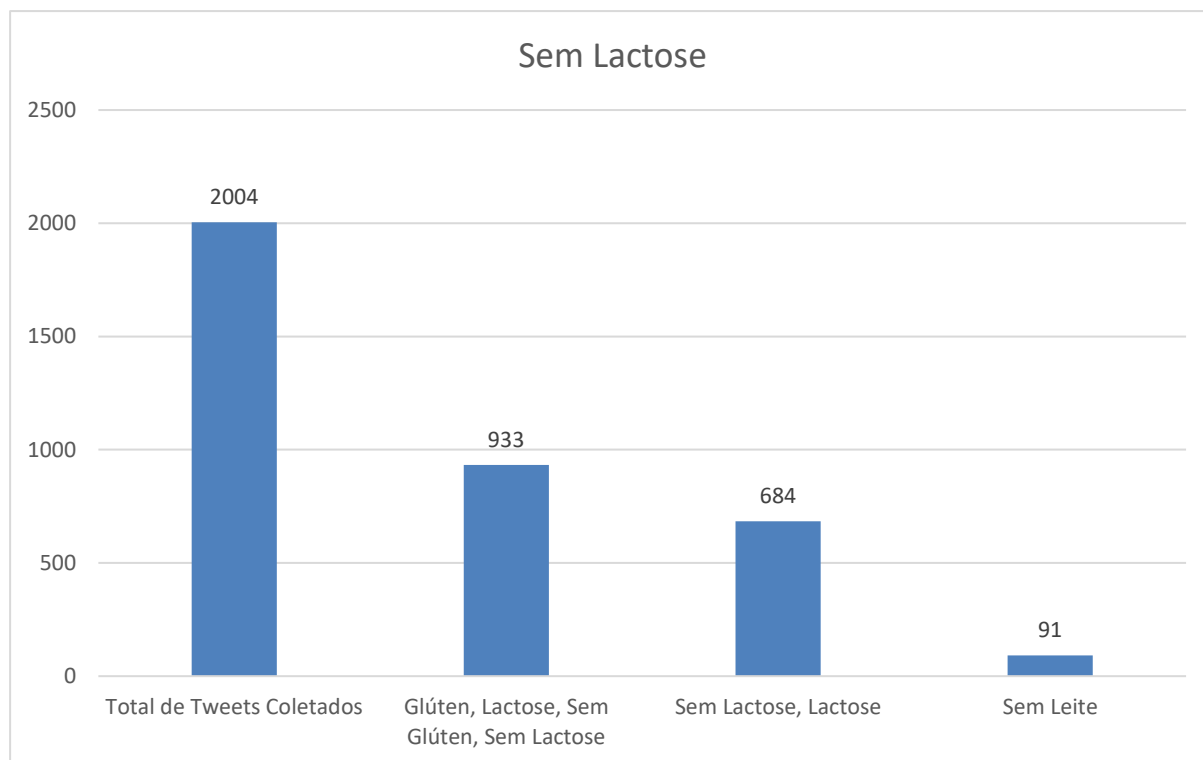
Sem Lactose

Final cluster centroids:

Cluster#													
Attribute	Full Data (2004.0)	0 (933.0)	1 (684.0)	2 (76.0)	3 (33.0)	4 (91.0)	5 (36.0)	6 (58.0)	7 (29.0)	8 (5.0)	9 (21.0)	10 (37.0)	11 (1.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semilactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
semgluten	Sim	Sim	Não	Sim	Não	Sim	Sim	Não	Não	Sim	Não	Não	Não
lactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
gluten	Sim	Sim	Não	Sim	Não	Sim	Sim	Não	Sim	Sim	Não	Não	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim	Sim	Não	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semleite	Não	Não	Não	Não	Sim	Sim	Não	Não	Não	Sim	Não	Não	Sim
semovos	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Sim

Clustered Instances

0	933 (47%)
1	684 (34%)
2	76 (4%)
3	33 (2%)
4	91 (5%)
5	36 (2%)
6	58 (3%)
7	29 (1%)
8	5 (0%)
9	21 (1%)
10	37 (2%)
11	1 (0%)



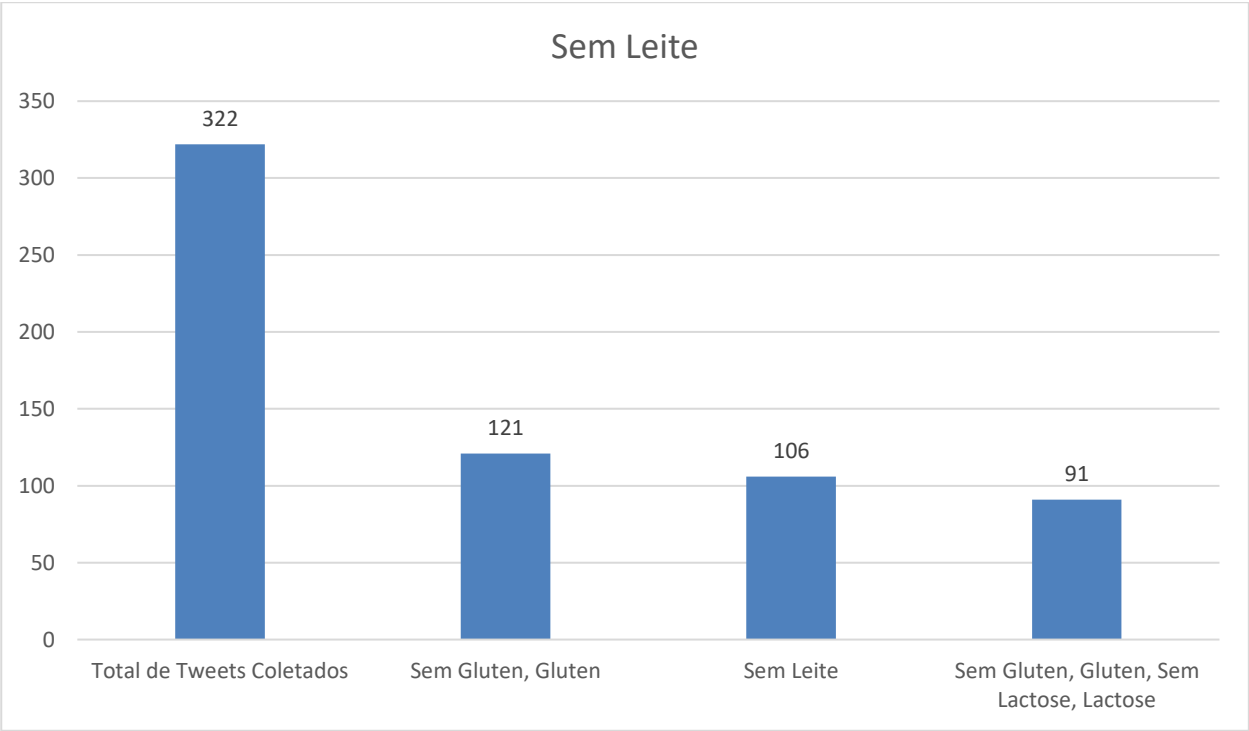
Sem Leite

Final cluster centroids:

Attribute	Cluster#												
	Full Data (322.0)	0 (121.0)	1 (106.0)	2 (9.0)	3 (8.0)	4 (20.0)	5 (2.0)	6 (10.0)	7 (1.0)	8 (1.0)	9 (34.0)	10 (1.0)	11 (9.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semilactose	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Sim	Não
semgluten	Sim	Sim	Não	Não	Sim	Sim	Não	Não	Não	Sim	Sim	Não	Não
lactose	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	Sim	Sim	Não
gluten	Sim	Sim	Não	Não	Sim	Sim	Não	Não	Não	Sim	Sim	Não	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Sim
semleite	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
semovos	Não	Não	Não	Sim	Não	Sim	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	121 (38%)
1	106 (33%)
2	9 (3%)
3	8 (2%)
4	20 (6%)
5	2 (1%)
6	10 (3%)
7	1 (0%)
8	1 (0%)
9	34 (11%)
10	1 (0%)
11	9 (3%)

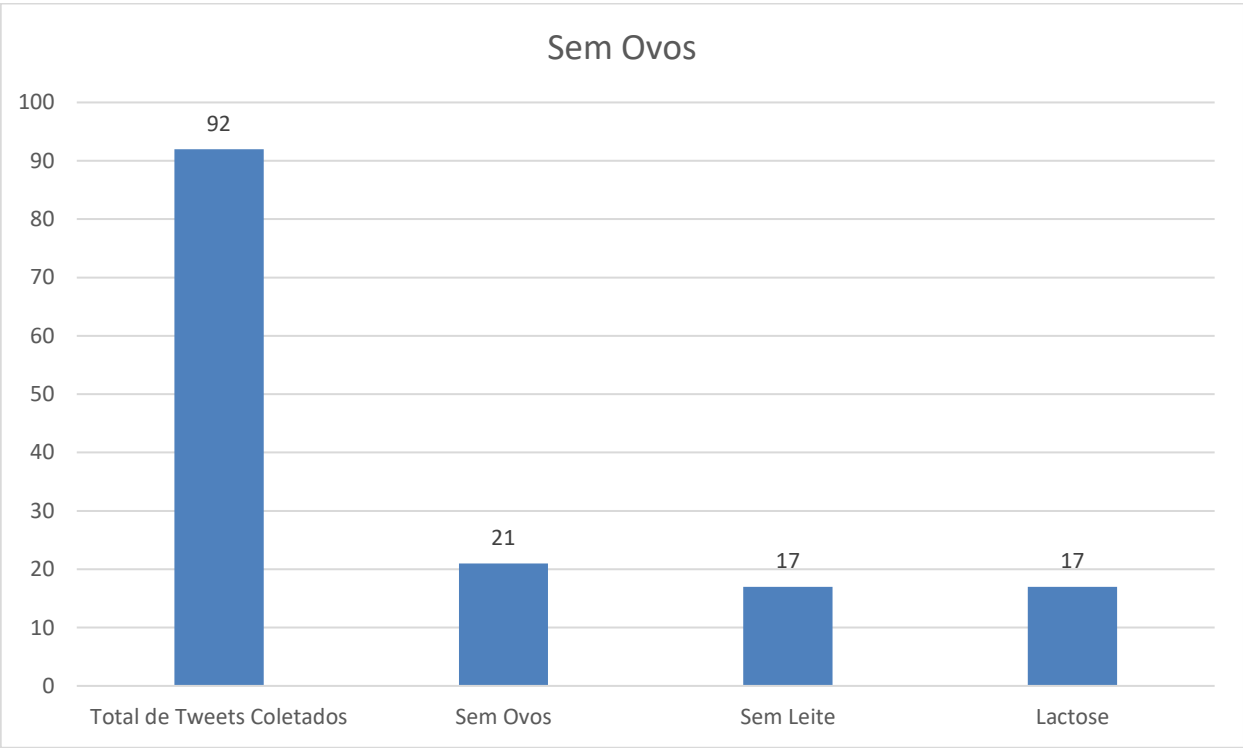


Sem Ovos

Final cluster centroids:													
Attribute	Full Data (92.0)	Cluster#											
		0 (17.0)	1 (21.0)	2 (17.0)	3 (7.0)	4 (9.0)	5 (6.0)	6 (9.0)	7 (2.0)	8 (1.0)	9 (1.0)	10 (1.0)	11 (1.0)
zerolactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
zerogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
sem lactose	Não	Não	Não	Sim	Não	Não	Sim	Não	Sim	Sim	Não	Sim	Sim
sem gluten	Não	Sim	Não	Não	Sim	Não	Sim	Não	Não	Sim	Sim	Não	Não
lactose	Não	Não	Não	Sim	Não	Não	Sim	Não	Sim	Sim	Não	Sim	Sim
gluten	Não	Sim	Não	Não	Sim	Não	Sim	Não	Sim	Sim	Sim	Não	Não
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaa lactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
intoleranciaa gluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	Sim
sem leite	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim	Não
sem ovos	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim
alimentacao saudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	17 (18%)
1	21 (23%)
2	17 (18%)
3	7 (8%)
4	9 (10%)
5	6 (7%)
6	9 (10%)
7	2 (2%)
8	1 (1%)
9	1 (1%)
10	1 (1%)
11	1 (1%)

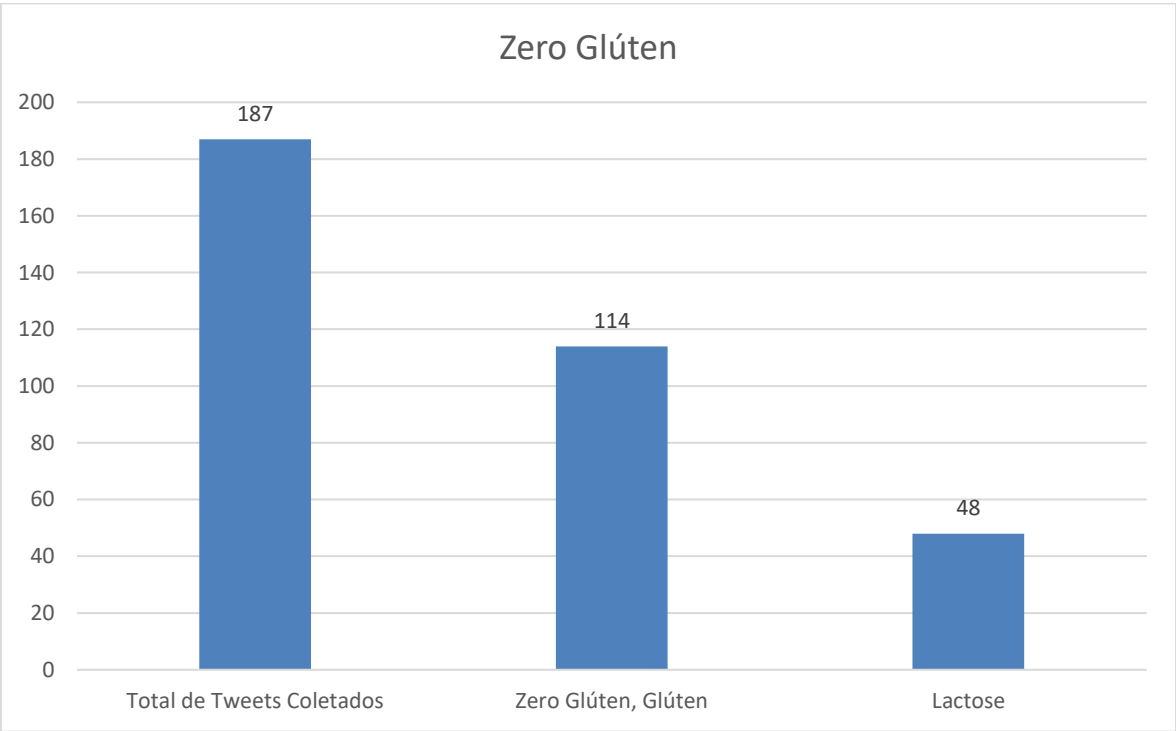


Zero Glúten

Final cluster centroids:														
Attribute	Full Data (187.0)	Cluster#												
		0 (114.0)	1 (48.0)	2 (9.0)	3 (3.0)	4 (1.0)	5 (2.0)	6 (1.0)	7 (4.0)	8 (1.0)	9 (1.0)	10 (1.0)	11 (2.0)	
zerolactose	Não	Não	Sim	Não	Sim	Não	Sim	Não	Não	Sim	Sim	Não	Não	
zerogluten	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	
sem lactose	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim	
sem gluten	Não	Não	Não	Sim	Não	Sim	Sim	Sim	Não	Não	Não	Não	Sim	
lactose	Não	Não	Sim	Não	Sim	Não	Sim	Não	Não	Sim	Sim	Sim	Sim	
gluten	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	
celiaco	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não	
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	
diet	Não	Não	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
light	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	
intoleranciaa lactose	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
intoleranciaa gluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
sem acucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
sem leite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
sem ovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	
alimentacao saudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	

Clustered Instances

0	114 (61%)
1	48 (26%)
2	9 (5%)
3	3 (2%)
4	1 (1%)
5	2 (1%)
6	1 (1%)
7	4 (2%)
8	1 (1%)
9	1 (1%)
10	1 (1%)
11	2 (1%)

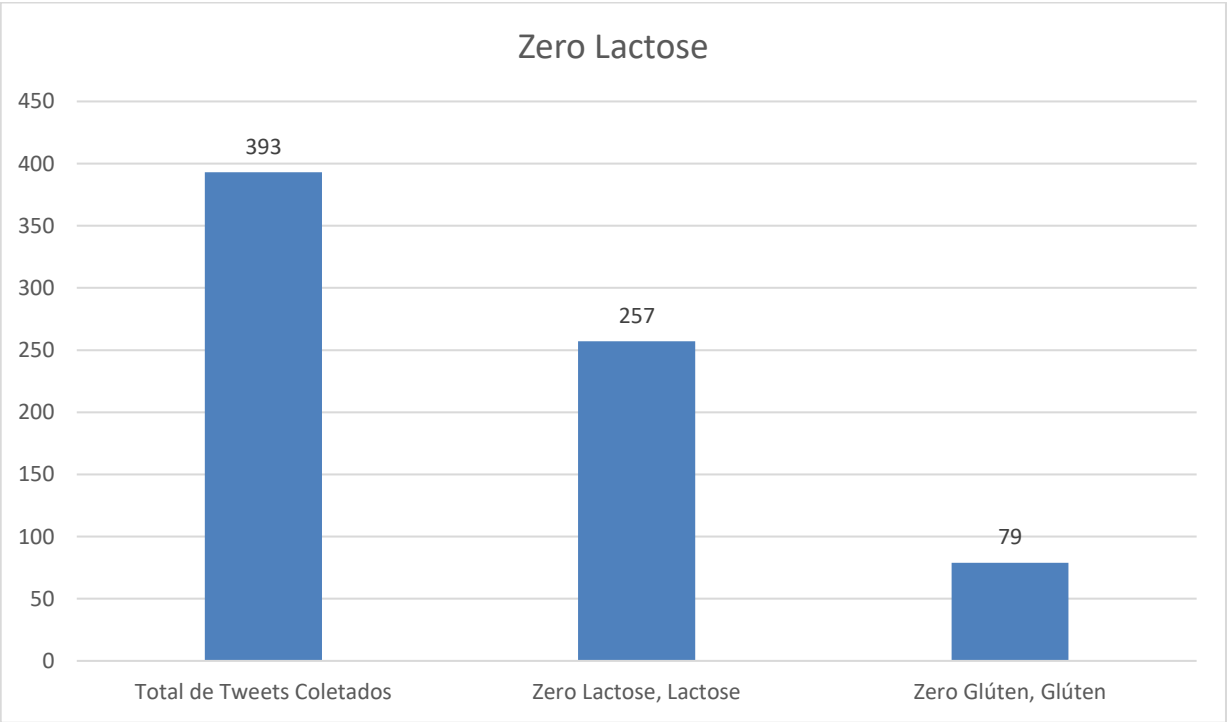


Zero Lactose

Final cluster centroids:													
Attribute	Cluster#												
	Full Data	0	1	2	3	4	5	6	7	8	9	10	11
	(393.0)	(208.0)	(9.0)	(20.0)	(79.0)	(5.0)	(3.0)	(1.0)	(10.0)	(49.0)	(6.0)	(2.0)	(1.0)
zerolactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
zerogluten	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Sim	Sim
sem lactose	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Sim	Sim	Não	Sim
sem gluten	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não
lactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
gluten	Não	Não	Sim	Não	Sim	Não	Não	Não	Sim	Não	Sim	Sim	Sim
celiaco	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diabete	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
diet	Não	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Não
dietetico	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
light	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não
intoleranciaalactose	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não
intoleranciaaogluten	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semacucar	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
semovos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
alimentacaosaudavel	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não

Clustered Instances

0	208 (53%)
1	9 (2%)
2	20 (5%)
3	79 (20%)
4	5 (1%)
5	3 (1%)
6	1 (0%)
7	10 (3%)
8	49 (12%)
9	6 (2%)
10	2 (1%)
11	1 (0%)



ALIMENTOS

Lactose

Final cluster centroids:

Attribute	Full Data (89203.0)	Cluster#										
		0 (77709.0)	1 (2777.0)	2 (1431.0)	3 (1736.0)	4 (3975.0)	5 (488.0)	6 (176.0)	7 (393.0)	8 (304.0)	9 (172.0)	10 (42.0)
lactose	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
docedeleite	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não
cremedeleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não
leitecondensado	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não
cappuccino	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
achocolatado	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
biscoitos	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
massas	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
pao	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não
biscoito	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	Não	Não
bolo	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim
barradecereais	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
iogurte	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não
frutas	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
verduras	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Sim
legumes	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	77709 (87%)
1	2777 (3%)
2	1431 (2%)
3	1736 (2%)
4	3975 (4%)
5	488 (1%)
6	176 (0%)
7	393 (0%)
8	304 (0%)
9	172 (0%)
10	42 (0%)

Glúten

Final cluster centroids:

Attribute	Full Data (26076.0)	Cluster#										
		0 (20673.0)	1 (2490.0)	2 (2301.0)	3 (146.0)	4 (75.0)	5 (2.0)	6 (289.0)	7 (28.0)	8 (28.0)	9 (1.0)	10 (43.0)
gluten	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
docedeleite	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não
cremedeleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
leitecondensado	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não
cappuccino	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
achocolatado	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
biscoitos	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não
massas	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
pao	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Sim	Não
biscoito	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Não	Não
bolo	Não	Não	Sim	Não	Não	Não	Sim	Não	Sim	Não	Sim	Sim
barradecereais	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
iogurte	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim
frutas	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não
verduras	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
legumes	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	20673 (79%)
1	2490 (10%)
2	2301 (9%)
3	146 (1%)
4	75 (0%)
5	2 (0%)
6	289 (1%)
7	28 (0%)
8	28 (0%)
9	1 (0%)
10	43 (0%)

Diet

Final cluster centroids:

Attribute	Full Data (3695.0)	Cluster#										
		0 (3474.0)	1 (16.0)	2 (16.0)	3 (110.0)	4 (29.0)	5 (1.0)	6 (2.0)	7 (24.0)	8 (5.0)	9 (2.0)	10 (16.0)
diet	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
dosedeleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não
cremedeleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
leitecondensado	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não
cappuccino	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não
achocolatado	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim
biscoitos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
massas	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
pao	Não	Não	Não	Não	Não	Sim	Sim	Não	Não	Não	Sim	Não
biscoito	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
bolo	Não	Não	Não	Não	Sim	Não	Sim	Não	Não	Não	Não	Não
barradecereais	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
iogurte	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não
frutas	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não
verduras	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
legumes	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	3474 (94%)
1	16 (0%)
2	16 (0%)
3	110 (3%)
4	29 (1%)
5	1 (0%)
6	2 (0%)
7	24 (1%)
8	5 (0%)
9	2 (0%)
10	16 (0%)

Light

Final cluster centroids:

Attribute	Cluster#											
	Full Data (12993.0)	0 (12591.0)	1 (70.0)	2 (8.0)	3 (172.0)	4 (74.0)	5 (5.0)	6 (37.0)	7 (14.0)	8 (14.0)	9 (4.0)	10 (4.0)
light	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
docedeleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim
cremedeleite	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não
leitecondensado	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
capuccino	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
achocolatado	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
biscoitos	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não
massas	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
pao	Não	Não	Não	Sim	Sim	Não	Não	Não	Não	Não	Sim	Não
biscoito	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	Não	Não	Não
bolo	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Não
barradecereais	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
iogurte	Não	Não	Sim	Sim	Não	Não	Não	Não	Não	Não	Não	Não
frutas	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
verduras	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não
legumes	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não

Clustered Instances

0	12591 (97%)
1	70 (1%)
2	8 (0%)
3	172 (1%)
4	74 (1%)
5	5 (0%)
6	37 (0%)
7	14 (0%)
8	14 (0%)
9	4 (0%)
10	4 (0%)

ANEXO B – APLICAÇÃO DO ALGORITMO APRIORI

Para este estudo de caso, foi aplicado o algoritmo Apriori nos dados coletados do *Twitter* por meio da ferramenta de aplicação Java. Foi possível aplicar o algoritmo por palavra-chave capturada e também por todos os dados em conjunto.

Abaixo é possível visualizar os resultados gerados após a aplicação do algoritmo Apriori.

Conjunto de Todos os Dados

```
Apriori
=====

Minimum support: 0 (1 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 17

Size of set of large itemsets L(2): 113

Size of set of large itemsets L(3): 257

Size of set of large itemsets L(4): 264

Size of set of large itemsets L(5): 143

Size of set of large itemsets L(6): 40

Size of set of large itemsets L(7): 4

Best rules found:

1. semlactose=y 39215 ==> lactose=y 39215    <conf:(1)> lift:(1.94) lev:(0.1) [19020] conv:(19020.2)
2. semgluten=y 28875 ==> gluten=y 28875    <conf:(1)> lift:(4.33) lev:(0.12) [22206] conv:(22206.1)
3. semgluten=y lactose=y 17467 ==> gluten=y 17467    <conf:(1)> lift:(4.33) lev:(0.07) [13432] conv:(13432.86)
4. semlactose=y gluten=y 17335 ==> lactose=y 17335    <conf:(1)> lift:(1.94) lev:(0.04) [8407] conv:(8407.89)
5. semlactose=y semgluten=y 16666 ==> lactose=y 16666    <conf:(1)> lift:(1.94) lev:(0.04) [8083] conv:(8083.4)
6. semlactose=y semgluten=y 16666 ==> gluten=y 16666    <conf:(1)> lift:(4.33) lev:(0.07) [12816] conv:(12816.86)
7. semlactose=y semgluten=y gluten=y 16666 ==> lactose=y 16666    <conf:(1)> lift:(1.94) lev:(0.04) [8083] conv:(8083.4)
8. semlactose=y semgluten=y lactose=y 16666 ==> gluten=y 16666    <conf:(1)> lift:(4.33) lev:(0.07) [12816] conv:(12816.86)
9. semlactose=y semgluten=y 16666 ==> lactose=y gluten=y 16666    <conf:(1)> lift:(7.88) lev:(0.08) [14551] conv:(14551.95)
10. semgluten=y diet=y 1675 ==> gluten=y 1675    <conf:(1)> lift:(4.33) lev:(0.01) [1288] conv:(1288.15)
```

```

11. semlactose=y diet=y 1409 ==> lactose=y 1409    <conf:(1)> lift:(1.94) lev:(0) [683] conv:(683.4)
12. semgluten=y semleite=y 1068 ==> gluten=y 1068   <conf:(1)> lift:(4.33) lev:(0) [821] conv:(821.34)
13. semgluten=y semacucar=y 917 ==> gluten=y 917    <conf:(1)> lift:(4.33) lev:(0) [705] conv:(705.21)
14. zerogluten=y 913 ==> gluten=y 913               <conf:(1)> lift:(4.33) lev:(0) [702] conv:(702.14)
15. semgluten=y lactose=y diet=y 899 ==> gluten=y 899 <conf:(1)> lift:(4.33) lev:(0) [691] conv:(691.37)
16. semlactose=y gluten=y diet=y 873 ==> lactose=y 873 <conf:(1)> lift:(1.94) lev:(0) [423] conv:(423.43)
17. semlactose=y semgluten=y diet=y 831 ==> lactose=y 831 <conf:(1)> lift:(1.94) lev:(0) [403] conv:(403.05)
18. semlactose=y semgluten=y diet=y 831 ==> gluten=y 831 <conf:(1)> lift:(4.33) lev:(0) [639] conv:(639.07)
19. semlactose=y semgluten=y gluten=y diet=y 831 ==> lactose=y 831 <conf:(1)> lift:(1.94) lev:(0) [403] conv:(403.05)
20. semlactose=y semgluten=y lactose=y diet=y 831 ==> gluten=y 831 <conf:(1)> lift:(4.33) lev:(0) [639] conv:(639.07)
21. semlactose=y semgluten=y diet=y 831 ==> lactose=y gluten=y 831 <conf:(1)> lift:(7.88) lev:(0) [725] conv:(725.59)
22. zerolactose=y gluten=y 634 ==> lactose=y 634    <conf:(1)> lift:(1.94) lev:(0) [307] conv:(307.51)
23. semlactose=y light=y 575 ==> lactose=y 575      <conf:(1)> lift:(1.94) lev:(0) [278] conv:(278.89)
24. zerogluten=y lactose=y 562 ==> gluten=y 562     <conf:(1)> lift:(4.33) lev:(0) [432] conv:(432.2)
25. semlactose=y semacucar=y 544 ==> lactose=y 544  <conf:(1)> lift:(1.94) lev:(0) [263] conv:(263.85)
26. semgluten=y lactose=y semacucar=y 518 ==> gluten=y 518 <conf:(1)> lift:(4.33) lev:(0) [398] conv:(398.36)
27. intoleranciaaogluten=y 511 ==> gluten=y 511    <conf:(1)> lift:(4.33) lev:(0) [392] conv:(392.98)
28. semlactose=y gluten=y semacucar=y 510 ==> lactose=y 510 <conf:(1)> lift:(1.94) lev:(0) [247] conv:(247.36)
29. semlactose=y semgluten=y semacucar=y 505 ==> lactose=y 505 <conf:(1)> lift:(1.94) lev:(0) [244] conv:(244.94)
30. semlactose=y semgluten=y semacucar=y 505 ==> gluten=y 505 <conf:(1)> lift:(4.33) lev:(0) [388] conv:(388.37)
31. semlactose=y semgluten=y gluten=y semacucar=y 505 ==> lactose=y 505 <conf:(1)> lift:(1.94) lev:(0) [244] conv:(244.94)
32. semlactose=y semgluten=y lactose=y semacucar=y 505 ==> gluten=y 505 <conf:(1)> lift:(4.33) lev:(0) [388] conv:(388.37)
33. semlactose=y semgluten=y semacucar=y 505 ==> lactose=y gluten=y 505 <conf:(1)> lift:(7.88) lev:(0) [440] conv:(440.94)
34. zerolactose=y zerogluten=y 481 ==> lactose=y 481 <conf:(1)> lift:(1.94) lev:(0) [233] conv:(233.3)
35. zerolactose=y zerogluten=y 481 ==> gluten=y 481 <conf:(1)> lift:(4.33) lev:(0) [369] conv:(369.91)
36. zerolactose=y zerogluten=y gluten=y 481 ==> lactose=y 481 <conf:(1)> lift:(1.94) lev:(0) [233] conv:(233.3)
37. zerolactose=y zerogluten=y lactose=y 481 ==> gluten=y 481 <conf:(1)> lift:(4.33) lev:(0) [369] conv:(369.91)
38. zerolactose=y zerogluten=y 481 ==> lactose=y gluten=y 481 <conf:(1)> lift:(7.88) lev:(0) [419] conv:(419.99)
39. semgluten=y light=y 438 ==> gluten=y 438       <conf:(1)> lift:(4.33) lev:(0) [336] conv:(336.84)
40. semlactose=y semleite=y 436 ==> lactose=y 436  <conf:(1)> lift:(1.94) lev:(0) [211] conv:(211.47)
41. gluten=y intoleranciaaolactose=y 423 ==> lactose=y 423 <conf:(1)> lift:(1.94) lev:(0) [205] conv:(205.17)
42. semgluten=y lactose=y semleite=y 322 ==> gluten=y 322 <conf:(1)> lift:(4.33) lev:(0) [247] conv:(247.63)
43. semlactose=y gluten=y semleite=y 294 ==> lactose=y 294 <conf:(1)> lift:(1.94) lev:(0) [142] conv:(142.6)
44. semlactose=y semgluten=y semleite=y 293 ==> lactose=y 293 <conf:(1)> lift:(1.94) lev:(0) [142] conv:(142.11)
45. semlactose=y semgluten=y semleite=y 293 ==> gluten=y 293 <conf:(1)> lift:(4.33) lev:(0) [225] conv:(225.33)

```

Alimentação Saudável

Apriori

=====

Minimum support: 0.1 (378 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 2

Size of set of large itemsets L(2): 1

Best rules found:

1. diet=y 464 ==> alimentacaosaudavel=y 464 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Celiaco

Minimum support: 0.1 (141 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 2

Size of set of large itemsets L(2): 1

Best rules found:

1. gluten=y 331 ==> celiaco=y 329 <conf:(0.99)> lift:(1.01) lev:(0) [1] conv:(1.33)

Diabetes

Apriori
=====

Minimum support: 0 (1 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 6

Size of set of large itemsets L(4): 1

Best rules found:

1. diet=y 160 ==> diabete=y 160 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. semacucar=y 22 ==> diabete=y 22 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. light=y 15 ==> diabete=y 15 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. lactose=y 11 ==> diabete=y 11 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. gluten=y 6 ==> diabete=y 6 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. intoleranciaalactose=y 5 ==> lactose=y 5 <conf:(1)> lift:(3347.09) lev:(0) [4] conv:(5)
7. intoleranciaalactose=y 5 ==> diabete=y 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. diabete=y intoleranciaalactose=y 5 ==> lactose=y 5 <conf:(1)> lift:(3347.09) lev:(0) [4] conv:(5)
9. lactose=y intoleranciaalactose=y 5 ==> diabete=y 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. intoleranciaalactose=y 5 ==> lactose=y diabete=y 5 <conf:(1)> lift:(3347.09) lev:(0) [4] conv:(5)

Diet

Apriori

=====

Minimum support: 0 (1 instances)

Minimum metric <confidence>: 0.1

Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 34

Size of set of large itemsets L(3): 45

Size of set of large itemsets L(4): 31

Size of set of large itemsets L(5): 12

Size of set of large itemsets L(6): 2

Best rules found:

1. light=y 164 ==> diet=y 164 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. lactose=y 47 ==> diet=y 47 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. gluten=y 35 ==> diet=y 35 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. semacucar=y 32 ==> diet=y 32 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. diabete=y 26 ==> diet=y 26 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. semgluten=y 25 ==> gluten=y 25 <conf:(1)> lift:(105.57) lev:(0.01) [24] conv:(24.76)
7. semgluten=y 25 ==> diet=y 25 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. semgluten=y diet=y 25 ==> gluten=y 25 <conf:(1)> lift:(105.57) lev:(0.01) [24] conv:(24.76)
9. semgluten=y gluten=y 25 ==> diet=y 25 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. semgluten=y 25 ==> gluten=y diet=y 25 <conf:(1)> lift:(105.57) lev:(0.01) [24] conv:(24.76)

Dietético

Apriori

=====

Minimum support: 0 (1 instances)

Minimum metric <confidence>: 0.1

Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 3

Best rules found:

1. light=y 15 ==> dietetico=y 15 <conf:(1)> lift:(1.02) lev:(0) [0] conv:(0.37)
2. gluten=y 7 ==> dietetico=y 7 <conf:(1)> lift:(1.02) lev:(0) [0] conv:(0.17)
3. semgluten=y 5 ==> gluten=y 5 <conf:(1)> lift:(205.29) lev:(0) [4] conv:(4.98)
4. semgluten=y 5 ==> dietetico=y 5 <conf:(1)> lift:(1.02) lev:(0) [0] conv:(0.12)
5. semacucar=y 5 ==> dietetico=y 5 <conf:(1)> lift:(1.02) lev:(0) [0] conv:(0.12)
6. semgluten=y dietetico=y 5 ==> gluten=y 5 <conf:(1)> lift:(205.29) lev:(0) [4] conv:(4.98)
7. semgluten=y gluten=y 5 ==> dietetico=y 5 <conf:(1)> lift:(1.02) lev:(0) [0] conv:(0.12)
8. semgluten=y 5 ==> gluten=y dietetico=y 5 <conf:(1)> lift:(205.29) lev:(0) [4] conv:(4.98)
9. diet=y light=y 3 ==> dietetico=y 3 <conf:(1)> lift:(1.02) lev:(0) [0] conv:(0.07)
10. lactose=y 1 ==> semlactose=y 1 <conf:(1)> lift:(1437) lev:(0) [0] conv:(1)

Light

Apriori

=====

Minimum support: 0 (1 instances)

Minimum metric <confidence>: 0.1

Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 27

Size of set of large itemsets L(3): 30

Size of set of large itemsets L(4): 16

Size of set of large itemsets L(5): 3

Best rules found:

1. diet=y 230 ==> light=y 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. gluten=y 19 ==> light=y 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. semgluten=y 13 ==> gluten=y 13 <conf:(1)> lift:(683.84) lev:(0) [12] conv:(12.98)
4. semgluten=y 13 ==> light=y 13 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. alimentacaosaudavel=y 13 ==> light=y 13 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. semgluten=y light=y 13 ==> gluten=y 13 <conf:(1)> lift:(683.84) lev:(0) [12] conv:(12.98)
7. semgluten=y gluten=y 13 ==> light=y 13 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. semgluten=y 13 ==> gluten=y light=y 13 <conf:(1)> lift:(683.84) lev:(0) [12] conv:(12.98)
9. lactose=y 12 ==> light=y 12 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. semlactose=y 11 ==> lactose=y 11 <conf:(1)> lift:(1082.75) lev:(0) [10] conv:(10.99)

Glúten

Minimum support: 0.1 (2285 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

1. semgluten=y 12830 ==> gluten=y 12830 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. lactose=y 5546 ==> gluten=y 5546 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. semgluten=y lactose=y 3745 ==> gluten=y 3745 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. semlactose=y 3187 ==> lactose=y 3187 <conf:(1)> lift:(4.12) lev:(0.11) [2413] conv:(2413.3)
5. semlactose=y 3187 ==> gluten=y 3187 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. semlactose=y gluten=y 3187 ==> lactose=y 3187 <conf:(1)> lift:(4.12) lev:(0.11) [2413] conv:(2413.3)
7. semlactose=y lactose=y 3187 ==> gluten=y 3187 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. semlactose=y 3187 ==> lactose=y gluten=y 3187 <conf:(1)> lift:(4.12) lev:(0.11) [2413] conv:(2413.3)
9. semlactose=y semgluten=y 3025 ==> lactose=y 3025 <conf:(1)> lift:(4.12) lev:(0.1) [2290] conv:(2290.63)
10. semlactose=y semgluten=y 3025 ==> gluten=y 3025 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Intolerância a Lactose

Minimum support: 0.1 (8 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 2

Size of set of large itemsets L(2): 1

Best rules found:

1. lactose=y 79 ==> intoleranciaalactose=y 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. intoleranciaalactose=y 80 ==> lactose=y 79 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.5)

Lactose

Minimum support: 0.1 (8920 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 7

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

1. semlactose=y 32598 ==> lactose=y 32598 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. intoleranciaalactose=y 18870 ==> lactose=y 18870 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. gluten=y 16094 ==> lactose=y 16094 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. semlactose=y gluten=y 11704 ==> lactose=y 11704 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. semgluten=y 11242 ==> semlactose=y 11242 <conf:(1)> lift:(2.74) lev:(0.08) [7133] conv:(7133.77)
6. semgluten=y 11242 ==> lactose=y 11242 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. semgluten=y 11242 ==> gluten=y 11242 <conf:(1)> lift:(5.54) lev:(0.1) [9213] conv:(9213.72)
8. semgluten=y lactose=y 11242 ==> semlactose=y 11242 <conf:(1)> lift:(2.74) lev:(0.08) [7133] conv:(7133.77)
9. semlactose=y semgluten=y 11242 ==> lactose=y 11242 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. semgluten=y 11242 ==> semlactose=y lactose=y 11242 <conf:(1)> lift:(2.74) lev:(0.08) [7133] conv:(7133.77)

Sem Açúcar

Minimum support: 0.2 (107 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 2

Best rules found:

1. gluten=y 165 ==> semacucar=y 165 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. semgluten=y 158 ==> gluten=y 158 <conf:(1)> lift:(3.24) lev:(0.2) [109] conv:(109.18)
3. semgluten=y 158 ==> semacucar=y 158 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. semgluten=y semacucar=y 158 ==> gluten=y 158 <conf:(1)> lift:(3.24) lev:(0.2) [109] conv:(109.18)
5. semgluten=y gluten=y 158 ==> semacucar=y 158 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. semgluten=y 158 ==> gluten=y semacucar=y 158 <conf:(1)> lift:(3.24) lev:(0.2) [109] conv:(109.18)
7. lactose=y 129 ==> semacucar=y 129 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. semlactose=y 125 ==> lactose=y 125 <conf:(1)> lift:(4.14) lev:(0.18) [94] conv:(94.8)
9. semlactose=y 125 ==> semacucar=y 125 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. semlactose=y semacucar=y 125 ==> lactose=y 125 <conf:(1)> lift:(4.14) lev:(0.18) [94] conv:(94.8)

Sem Glúten

Minimum support: 0.35 (1065 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 13

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

1. gluten=y 3044 ==> semgluten=y 3044 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. semgluten=y 3044 ==> gluten=y 3044 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. lactose=y 1146 ==> semgluten=y 1146 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. lactose=y 1146 ==> gluten=y 1146 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. lactose=y gluten=y 1146 ==> semgluten=y 1146 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. semgluten=y lactose=y 1146 ==> gluten=y 1146 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. lactose=y 1146 ==> semgluten=y gluten=y 1146 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. semlactose=y 1088 ==> semgluten=y 1088 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. semlactose=y 1088 ==> lactose=y 1088 <conf:(1)> lift:(2.66) lev:(0.22) [678] conv:(678.39)
10. semlactose=y 1088 ==> gluten=y 1088 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Sem Lactose

Minimum support: 0.55 (1102 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 9

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

1. lactose=y 2004 ==> semlactose=y 2004 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. semlactose=y 2004 ==> lactose=y 2004 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. gluten=y 1172 ==> semlactose=y 1172 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. gluten=y 1172 ==> lactose=y 1172 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. lactose=y gluten=y 1172 ==> semlactose=y 1172 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. semlactose=y gluten=y 1172 ==> lactose=y 1172 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. gluten=y 1172 ==> semlactose=y lactose=y 1172 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. semgluten=y 1137 ==> semlactose=y 1137 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. semgluten=y 1137 ==> lactose=y 1137 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. semgluten=y 1137 ==> gluten=y 1137 <conf:(1)> lift:(1.71) lev:(0.24) [472] conv:(472.05)

Sem Leite

Minimum support: 0.1 (32 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 10

Size of set of large itemsets L(4): 5

Size of set of large itemsets L(5): 1

Best rules found:

1. gluten=y 185 ==> semleite=y 185 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. semgluten=y 181 ==> gluten=y 181 <conf:(1)> lift:(1.74) lev:(0.24) [77] conv:(77.01)
3. semgluten=y 181 ==> semleite=y 181 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. semgluten=y semleite=y 181 ==> gluten=y 181 <conf:(1)> lift:(1.74) lev:(0.24) [77] conv:(77.01)
5. semgluten=y gluten=y 181 ==> semleite=y 181 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. semgluten=y 181 ==> gluten=y semleite=y 181 <conf:(1)> lift:(1.74) lev:(0.24) [77] conv:(77.01)
7. lactose=y 47 ==> semleite=y 47 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. semlactose=y 45 ==> lactose=y 45 <conf:(1)> lift:(6.85) lev:(0.12) [38] conv:(38.43)
9. semlactose=y 45 ==> semleite=y 45 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. semlactose=y semleite=y 45 ==> lactose=y 45 <conf:(1)> lift:(6.85) lev:(0.12) [38] conv:(38.43)

Sem Ovos

Minimum support: 0.3 (28 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 2

Best rules found:

1. gluten=y 34 ==> semovos=y 34 <conf:(1)> lift:(1.11) lev:(0.04) [3] conv:(3.33)
2. semgluten=y 32 ==> gluten=y 32 <conf:(1)> lift:(2.71) lev:(0.22) [20] conv:(20.17)
3. semgluten=y 32 ==> semovos=y 32 <conf:(1)> lift:(1.11) lev:(0.03) [3] conv:(3.13)
4. semgluten=y semovos=y 32 ==> gluten=y 32 <conf:(1)> lift:(2.71) lev:(0.22) [20] conv:(20.17)
5. semgluten=y gluten=y 32 ==> semovos=y 32 <conf:(1)> lift:(1.11) lev:(0.03) [3] conv:(3.13)
6. semgluten=y 32 ==> gluten=y semovos=y 32 <conf:(1)> lift:(2.71) lev:(0.22) [20] conv:(20.17)
7. lactose=y 28 ==> semlactose=y 28 <conf:(1)> lift:(3.29) lev:(0.21) [19] conv:(19.48)
8. semlactose=y 28 ==> lactose=y 28 <conf:(1)> lift:(3.29) lev:(0.21) [19] conv:(19.48)
9. semlactose=y 28 ==> semovos=y 28 <conf:(1)> lift:(1.11) lev:(0.03) [2] conv:(2.74)
10. lactose=y 28 ==> semovos=y 28 <conf:(1)> lift:(1.11) lev:(0.03) [2] conv:(2.74)

Zero Glúten

Minimum support: 0.25 (47 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

1. gluten=y 187 ==> zerogluten=y 187 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. zerogluten=y 187 ==> gluten=y 187 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. lactose=y 60 ==> zerogluten=y 60 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. lactose=y 60 ==> gluten=y 60 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. lactose=y gluten=y 60 ==> zerogluten=y 60 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. zerogluten=y lactose=y 60 ==> gluten=y 60 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. lactose=y 60 ==> zerogluten=y gluten=y 60 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. zerolactose=y 55 ==> zerogluten=y 55 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. zerolactose=y 55 ==> lactose=y 55 <conf:(1)> lift:(3.12) lev:(0.2) [37] conv:(37.35)
10. zerolactose=y 55 ==> gluten=y 55 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Zero Lactose

Minimum support: 0.2 (79 instances)
 Minimum metric <confidence>: 0.9
 Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

1. lactose=y 391 ==> zerolactose=y 391 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. gluten=y 109 ==> zerolactose=y 109 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. gluten=y 109 ==> lactose=y 109 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.55)
4. lactose=y gluten=y 109 ==> zerolactose=y 109 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. zerolactose=y gluten=y 109 ==> lactose=y 109 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.55)
6. gluten=y 109 ==> zerolactose=y lactose=y 109 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.55)
7. zerogluten=y 84 ==> zerolactose=y 84 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. zerogluten=y 84 ==> lactose=y 84 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.43)
9. zerogluten=y 84 ==> gluten=y 84 <conf:(1)> lift:(3.61) lev:(0.15) [60] conv:(60.7)
10. zerogluten=y lactose=y 84 ==> zerolactose=y 84 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

ANEXO C – QUESTIONÁRIO – ENTENDIMENTO DO NEGÓCIO

Para o entendimento do negócio, primeira etapa do plano de trabalho deste estudo de caso, foi realizada uma pesquisa juntamente com a empresa ALLPS para alinhar o conhecimento atual da empresa com as coletas, análise e possíveis resultados deste trabalho.

Abaixo apresenta-se as perguntas realizadas a empresa ALLPS e as respostas retornadas pela empresa.

Perguntas Autor	Respostas ALLPS
Quais tipos de restrições tendem a ocorrer de forma conjunta?	Glúten e Lactose, Glúten e Diabetes
Qual restrição alimentar ocorre com mais frequência nas solicitações dos seus clientes?	Glúten e Lactose
Quais tipos de restrição alimentar a empresa mais se preocupa na hora escolher seus produtos?	Glúten e Lactose
Qual tipo de restrição alimentar é mais importante para o negócio da empresa?	Glúten
Quais as principais características dos produtos vendidos pela empresa?	Produtos Sem Glúten, Sem Lactose, com Baixo Teor de Sódio, Sem Conservantes, Com Fibras
Qual alimento as pessoas mais procuram sem lactose?	Doce de Leite, Creme de Leite, Leite Condensado, Cappuccino, Achocolatado, Biscoitos, Iogurte
Qual alimento as pessoas mais comentam sem glúten?	Massas, Pães, Biscoitos, Bolos, Barra de Cereais, Salgados.

Que tipo de informação a empresa gostaria de obter, para ajudar no marketing digital do produto?	Quais outros tipos de restrições são tão solicitados quanto Glúten e Lactose? Quais tipos de restrições, além do Glúten e Lactose ocorrem com alguma frequência?
Quais palavras-chave a empresa gostaria de obter uma análise?	Lactose, Glúten, Sem Lactose, Sem Glúten, Zero Lactose, Zero Glúten, Diet, Light, Sem Leite, Sem Ovos, Sem Açúcar, Diabete, Dietético, Celíaco, Intolerância a Lactose, Intolerância a Glúten, Alimentação Saudável

ANEXO D – QUESTIONÁRIO DE RESULTADO

Após as análises concluídas, os grafos, gráficos e resultados foram apresentados a empresa ALLPS Alimentos Saudáveis. Após, foi entregue a empresa um questionário para que a sua direção analisasse os resultados e respondesse. Na tabela abaixo encontrasse as perguntas e o retorno da empresa.

Perguntas Autor	Respostas ALLPS
1. Os resultados, grafos e gráficos apresentados após a aplicação de algoritmos de agrupamento e associação ficaram claros para a empresa?	Sim, os resultados ficaram claros e foram esclarecedores com relação a algumas informações que não obtínhamos conhecimento. Foi importante também ter conhecimento de quais alimentos são mais procurados por tipo de restrição, dessa forma fica mais fácil definir quais produtos estão presentes em nossas caixas, com mais frequência.
2. Os resultados apresentados trouxeram alguma novidade a empresa?	Sim, principalmente com relação a demanda em produtos Light e Diet. Não tínhamos conhecimento total das restrições mais procuradas pelo nosso público-alvo.
3. Houve algum resultado que apresentou a empresa alguma ideia de mudança do foco atual?	Sim, estamos estudando a possibilidade de oferecermos caixas focadas em pessoas que procuram produtos Diet e Light. Hoje trabalhamos com caixas sem lactose e sem glúten apenas.
4. Houve algum resultado que alterará a análise SWOT da empresa?	Sim, agora obtemos mais conhecimento do nosso público alvo e mudaremos o tipo de restrição alvo da empresa, que não

	será apenas produtos sem glúten e sem lactose.
5. Os resultados realizados após a coleta de <i>tweets</i> referente ao glúten, lactose, diet e light trouxeram alguma informação que a empresa não tinha conhecimento?	Sim, o foco inicial da empresa sempre foi um público com restrições a glúten e lactose. A quantidade, mesmo que inferior, de citações com Diet e Light nos fez perceber que há u público ainda não atingido que podem ter interesse em nosso produto.
6. Os resultados referentes aos grãos e gráficos de redes das concorrentes, coletados do <i>Facebook</i> , trouxeram alguma informação importante a empresa?	Sim, foi importante analisar que nossos maiores concorrentes também possuem um foco muito parecido ao nosso. Com isso, alterando nossos produtos, poderemos dar um passo a frente de nossos concorrentes e atingir um público maior.
7. O volume de postagens realizadas no Instagram, apresentados por meio de um grafo de rede, referente ao glúten, lactose, diet e light, ajudaram a empresa de alguma forma?	Sim, assim como foi com o <i>twitter</i> , foi muito importante para nós visualizarmos que também no Instagram as pessoas procuram muito por produtos Light e Diet. O Instagram é um dos principais meios de marketing digital que a empresa trabalha hoje. Assim poderemos patrocinar posts referentes a produto Diet e Light sabendo que há um público a ser atingido.
8. A Nuvem de <i>Tags</i> apresentada, referente a todas coletas realizadas no <i>Twitter</i> , auxiliaram a empresa?	Sim, foi possível identificar outros tipos de restrições e até mesmo tipos de alimentos que poderemos incluir em nossas caixas.

9. A empresa sentiu a necessidade de mais alguma informação não apresentada neste Estudo de Caso?	Gostaríamos de obter mais informações dos usuários dessas redes sociais, como gênero, idade e se possível até mesmo localidade.
10. Em termos gerais, como este estudo de caso ajudou, ou não, a empresa?	Sim, em termos gerais nos auxiliou a entender que a empresa está no caminho correto, mas que precisa incluir alguns novos produtos para que consigamos atingir um público maior e ter um diferencial a mais de seus concorrentes. Foi importante também pois dessa forma saberemos focar melhor nosso marketing digital, sem ter gastos desnecessários, focando diretamente em restrições e produtos de interesse do nosso público, atingindo assim o marketing <i>one-to-one</i> .